
بسم الله الرحمن الرحيم

سنجه‌های وبی

برای متخصصان کتابداری و اطلاع‌رسانی

دیوید استوارت

مترجم:

مریم خسروی

با همکاری:

فاطمه بهمنی

ویراستار:

حمید کشاورز

وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

نشر چاپار

سرشناسه	:	استوارت، دیوید (David Stuart, David Patrick)
عنوان و نام پدیدآور	:	سنجه‌های وبی: برای متخصصان کتابداری و اطلاع‌رسانی/دیوید استوارت ؛ مترجم مریم خسروی با همکاری فاطمه بهمنی ؛ ویراستار حمید کشاورز.
مشخصات نشر	:	تهران: پژوهشگاه علوم و فناوری اطلاعات ایران: چاپار، ۱۳۹۶.
مشخصات ظاهری	:	۲۰۰ ص.
شابک	:	978-600-98635-0-1
وضعیت فهرست نویسی	:	فیپا
یادداشت	:	عنوان اصلی: Web metrics for library and information profesionals, 2014.
یادداشت	:	واژه‌نامه.
یادداشت	:	کتابنامه: ص. ۲۵۹.
یادداشت	:	نمایه.
موضوع	:	وب‌سنجی
موضوع	:	Webometrics
موضوع	:	کتاب‌سنجی
موضوع	:	Bibliometrics
موضوع	:	وب -- تحقیق
موضوع	:	World Wide Web -- Research
موضوع	:	شبکه‌های اجتماعی پیوسته
موضوع	:	Online social networks
شناسه افزوده	:	خسروی، مریم، ۱۳۳۸ -، مترجم
شناسه افزوده	:	بهمنی، فاطمه، ۱۳۶۹ -، مترجم
شناسه افزوده	:	کشاورز، حمید، ۱۳۵۸ -، ویراستار
شناسه افزوده	:	پژوهشگاه علوم و فناوری اطلاعات ایران
رده بندی کنگره	:	۱۳۹۶ ۹س۵الف/Z۶۶۹/۸
رده بندی دیویی	:	۰۲۰/۷۲۷
شماره کتابشناسی ملی	:	۴۹۷۲۸۶۲

این کتاب از سلسله انتشارات پژوهشگاه علوم و فنّاوری اطلاعات ایران است که با همکاری انتشارات چاپار منتشر شده است.

سنجه‌های وبی

برای متخصصان کتابداری و اطلاع‌رسانی

دیوید استوارت

مترجم: **مریم خسروی**

با همکاری: **فاطمه بهمنی**

ویراستار: **حمید کشاورز**

صفحه‌آرا و طراح جلد: رحیم کبیرصابر

نوبت چاپ: اول ۱۳۹۶| شمارگان: ۵۰۰ نسخه

تهران- خیابان ولی‌عصر (عج) - بالاتر از زرتشت

کوچه نوربخش- پلاک ۴۰ - نشر چاپار

تلفن پخش: ۸۸۸۹۹۶۸۰ - ۰۹۱۲۱۹۶۰۲۱۴

بها: ۲۵۰.۰۰۰ ریال

فهرست مطالب

فهرست اشکال

فهرست جداول

۱

مقدمه

همه آنچه قابل شمارش است ارزشمند شمرده نمی شود، و همه آنچه ارزشمند است قابل شمارش نیست.

ویلیام بروس کامرون[1] (۱۹۵۷)

همه سازمان ها، مشاغل و خدمات بیش از پیش در معرض سنجه‌ها و شاخص‌های عملکردی قرار می‌گیرند. جداول لیگ ها تنها برای برنده و بازنده‌های گروه های ورزشی استفاده نمی‌شوند، بلکه برای مقایسه سازمان‌های گوناگون از بیمارستان‌ها و مدارس گرفته تا مؤسسات آموزشی و کشورها به کار گرفته می‌شوند. گسترش و کاربرد وب سبب گسترش و کاربرد سنجه‌ها و شاخص‌های جدید شده است. وب سبب تسهیل به اشتراک‌گذاری سنجه‌ها و داده‌های مرسومی مانند داده‌های عمومی گوگل[2] می‌شود که در حال حاضر زمینه کاربرپسندی را برای طیف وسیعی از داده های عمومی فراهم می کند به گونه‌ای که امکان مقایسه‌ای ساده بین کشورها برمبنای ویژگی‌هایی مانند تعداد فروش روزانه روزنامه‌ها تا میزان مشارکت در زمینه انرژی تجدیدپذیر برای تأمین انرژی یک کشور را امکان‌پذیر می‌سازد. وب امکان عرضه داده های جدیدی را میسّر می‌سازد مانند خدمات وب ۲/۰ که امکان رتبه‌بندی

1. William Bruce Cameron 2. Google Public Data (www.google.com/publidata)

سیاست‌مدارها را بر اساس جدّیت[1] آنها و اساتید را بر اساس کارآمدی آنها فراهم می‌کند. از همه مهم‌تر اینکه وب رسانه‌های جدیدی را نیز برای تحلیل ارائه می‌کند مانند همان روشی که رسانه‌های سنتی کتاب و مجله مبنایی برای سنجه‌های جدیدی از قبیل تعداد کتاب‌های منتشر شده یک کشور یا تعداد استنادهای دریافت شده یک مجله را فراهم می‌کردند. بدین‌ترتیب، رسانه‌های جدیدی چون صفحه‌های وبی و وب‌نوشت‌ها[2] پایه‌هایی را برای سنجه‌های جدید فراهم می‌کنند. این سنجه‌های وبی می‌توانند بسیار غنی‌تر از سنجه‌هایی مربوط به رسانه‌های قدیمی باشند زیرا داده‌های متنوع‌تری را را در سطوح بسیار قطعه‌ای[3] گردآوری می‌کنند.

این کتاب نشان می‌دهد که چگونه شماری از سنجه‌های وبی جدید می‌توانند افزوده ارزشمندی بر مجموعه مهارت‌های متخصصان کتابداری و اطلاع‌رسانی باشند. سنجه‌های وبی می‌توانند امکاناتی را برای بهبود خدمات برخطی که کتابداران برای کارگزاران خود عرضه می‌کنند فراهم کرده و تأثیر خدمات آنها را برای مدیران و سیاست‌گذاران به نمایش بگذارند. از سنجه‌های وبی می‌توان برای شناسایی مرتبط‌ترین منابع یک حوزه خاص و اهمیت خدمات برخط آنها استفاده کرد. در حال حاضر حضور برخط متخصصان کتابداری و اطلاع‌رسانی بسیار گسترده است. کتابخانه‌های بسیاری دارای حساب کاربری فیسبوک[4] و توییتر[5] هستند. برخی از آنها میزبان وب‌نوشت‌ها و ویکی‌ها[6] هستند و تعداد اندکی هم از خدمات رسانه‌های اجتماعی جدید مانند پینترست[7] یا گوگل پلاس[8] استفاده می‌کنند. این کتاب روش‌هایی را برای ایجاد و تحلیل سنجه‌ها ارائه کرده و مشکلات موجود را مشخص می‌کند. این فصل مقدمه‌ای درباره موضوع کتاب و نیز نگاهی کلی بر ادامه کتاب ارائه می‌کند.

در سرتاسر کتاب اصطلاح «کتابدار[9]» برای کوتاه‌سازی اصطلاح طولانی‌تر «متخصص کتابداری و اطلاع‌رسانی» بکار رفته است. از آنجا که مخاطبان مورد نظر این کتاب بسیاری از متخصصان دانش و اطلاعات هستند که نه در کتابخانه‌های

1. hotness 2. blogs 3. granularity 4. Facebook
5. Twitter 6. Wiki 7. Pinterest 8. Google Plus 9. librarian

سنتی کار می کنند و نه خود را در شمار کتابداران می‌دانند، انتخاب یک اصطلاح کوتاه‌تر برای واضح‌تر نشان دادن ضروری بود. همانگونه که در بحث‌های اخیر پیرامون تغییر علامت تجاری مؤسسه اعطای امتیاز متخصصان کتابداری و اطلاع‌رسانی[1] دیده می‌شود، هیچ اصطلاح خاصی نیست که با آن بتوان کل حرفه را شناسایی کرد. از این رو کتاب از رایج‌ترین اصطلاح استفاده کرده است.

سنجه‌ها

از آنجا که «سنجه‌ها» پایه و اساس کتاب حاضر را تشکیل می‌دهند، ضروری است ابتدا با درک روشنی از معنای آن شروع کنیم. اصطلاح «سنجه» در سراسر این کتاب دلالت بر یک *استاندارد قابل سنجش اندازه‌گیری*[2]، یا بنا به تعریف واژه‌نامه انگلیسی آکسفورد[3] «نظام یا استاندارد اندازه‌گیری؛ یک یا چند معیار بیان شده در قالبی قابل اندازه‌گیری» (اُ اِد[4] ۲۰۰۱) دارد. یک استاندارد قابل ارزیابی نه تنها می تواند به عدد مشخصی تقلیل یابد، بلکه می‌تواند در موقعیت های مختلف برای مقایسه بین موارد بررسی مشابه بکار رود.

اندازه‌گیری‌های کمّی شالوده برخی از پیشینه‌های اولیه انسان را تشکیل می دهند: تراش چوب خط روی استخوان ها برای ثبت کارهای انجام شده، و نیز تملک گله گاو در جوامع پیش از ظهور خط (یو[5]، ۲۰۱۰). در چنین مواقعی که در آن تراشه‌ها اساس کار بودند احتمالاً استانداردهای بسیار ساده‌ای مطرح بودند. برای مثال، هنگامی که کشاورز تعداد گله گاو خود را مشخص می کرد، اغلب مردم می‌دانستند که منظورش موش صحرایی نیست بلکه تعدای گاو است. با اینهمه، هرچه جامعه پیچیده‌تر شده و تجارت به مسیرهای طولانی بسط پیدا می‌کند، ناگریز باید استاندارها مشخص‌تر شوند. تخصصی‌تر کردن بدین معنی است که دیگر نمی توان هنجارهای یکسانی را بین جوامع مختلف پذیرفت. پایان چنین شرایطی را می توان در

1. Chartered Institute of Library and Information Professionals
2. quantifiable standard of measurement 3. Oxford English Dictionary 4. Oed 5. Yeo

اواخر قرن هجدهم در فرانسه مشاهده کرد که در سراسر کشور به اندازه ربع یک میلیون شاخص وزن و ارزیابی مورد استفاده بودند (راسل[1]، ۲۰۰۵). این قبیل گوناگونی وزن و معیار ها پیامدهای روشنی را برای جریان آزاد تجارت در بردارند، به همان شکلی که اکنون استانداردی برای سنجش وب وجود ندارد. اگر شرکتی تمایل داشته باشد بر روی وبگاه شرکت دیگری تبلیغات کند ولی برای بررسی تعداد بازدید های صفحه وبگاه هیچ روش استانداردی نداشته باشد گفتگوهای پیچیده ای مورد نیاز خواهد بود. به همین ترتیب، چنانچه یک سازمان بزرگ چند ملیتی بخواهد با خدمات یک شرکت بازاریابی رسانه های اجتماعی تعامل داشته باشد، اما هیچ استانداردی برای اندازه‌گیری تأثیر در کار نباشد، تعیین اینکه آیا شرکت بازاریابی اهداف آن سازمان را تحقق بخشیده یا بدون تحقق اهداف دستمزد پرداخت شده مشکل خواهد بود. از این رو توسعه استانداردهای قابل قبول به نفع افراد بوده و این استاندارها به صورت برخط در حال ظهورند.

استاندارها یا از طریق اتفاق نظر، یا اعمال نفوذ یک صاحب نظر، و یا با ترکیب این دو روش ظهور می‌کنند (راسل، ۲۰۰۵). در جایی که گوناگونی استانداردها بر اقتصاد یا بر امنیت افراد تأثیر چشمگیری داشته باشند، استانداردسازی می تواند توسط دولت اعمال شود. در بیش‌تر موارد استاندارسازی در حوزه‌های بسیاری به اتفاق نظر محول می‌شود تا از اجتماع افراد ذینفع پدیدار شود. این موضوع مخصوصاً در محیط هایی که تغییرهای سریع در کار با فناوری های جدید موجب اختلال سیستم‌های فعلی می‌شوند می‌تواند دشوار باشد. چنین گسستی از فناوری به جزء لاینفکی از کار کتابداران تبدیل شده است.

جامعه کتابداران در تعامل با سنجه ها دارای تاریخچه طولانی است. خود حوزه کتاب‌سنجی توسط آلن پریچارد[2] در سال‌های ۱۹۶۰ به عنوان «کاربرد ریاضیات و روش‌های آماری برای کتاب‌ها و سایر روش‌های ارتباطی» تعریف شده است

1. Russell 2. Alan Pritchard

(۱۹۶۹، ۳۴۹). شاید ریشه‌های آن به زمان‌های بسیار پیش‌تر برگردد. درحالی که کاربرد کتاب‌سنجی توسط گارفیلد[1] و پرایس[2] در دهه ۱۹۵۰ همه‌گیر شد، گودین[3] (۲۰۰۶) ریشه‌های کتاب‌سنجی را در روانشناس‌های اوایل قرن بیستم دنبال می‌کند، و شاپیرو[4] (۱۹۹۲) آغاز آن را تا کتاب سنجی حقوقی قرن نوزدهم به عقب برده است. برودوس[5] (۱۹۸۷) حتی پا را از این نیز فراتر گذاشته و پیش درآمد کتاب‌سنجی را شمارش ساده کتاب‌ها یا مواردی از این قبیل که می‌توان در قرن سوم قبل از میلاد مسیح ردگیری کرد و شناسایی ۴۹۰/۰۰۰ کتیبه در کتابخانه الکساندریا[6] دانسته است. با اینهمه، حتی با تاریخچه‌ای بیش از ۲۰۰۰ سال قبل، پاسخگویی به سؤالات کتاب‌سنجی مسئله ساده‌ای نیست، خواه مربوط به تعداد منابع موجود در یک کتابخانه باشد یا حتی به صورت خاص‌تر تعداد کتاب‌های یک کتابخانه باشد.

این پرسش که «چه تعداد کتاب در یک کتابخانه نگهداری می‌شود؟» سریعاً مسائلی را در خصوص معنای کتاب و اینکه در عصر کتاب‌های الکترونیکی، نگهداری کتاب برای یک کتابخانه به چه معناست را به ذهن متبادر می‌سازد. با وجودی که همه ما ایده مبهمی از معنای کتاب داریم که برای بیش تر تعاملات روزانه ما کافی است، اما برای استانداردسازی نیاز به تعریفی صریح و واضح وجود دارد لیکن برای «کتاب» هیچ تعریفی واحدی که مورد وفاق جهانی باشد وجود ندارد. سازمان یونسکو[7] (۱۹۸۵) با هدف گردآوری آمارهایی درباره تولید و توزیع انتشارات چاپی، کتاب را اینگونه تعریف می‌کند: «انتشارات غیر ادواری با حداقل ۴۹ صفحه به‌جز صفحه‌های جلد، که در کشوری منتشر شده و دسترسی آن برای عموم آزاد است». اما خدمات پستی ایالات متحده[8] (۲۰۱۲) کتاب را به عنوان «هر چیزی که بیش از ۲۴ صفحه داشته باشد» تعریف می‌کند، بدون اینکه به قابلیت دسترسی عموم به آن اشاره‌ای داشته باشد. انتخاب یکی از این دو تعریف می‌تواند موجب شود با

1. Garfield 2. Price 3. Godin 4. Shapiro
5. Broadus 6. Alexandria 7. UNESCO 8. US Postal Service

یک حرکت قلم، حجم موجودی یک کتابخانه تقلیل یا افزایش یابد. گرچه می‌توان به راحتی به تصمیم مقامات رسمی که استانداردهای محدودتری را برای کتاب برمی‌گزینند انتقاد داشت، و بدین‌ترتیب اندازه کتابخانه‌ها را بدون هیچ هزینه اضافی افزایش داد، اما می‌توان به راحتی وضعیتی را تصور کرد که تعریفی سخت‌تر هم مناسب نیست. برای مثال، در کتابخانه های کودکان که بسیاری از آنچه که ما به عنوان «کتاب» در نظر می‌گیریم کم‌تر از ۴۹ صفحه هستند، ظاهراً تعریف اداره خدمات پستی ایالات متحده از کتاب تعریف مناسب تری است تا اینکه بسیاری از ما چنین کتاب‌هایی را به پیروی از یونسکو جزوه[1] به حساب می‌آوریم.

پیدایش کتاب‌های الکترونیکی در سال‌های اخیر موجب چالش بزرگ‌تری در تعیین تعداد کتاب های یک کتابخانه شده است. از آنجایی که کتابخانه ها بیش از پیش مشترک انتشارات الکترونیک می‌شوند، تمرکز تنها بر انتشارات چاپی مسلماً بخشی از اوضاع است. اما اگر در صدد مقایسه کتاب‌های دو مؤسسه باشیم چگونه می‌توانیم کتاب‌های الکترونیکی را بحساب آوریم؟ آیا اشتراک ۱۰/۰۰۰ بسته نرم‌افزاری کتاب الکترونیکی مهم‌تر از اشتراک ۱/۰۰۰ عنوانی است که برگزیده شده‌اند؟ آیا با یک اَبَرپیوند[2] به پروژه گوتنبرگ[3] از وبگاه یک کتابخانه می‌توان ادعای ۳۹/۰۰۰ کتاب الکترونیکی اضافی را داشت؟ اگر مجموعه پروژه گوتنبرگ در فهرست کتابخانه ثبت شده باشد و کتابداران از دانش تخصصی خود برای ایجاد ارزش افزوده استفاده کرده باشند چطور؟ چنانچه چند کاربر بتوانند به طور هم‌زمان به یک کتاب الکترونیکی دسترسی داشته باشند آیا می‌توان آن‌ها را به عنوان چند نسخه شمارش کرد؟ پاسخ اصلاً ساده نیست، هرچند استانداردهایی برای رسانه های جدید و قدیمی در حال تدوین هستند. کانتر[4] یک اقدام بین المللی برای تدوین استانداردهایی در تبادل اطلاعات مورد استفاده بین ناشران و کتابخانه‌ها در خصوص اشتراک مجلات، پایگاه داده ها و کتاب است، در حالیکه انجمن تحلیل وب (پیشگام در انجمن تحلیل

1. pamphlet
2. hyperlink
3. Project Gutenberg (www.gutenberg.org)
4. COUNTER (www.projectcounter.org)

دیجیتال[1])، به منظور تسهیل امر ارتباطات و تحلیل بهتر کاربرد وب، تعاریفی را برای طیف گسترده‌ای از مفاهیم مانند بازدیدهای صفحه [2] و ارجاع[3] منتشر کرده است.

وجود استانداردها و تعاریف به این معنی نیست که سنجه های بسیار مناسبی برای یک موقعیت خاص شناسایی شده‌اند. روش های متعددی برای سنجش محتوای وب وجود دارد. دیوید برکویتز[4] (۲۰۰۹) برای ارزیابی رسانه های اجتماعی در وب‌نوشت بازاریابی خود صد روش را فهرست می کند. او سنجه های مستقیم بسیاری را که شاید کاملاً بدیهی باشند مانند دوستان[5] (فیسبوک)، دنبال‌کننده‌ها[6] (توییتر) و نمایش صفحات (وب‌نوشت‌ها)، و سنجه‌هایی که از صراحت و روشنی کم‌تری برخوردارند مانند تعداد مشتری ها و درخواست‌های کاری ذکر کرده است. یک سنجه مناسب در درجه نخست بستگی به هدف سنجه دارد، اما در هنگام مقایسه بین افراد یا سازمان‌ها، گزینش یک سنجه خاص قطعاً بحث‌انگیز است زیرا ممکن است آن سنجه به نفع یک گروه سوگیری داشته باشد.

شاخص‌ها

سنجه‌ها نه برای خودشان، بلکه برای هدفی در نظر گرفته می‌شوند. برای مثال، وقتی‌که کتابخانه‌ای پس از شمارش کتاب‌هایش اظهار می‌دارد که مجموعه‌ای با یک میلیون کتاب را در اختیار دارد، به عنوان شاخصی از توانایی کتابخانه در مواجه با نیازهای کاربرانش تلقی می‌شود. استفاده از چنین سنجه‌ای به عنوان یک شاخص در صورتی می‌تواند اشاره شود که مشخص و مرسوم باشد. برای مثال، بعید است که یک کتابخانه اعلام اندازه مجموعه خود به عنوان شاخصی از توانایی خود در رویارویی با نیازهای کاربرانش را ضروری بداند. با اینهمه، هنگامی که کاربرد یک سنجه به عنوان شاخص برای چیزی بزرگ‌تر کاملاً مشخص و یا مورد پذیرش همگان نباشد می‌تواند مشکل‌آفرین باشد، بویژه وقتی که امکان پیامدهای چشمگیری وجود داشته باشد.

1. Web Analytics Association (www.digitalanalyticsassociation.org)
2. Page view 3. referral 4. David Berkowitz 5. friends 6. followers

روش های کتاب سنجی تثبیت شده در عرصه علوم کتابداری و اطلاع‌رسانی بطور پیوسته در درون جامعه دانشگاهی وسیع تر پذیرفته می‌شوند، تا شاخص های تعالی فردی و مؤسسه پژوهشی عرضه شود، هرچند که چندان مورد استقبال مخاطبان چنین بررسی‌هایی قرار نمی‌گیرند. متعاقب اعلام این موضوع که کتاب‌سنجی به چارچوب ارتقای پژوهش ۲۰۱۳[1] کمک می‌کند (روشی که کیفیت پژوهش را در مؤسسه های آموزش عالی انگلستان ارزیابی کرده و شیوه اختصاص میلیون‌ها پوند سرمایه‌گذاری را نشان می‌دهد)، کاربرد کتاب‌سنجی موجی از انتقادهای منفی را به دنبال داشت: «سنجه ها ادعای تنوع را از بین می‌برند» (لیپست[2]، ۲۰۰۶)؛ «پیروزی نبوغ را تحت الشعاع قرار دهد» (لاورنس[3]، ۲۰۰۷)؛ «گزارش: کتاب‌سنجی‌ها می‌توانند ارزیابی پژوهش را تحریف کنند» (لیپست، ۲۰۰۷).

دل نگرانی‌هایی که برای کاربرد کتاب‌سنجی وجود دارند حاکی از کاربرد گسترده‌تر آنهاست. سه چالش اصلی در پذیرش گسترده کتاب‌سنجی برای شاخص های برتری پژوهش همواره وجود دارند: نبود توافق در اینکه ارتقای پژوهش می تواند کمّی باشد؛ نگرانی از ابزارهای موجود؛ و نگرانی تأثیر شاخص‌ها بر فرآیند نشر و پژوهش.

چنین نگرانی‌هایی تنها به کاربرد کتاب‌سنجی محدود نیستند، بلکه در بکارگیری سایر سنجه های مرتبط نیز به همان اندازه وجود دارند، بویژه وقتی که منتهی به برتری سنجه در برابر معنای حقیقی شاخص شود. برای مثال بانرجی و دوفلو[4] (۲۰۱۲) مدرسه ای در کلکته[5] که سابقه قبولی سالانه خوبی داشتند را در نظر گرفتند. متأسفانه از آنجا که این سابقه از طریق اتخاذ سیاست اخراج سالانه دانش آموزهای تنبل کلاس بدست آمده بود؛ تمایل به یک نرخ قبولی خوب بر تمایل به آموزش خوب که هدف طراحی سنجه بود پیشی گرفته بود. این امر به منزله بی استفاده بودن چنین شاخصه‌هایی بی استفاده نیست بلکه به این معناست که باید با آن ها با احتیاط

1. 2013 Research Excellence Framework
2. Lipsett
3. Lawrence
4. Banerjee & Duflo
5. Calcutta

رفتار شود، مخصوصاً در سنجه‌های وبی که در آن‌ها ایجاد محتوا ارزان بوده و بسیاری از اطلاعات موقتی هستند.

سنجه‌های وبی و قانون رانگاناتان برای علوم کتابداری

اصطلاح «سنجه‌های وبی» در سرتاسر این کتاب دلالت بر ارزیابی کمّی ایجاد و کاربرد محتوای وب دارد. تعیین تأثیر بیش از حد وب بر زندگی افراد مشکل است. وب علاوه بر اینکه جایی است که مردم می‌توانند از آن اطلاعات بدست آورند، جایی است که مرتباً در آن با یکدیگر تعامل چشمگیر داشته و محتوای خود را ایجاد می کنند. در طول دو دهه گذشته وب موجب تغییر نشر شکل‌های قدیمی رسانه‌ای شده و سبک‌های جدیدی از رسانه های دیجیتالی را معرفی کرده است. صفحه‌های خانگی[1] و وبگاه‌ها از طریق وب نوشت‌ها و ویکی‌ها و سایت‌های شبکه اجتماعی عظیمی که میلیون‌ها کاربر را جذب خود کرده‌اند بهم پیوند خورده‌اند. درحال حاضر توییتر، فیسبوک، لینکداین[2] و فلیکر[3] بسترهای اصلی بسیاری از افراد و سازمان ها بوده و افرادی که چنین بسترهایی را نادیده می‌گیرند، به طور بالقوه فرصت مشارکت خود را با تعداد زیادی از سهامداران واقعی یا احتمالی از دست می‌دهند. انواع مطالب منتشر شده نیز گسترده شده است از مدرک تا داده ها و از متن تا انواع غنی رسانه. تمام این رسانه‌ها و بسترهای جدید شیوه‌های جدیدی را برای کتابداران فراهم می‌کنند تا اطلاعات را به اشتراک گذاشته، سنجه‌ها و شاخص‌های جدیدی را جهت ارزیابی ارائه کنند. با چنین حجمی از شاخص‌ها، کتابداران باید فلسفه بنیادی علوم کتابداری و اطلاع‌رسانی و نیز نقش آن در نظام زیستی[4] اطلاعات را همواره در ذهن داشته باشند. کتابدارانی که به شایعه پراکنی درباره یک شخص سرشناس دامن بزنند شاید بتوانند تأثیر برخط خود را به سرعت افزایش دهند، هرچند چنین رفتاری به فلسفه اساسی حرفه کتابداری ارتباطی پیدا نمی‌کند.

1. Home pages 2. LinkedIn 3. Flickr 4. ecosystem

اهداف حرفه کتابداری و اطلاع‌رسانی الزاماً شبیه اهداف سایر کسب و کارها نبوده و نوع سنجه های آن باید متناسب با آن اهداف باشد. جیم استرن[1] (۲۰۱۰) در کتاب *سنجه‌های رسانه‌های اجتماعی*[2] بر اهمیت شناسایی سنجه‌هایی تأکید می‌کند که در راستای سه هدف بزرگ این حرفه هستند: افزایش درآمد، کاهش هزینه ها و ارتقای رضایت مشتری. هر یک از این سه هدف جایگاه خاصی در کتابخانه دارند. کتابخانه باید راه‌هایی را برای افزایش درآمد و کاهش هزینه های خود پیدا کند، اما عامل اصلی رضایت مشتری است، و برای تحقق آن لازم است به فلسفه علوم کتابداری که در پنج قانون علوم کتابداری رانگاناتان (۱۹۳۱) به صراحت بیان شده توجه کنیم:

۱. کتاب‌ها برای استفاده‌اند.
۲. هر خواننده کتاب خود را دارد.
۳. هر کتاب خواننده خود را دارد.
۴. در زمان خوانندگان صرفه‌جویی شود.
۵. کتابخانه یک نظام پویاست.

این قوانین پیوسته بازتفسیر می‌شوند تا فناوری های جدید و انواع محتوای جدید، چه انواع جدید رسانه‌های (سیمپسون[3]، ۲۰۰۸) باشد و چه داده‌های اصلی (استوارت[4]، ۲۰۱۱)، را دربر بگیرد. در نظر داشتن این قوانین هنگام شناسایی سنجه‌های وبی مناسب ضروری است. سنجه های وبی می توانند به ما کمک کنند تا موضوعاتی مثل اینکه آیا کتاب ها مورد استفاده هستند یا خیر، خوانندگان دسترسی به اطلاعات مورد نیاز دارند یا خیر، اطلاعات به اشخاصی که به آن نیاز دارند عرضه می شوند یا خیر، و اینکه آیا ما در وقت خوانندگان صرفه جویی می‌کنیم یا خیرو اینکه آیا کتابخانه یک نظام پویاست یا خیر را تعیین کنیم.

1. Jim Sterne 2. *Social Media Metrics* 3. Simpson 4. Stuart

سنجه‌های وبی برای متخصصان کتابداری و اطلاع‌رسانی

درحالی که رانگاناتان به ما در شناسایی اهداف کتابداران کمک می‌کند، سنجه‌های وبی نیز می‌توانند به شیوه‌هایی ما را در رسیدن به این اهداف یاری دهند. بِهن[1] (۲۰۰۳) هشت هدف را برای سنجش عملکرد مدیران سازمان‌های عمومی شناسایی کرد: ارزیابی[2]، کنترل[3]، بودجه‌بندی[4]، انگیزه بخشی[5]، ارتقا[6]، تجلیل[7]، آموختن[8] و اصلاح[9]. هر یک از این اهداف می توانند موجب تشویق کتابداران شده تا یک سنجه وبی را برای بیش از یک هدف توسعه دهند.

ارزیابی معمولا برای سنجش عملکرد و ارزیابی پژوهش که کتاب‌سنجی‌ها مبتنی بر آن است انجام می‌شود (برای مثال موئد[10]، ۲۰۰۵). با این وجود، ارزش ارزیابانه سنجه‌های وبی ضرورتاً به اندازه اهمیت شاخص‌های کتاب‌سنجی روشن نیست. درحالی که برای برخی سازمان‌ها شاید بازدیدکنندگان یک فروشگاه برخط دارای الویت باشد، اما برای سایرین می‌تواند از اولویت کم‌تری برخوردار باشد. برای مثال، موفقیت یک گروه پژوهشی به اندازه حضور آن در وب در قالب انتشارات پژوهشی ضروری نیست، با این حال باز هم برخی مطالعات نشان داده اند که تعداد پیوند به یک وبگاه با ارتقای پژوهشی مؤسسه و موفقیت وبگاه گره خورده است (وُگان و یانگ[11]، ۲۰۱۲). تعداد دنبال‌کننده‌های توییتر یا لایک[12] صفحه فیسبوک یک سازمان نیز ممکن است به عنوان موفقیت یک برند در وب اجتماعی بنظر آید.

کنترل بر اطمینان از رفتار مناسب افراد دلالت دارد. درحالی که ممکن است یک کتابخانه برای ارزیابی تأثیر خدمات خود علاوه بر این که از سنجه تعداد دنبال‌کننده‌های توییتری خود استفاده کند، باید اطمینان پیدا کند که با آنها به بهترین شکل تعامل دارد. شاید کتابخانه برای تعداد توییت هایی[13] که در یک روز یا در هفته خاصی می‌فرستد حد پایین یا بالایی را برای عرضه خدمات قائل شود.

---

1. Behn
2. Evaluate
3. Control
4. Budge
5. Motivate
6. Promote
7. Celebrate
8. Learn
9. Improve
10. Moed
11. Vaughan and Yang
12. likes
13. tweets

بودجه‌بندی بر تخصیص مناسب منابع دلالت می‌کند. درحالی که دسترسی کتابداران معمولاً بسیار محدود است اما آنها می‌توانند در وب از راه های زیادی برای برقراری ارتباط با کاربران خود استفاده کنند. سنجه‌ها می‌توانند به کتابداران کمک کنند تا کارآمدترین فناوری وب را برای اهداف خاص خود مشخص نمایند. احتمال دارد که راه‌اندازی خدمات برخط جدید زمان بر باشند، اما در صورتی که پس از شش ماه نشانه های پیشرفت مشاهده نشد، می تواند به معنای نامناسب بودن این خدمات باشد.

انگیزه‌بخشی بر تشویق کاربران در نیل به اهداف دلالت می‌کند. وب فرصت‌های زیادی را فراهم می کند، اما سطح آن می تواند موجب اضطراب شود. هنگامی که در خصوص افراد مشهور با میلیون‌ها دنبال‌کننده توییتری گزارش‌هایی منتشر می‌شود یا ویدیویی که در یوتیوب ویروسی شده است و میلیارد ها بار دیده شده است، شاید آنچه که رسانه های اجتماعی یک کتابخانه بازنمون می کند بسیار جزئی به نظر برسد. بنابراین، جهت ایجاد انگیزه ضروری است اهداف تحقق‌پذیر ویژه‌ای را در ذهن داشت و آن را با سازمان‌های مشابه یا تلاش‌های مشابه سازمان‌های همانند مقایسه کرد. شاید وب‌نوشت یک کتابخانه دنبال این باشد که صدها خواننده یا شاخص مشارکت چشمگیرتری مانند تعداد اظهار نظرهای[1] گذاشته شده بدست آورد.

ارتقا بر این نکته دلالت می‌کند که عموم مردم یا مقامات رده اول سازمان متقاعد شوند که به خوبی کار می کنند. بودجه کتابخانه‌ها دائماً تحت بررسی بوده و کتابداران باید اهمیت خود را نشان بدهند، هرچند شاید در زمانی که در یک مؤسسه عمومی از فناوری جدیدی استفاده شده اهمیت آن هنوز برای همگان مشخص نباشد لذا این کار باید با احتیاط انجام شود. علیرغم اینکه یک وب‌نوشت می‌تواند در چند دقیقه ایجاد شود، ایجاد یک حضور برخط موفق چیزی نیست که در یک شب بدست آید بلکه شاید ماه‌ها یا حتی سال‌ها طول بکشد. حتی در صورتی که کتابخانه‌ای حضور

1. comments

موفقی در یک خدمت برخط ایجاد کرده باشد، هیچ تضمینی نیست که این کار چیزی در حد اتلاف وقت باشد.

تجلیل بر نمایان ساختن دستاوردهای سازمانی دلالت دارد. در هنگام تحقق اهداف مشترک، فرصتی برای گروه ها فراهم می‌شود تا حول این دستاورد در کنار هم باشند. این هدف می تواند شامل هزارمین دنبال کننده توییتر کتابخانه، پانصدمین بارگیری[1] پادکست[2]، یا یک میلیون بازدیدکننده از وب‌سایت باشد.

یادگیری بر درک تأثیر مشارکت های کتابداران دلالت می‌کند. در دنیایی که فناوری‌های وبی به سرعت در حال تغییرند تعیین تأثیر یک فناوری خاص بسیار ضروری است. تنها از طریق سنجش است که مشکل‌ها شناسایی شده و با آن‌ها برخورد می‌شود. آیا میزان تعامل حساب توییتری یک سازمان هنگامی که توسط یک کاربر کنترل می شود بیش‌تر از زمانی است که توسط کاربر دیگری کنترل می‌شود؟ آیا طراحی مجدد وبگاه، که بسیار مورد علاقه مدیرکل است موجب تغییر تعداد بازدیدهای وب‌سایت شده است؟

اصلاح بر تلاش در عرضه خدمات بهتر دلالت دارد. تنها دانستن این که محتوایی خاص موجب واکنش مطلوب تری می‌شود کافی نیست بلکه کتابداران باید با استفاده از این اطلاعات به اصلاح خدماتی بپردازند که سازمان آنها عرضه می‌کند. اگر سنجه‌ها نشان دهند که طراحی جدید وبگاه مفدور نیست سازمان بجای بی توجهی و ادامه کار با آن وبگاه، باید از آن تجربه درس بگیرد.

هشت هدف بهن (۲۰۰۳) بر سنجش عملکرد داخلی معطوف شده و به سازمان کمک می‌کنند تا از کارکرد و دستاوردهای خودش درک بهتری بدست آورد. سنجه های وب همچنین فرصتی را برای کتابداران فراهم می کنند تا سنجه ها را به دو شیوه فراتر از سازمان اعمال کنند: پاک‌سازی[3]؛ پژوهش[4].

پاک‌سازی به معنای کاربرد سنجه‌های وبی در مواجهه با مسئله اضافه بار اطلاعاتی است. اضافه بار اطلاعاتی مسئله جدیدی نبوده و در واقع ماجرای پیشرفت

1. download 2. podcast 3. Filter 4. research

علمی مرتباً توسط ابزارهای جدید برای مقابله با این مشکل با موانعی همراه بوده است. پیدایش مجله های علمی در قرن هفدهم موجب شد که مسئله اجبار دانشمندان برای به اشتراک گذاری چندباره نتایج مشابه با همکاران مختلف حل شود. هم راستا با افزایش تعداد مجله ها، تعداد ابزارهایی به منظور کمک به پژوهشگران جهت مدیریت آنها نیز افزایش یافت. در قرن هجدهم هم ترازخوانی[1] (کرونیک[2]، ۱۹۹۰) و مجله های چکیده اختصاصی[3] (اسکولنیک[4]، ۱۹۷۹) پدیدار شدند، در حالی که رایانه ها نمایه های استنادی و نمایه‌سازی متن کامل را در قرن بیستم در مقیاس بالا فراهم نمودند. وب امکان انتشار میلیاردها مدرک را میسّر کرده و روش های جدیدی برای کنترل انفجار اطلاعات را ضروری ساخته است مگر اینکه ما بخواهیم «مجلد مهملات غیرمسئولانه[5]» را با هر زحمتی که است بخوانیم، اصطلاحی که زیمان[6] (۱۹۶۹) بسیار پیش تر از اختراع وب بیان کرده بود. به همان روشی که ضریب تأثیر مجله به شناسایی مجله های هسته کمک کرد، رتبه‌صفحه[7] نیز به رتبه‌بندی صفحه‌ها در وب کمک کرد و انتظار می‌رود الگوریتم های جدید به پژوهشگران در کار پاک‌سازی مجموعه متنوع مطالب برخط رو به افزایش کمک کنند.

انجام پژوهش به معنای اعمال روش های کمّی سنجه های وبی برای اهداف پژوهشی است. همانگونه که در جای دیگر نیز گفته شده است (برای مثال استوارت، ۲۰۱۱) هرچند که کتابداران تا آینده نزدیک درگیر آماده‌سازی اسناد قدیمی (کتاب‌ها، مجله‌ها و مقاله‌ها) هستند، اما با توجه به اینکه نقش‌های سنتی کتابداران خودکار[8] شده و وب از وب اسناد به وب داده ها تغییر پیدا کرده است، تغییر نقش کتابدارها نیز حائز اهمیت است. این امر تنها به معنای دسترسی کاربران به انواع قدیمی اسناد نیست، بلکه تسهیل دسترسی کاربر به حجم عظیم داده های است که همه روزه به طرز روزافزونی به صورت برخط در دسترس قرار می‌گیرند. برخی کتابخانه ها از قدیم یک خدمت پژوهشی برای داده‌های کتاب‌سنجی ارائه کرده‌اند از تحلیل

1. peer review
2. Kronick
3. dedicated abstract journals
4. Skolinik
5. tomes of irresponsible nonsence
6. Ziman
7. PageRank
8. automated

استنادی گرفته تا نگاشت پژوهشی در یک حوزه خاص. سنجه‌های وبی می‌توانند از چنین مطالعات کتاب‌سنجی قدیمی اطلاعات بیش‌تری ارائه کرده و این امکان را فراهم کنند که در خصوص گستره وسیعی از پرسش‌های پژوهشی اطلاعاتی به دست آید، از قبیل افکار همگانی بر روی موضوعاتی مانند محصولات کشاورزی اصلاح شده ژنتیکی و آخرین نسخه بازی‌های رایانه‌ای (تلوال[1]، ۲۰۰۹) تا پیش‌بینی برنده نمایش‌های مشاهیر (تنسر[2]، ۲۰۰۹) و شیوع بیماری‌ها (ایسنباچ[3]، ۲۰۰۶).

هدف این کتاب

هدف این کتاب نمایش نقش سنجه های وبی در کار کتابداران است. تمرکز اولیه بر روی ابزارهایی است که بدون محدودیت قابل استفاده بوده و حداقل به صورت رایگان دارای کارآمدی مفیدی باشند حتی اگر کاربرد بهتر آن نیازمند اشتراک باشد. بیش‌تر کتابداران بودجه کافی برای پرداخت عضویت در تعداد روزافزون شرکت‌هایی که دسترسی به سنجه‌های وبی را به بهای گزافی فراهم می‌کنند، ندارند اما این امر به معنی این نیست که کتابداران نمی‌توانند از ابزارهایی که به رایگان در دسترس است آگاهی داشته باشند. در حال حاضر اجماع اندکی برای استانداردهای مناسب و سنجه‌های مورد استفاده اهداف خاص در بسیاری از رسانه‌های مختلفی که مورد بررسی این کتاب خواهند بود، و یا توقع وضع استانداردهای وسیعی توسط یک شخص با نفوذ در آینده، نیست. برعکس، کتابداران معمولاً باید از مسیری پر پیچ و خم راه خود را برای نیل به نزدیک‌ترین سنجه قابل دسترسی پیدا کنند، چه این سنجه‌ها محسوس‌ترین سنجه خدماتی مورد استفاده باشند (برای مثال تعداد دوستان در فیسبوک یا تعداد دنبال‌کننده‌ها در توییتر) یا یکی از تعداد بیشمار گروه‌های ثالثی که وعده راه حل ساده‌ای را به نیازهای سنجه‌ای یک شخص می‌دهد (برای مثال Klout.com یا Alexa.com). هدف این کتاب معرفی چنین سنجه‌هایی به جامعه کتابداران و کمک به آنها در شناخت دلایل مفیدتر بودن یک سنجه از سنجه

1. Thelwall 2. Tancer 3. Eysenbach

دیگر و نیـز محـدودیت ابزارهـای موجـود اسـت. سـنجه‌هـای جدیـد و بسـیاری در چارچوب نوآوری بازار پدیدار و پذیرفته می شوند. این کتاب به منظور ارائه شـیوه ها و ایده‌هایی که دانشگاهیان و متخصصـان بـا کـاربرد اسـتانداردها و روش‌شناسـی‌هـای حوزه‌های جدید مطرح کرده‌اند و نیز با تاکید بر نقش کتابداران تدوین شده است.

در این کتاب محدودیت های سنجه های وبی یا پیشینه آن‌هـا نادیـده گرفتـه نمـی شود بلکه ظرفیت آن‌ها به عنوان سنجه‌های نیمـه- ارزیابانـه[1] و معیارسـازی[2] ضـعیف بررسـی خواهنـد شـد (تلـوال، ۲۰۰۴). الزامـاً اگـر یـک وبگـاهی در یـک مـاه هـزار بازدیـدکننـده دارد در حـالی کـه وبگـاه مؤسسـه مشـابهی در یـک مـاه یـک میلیـون بازدیدکننده دارد نادرست نیست، اما این موضـوع تحقیقـات بـیش‌تـری را مـی‌طلبـد. سنجه‌ها بجای این‌که در انتهای یک مذاکره باشند باید در ابتدای آن مورد بحـث قـرار بگیرند. همه چیز را نمی توان به راحتی در اعداد خلاصه کرد و ممکن اسـت عملکـرد شخص یا مؤسسه‌ای به اندازه عملکرد دیگری خوب نباشد زیرا سـنجه انتخـاب شـده توانایی های خاص آن‌ها را نشان نداده است. در چنین اوضاعی اگر مـدیر یـا شـخص قانون گذاری بپرسد که چرا تأثیر مورد انتظار میسّر نشده است امـر منطقـی اسـت امـا لازم است که فضایی را برای شخص یـا مؤسسـه در نظـر بگیرنـد تـا در آن فضـا بـه فعالیت بپردازد.

ساختار ادامه کتاب

فصل ۲. کتاب‌سنجی[3] وب‌سنجی[4] و سنجه‌های وبی[5]

فصل دوم نگاه منسجم‌تری به انواع سنجه های مورد اسـتفاده جامعـه کتابـداری و نیز نحوه ارتباط آن‌ها با سنجه‌های وبی دارد. حوزه‌هـایی ماننـد کتـاب‌سـنجی در ایـن عرصه تاریخچه دیرینه ای داشته و شاید با بسیاری از ایرادهـای وارده بـر سـنجه هـای وبی مواجه شده باشند. این فصل علیرغم اذعـان بـر ایـن محـدودیت‌هـا، بـر ظرفیـت

1. semi-evaluative
2. benchmarketing
3. Bibliometrics
4. Webomrtrics
5. web metrics

گستره سنجه‌های جامعه متخصصان کتابداری و اطلاع‌رسانی و ارزش سنجش برای کار کتابداران و درس‌های آموخته شده قدیمی و قابل کاربرد در وب تأکید می‌کند.

فصل ۳. ابزارهای گردآوری داده‌ها

سنجه‌های وبی به شدت به ابزارها و داده‌های موجود وابسته هستند و فصل سوم به نحوه تغییر و گسترش این ابزارها در حوزه سنجه‌های وبی در دو دهه اخیر می‌پردازد. چهار بازه زمانی متمایز در تطبیق‌پذیری پژوهشگران با ماهیت متغیر وب و ابزارهای موجود در آن، برای پژوهش‌های وب‌سنجی مطرح بوده است. این بازه‌ها در خصوص محدودیت سنجه‌های وبی به دلیل ساختار وب و نیز ماهیت متغیر ابزارها و تغییرات احتمالی در آینده بوجود آمده‌اند.

فصل ۴. ارزیابی تأثیر در وب

با وجود افزایش خدمات رسانه‌های اجتماعی گروه ثالث، مطالب خود- میزبانی[1] هم‌چنان بخش مهمی از حضور وبی بسیاری از کتابخانه‌ها بوده و به عنوان یک منبع اطلاعاتی در خصوص سایر سازمان‌ها و جامعه عمومی‌تر به شمار می‌روند. فصل چهارم به بررسی سنجه‌های اندازه‌گیری تأثیر وبگاه‌ها، وب‌نوشت‌ها و سایر مطالب می‌پردازد از خدمات تحلیلی تا ارجاعات موجود در وب. این فصل هم‌چنین در خصوص استفاده از تحلیل محتوا برای بدست آوردن اطلاعات بیش‌تر درباره این منبع داده‌ای ساختارنیافته و بزرگ به بحث می‌پردازد.

فصل ۵. ارزیابی تأثیر رسانه‌های اجتماعی

خدمات رسانه‌های اجتماعی گروه- ثالث[2] نقش فزاینده چشمگیری در میزبانی مطالب دارند از جمله ایجاد فرصت برای راه‌اندازی سنجه‌های جدید و مشکلات جدید در گردآوری داده. فصل پنجم به انواع سنجه‌های مورد استفاده در بررسی وبگاه‌های شبکه اجتماعی و تحلیل عاطفی[3] می‌پردازد.

1. self-hosted 2. third-party 3. Sentiment analysis

فصل ۶. بررسی رابطه بین کنشگرها

سنجه‌های وبی محدود بـه اهـداف ارزیابانـه نیسـتند، زیـرا ممکـن اسـت بـرای عرضـه اطلاعات ارتباطی بر روی وب و وب اجتماعی مورد استفاده قرار گیرند. فصل ششم به برخـی ابزارها و روش‌های در دسترس نگاشت و تحلیل ارتباطات موجودیت‌های برخط می‌پردازد.

فصل ۷. کاوش انتشارات قدیمی در محیطی جدید

هـر چـه بـر اهمیـت انـواع جدیـد رسـانه‌هـای بـرخط افـزوده مـی‌شـود، کتاب‌شناسی[1]های قدیمی هم‌چنان مهم‌ترین بخش کـاری اغلـب کتابـداران محسـوب می‌شوند از انواع فیزیکی قدیمی موجود در قفسه ها گرفته تا نسـخه‌هـای الکترونیکـی آن‌ها. با این‌همه، فناوری‌های جدید راه‌های جدیدی را برای بررسی تأثیر فرمـت‌هـای قدیمی، چه به صورت ذکر برخط متون، یا شمارش تعداد اسناد بارگیری شده[2]، و هـم نشانه‌گذاری‌های[3] مدیریت مرجع مندلی[4]، عرضه می‌کنند.

فصل ۸. سنجه‌های وبی و وب داده‌ها

وب پیوسته در حال حرکت از وب اسناد به سمت یک وب عاطفی‌تـر اسـت. نـه تنها انتظار می‌رود واسپارگاه‌های[5] کتابخانه‌ها علاوه بر میزبانی اسـناد میزبـان داده‌هـای خام نیز باشند، بلکه وبگاه‌ها نیز بـه طـرز روزافزونـی داده‌هـا را در صـفحه‌هـای وب نشانه‌گذاری مـی‌کننـد. اگـر مـی‌خـواهیم تـاثیر داده‌هـای در دسـترس را بـدانیم بایـد سنجه‌های جدیدی را توسعه داد. این فصل برخی از چالش های موجـود و برخـی راه حل‌های احتمالی را مورد بحث قرار می‌دهد.

فصل ۹. آینده سنجه‌های وبی و متخصصان و کتابداری و اطلاع‌رسانی

به احتمال زیاد در آینده فناوری‌های جدید، افزایش فشار بر بودجه کتابخانه‌هـا و تأکید بیش‌تر بر استفاده از سنجه‌ها را شاهد خواهیم بود. فصـل آخـر بـه چـالش‌هـا و مسائلی که این موضوع برای کتابداران به دنبال دارد و راه‌حل‌های احتمالی بـرای آنـان در جهت رویارویی با چالش‌های آتی می‌پردازد.

1. bibliographic 2. downloads 3. Bookmarks 4. Mendeley 5. repositories

۲

کتاب‌سنجی، وب‌سنجی و سنجه‌های وبی

مقدمه

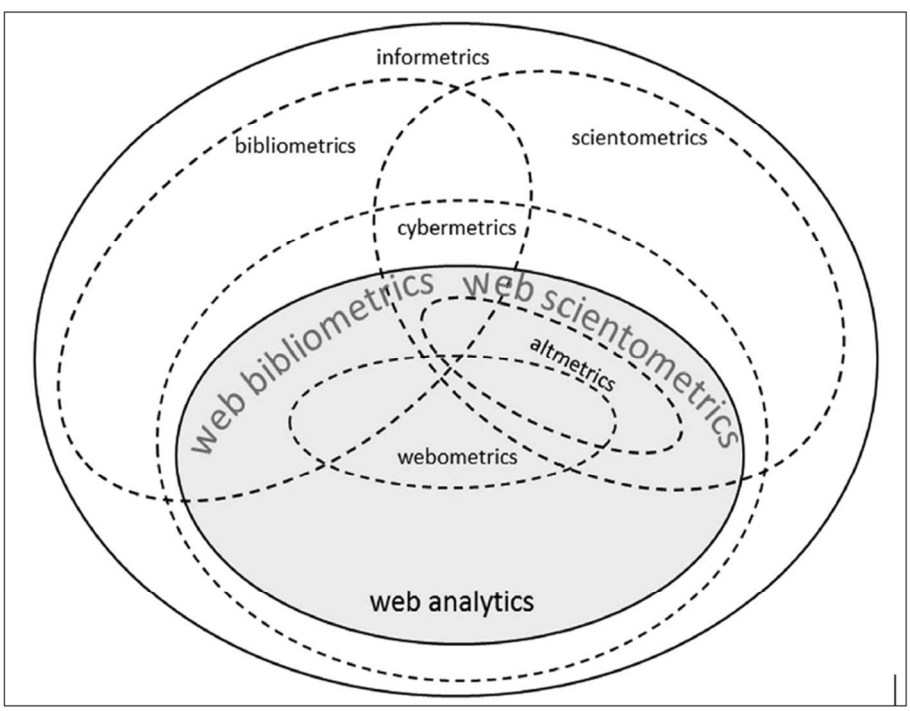
سنجه‌های وبی روش‌ها و ابزارهای جامعه علم اطلاعات را با اهداف و کاربردهای جامعه بازاریابی برای کتابداران ترکیب می‌کنند. بخش اول این فصل ارتباط سنجه‌های وبی و اصطلاحات مرتبط جامعه علم اطلاعات (برای مثال کتاب‌سنجی، علم‌سنجی و وب‌سنجی)، و همین طور ارتباط آن را با تحلیل وبی جامعه بازاریابی مورد بررسی قرار می‌دهد. بررسی‌های سنجه‌های وبی در این حوزه‌های مختلف را می‌توان کلاً به دو دسته ارتباطی[1] یا ارزیابانه[2] تقسیم کرد. بخش دوم فصل نگاه دقیق‌تری دارد بر اصول نظری و بررسی‌های عملی که می‌توانند تحت این عناوین مدّ نظر باشند و نیز شیوه احتمالی اعمال چنین بررسی‌هایی توسط کتابداران. بخش پایانی این فصل به اعتباریابی یافته‌های سنجه وبی می‌پردازد.

سنجه‌های وبی

اصطلاح «سنجه‌های وبی» در سراسر این کتاب بر سنجش کمّی ایجاد و کاربرد محتوای وب دلالت دارد. این واژه به عنوان اصطلاحی کلی برگزیده شده و

1. relational 2. evaluative

کاربردهای گوناگونی را در برمی‌گیرد که سنجه‌های وبی برای آن‌ها طراحی شده و نیز واژگان مختلفی که در طی سال‌ها برای آن‌ها پدیدار شده است. در نمودار بیجورنبورن[1] و اینگورسن[2] (۲۰۰۴)، در شکل ۲-۱ دامنه تداخل واژگانی این کتاب و سنجه‌های وب را نشان می‌دهد. بخش هاشور خورده در نمودار کتاب‌سنجی وب، علم‌سنجی وب، دگرسنجه‌ها[3]، وب‌سنجی و تحلیل وبی را نشان می‌دهد. اندازه بیضی‌ها بازنمون اندازه حوزه‌های پژوهشی نیستند بلکه برای شفاف‌سازی ارائه شده‌اند. هر یک از بخش‌هایی که سنجه‌های وب را شکل می‌دهند در ادامه با جزئیات بیش‌تر پرداخته خواهد شد.



شکل ۲-۱: واژه‌شناسی «سنجه» در کتابداری و اطلاع‌رسانی

1. Bjorneborn 2. Ingwersen 3. altmetrics

سنجه‌های علم اطلاعات

در علم اطلاعات طی سال‌ها تعدادی واژه‌های آمیخته با «سنجه» ابداع شده است. از برخی از آن ها استقبال فراوانی شده است (مانند کتاب سنجی)، برخی عمر کوتاهی داشته‌اند (مانند اینترنت‌سنجی[1])، و برخی تنها در حوزه‌های خاص بررسی می‌شوند (برای مثال دیسک‌سنجی[2]). تعریف دقیق و روشن واژگان از این جهت مهم است که پژوهش گران و سیاست‌گذاران علمی بتوانند با یکدیگر تعامل کنند (لازارو[3]، ۱۹۹۶). از طریق هم‌پوشانی واژگان سنجشی است که می‌توانیم نحوه مشارکت حوزه‌های مختلف با سنجه‌های وبی و توانایی سنجه‌های وبی را در فعالیت‌های تخصصی یک کتابدار شناسایی کنیم. واژگان سنجشی مختلفی که توسط جامعه علم اطلاعات ابداع شده غالبا در دو دسته کلی جای می‌گیرند: آن هایی که به موضوعات مورد سنجش مرتبطند، و دیگر آن هایی که مربوط به هدف سنجش موضوعات هستند. البته بین این دو هم‌پوشانی فراوانی وجود دارد.

اطلاع‌سنجی

اطلاع‌سنجی[4]، شامل گسترده‌ترین سنجش محتوا و مطالعه جنبه‌های کمّی هر نوع اطلاعات در هر شکلی می‌شود (تاگو- سوتکلیف[5]، ۱۹۹۲). از آنجا که شاید اطلاعات جزء دارایی بنیادین جهان مجسوب شود (استونیر[6]، ۱۹۹۹) حوزه‌های فراوانی برای اطلاع‌سنجی وجود دارد هرچند کاربرد اطلاع‌سنجی معمولاً محدود به مطالعه کمّی ارتباطات بین افراد است. اما اکثر تحقیقات اطلاع‌سنجی بر زیر مجموعه کتاب‌سنجی معطوف شده‌اند.

رسانه: از کتاب‌سنجی تا وب‌سنجی

آلن پریچارد[7] کتاب‌سنجی را «کاربرد روش‌های آماری و ریاضی برای کتاب‌ها و سایر روش های ارتباطی»(۱۹۶۹، ۳۴۹) تعریف کرده است. مجلات و کتاب ها در

1. Internetormetrics 2. discometrics 3. Lazarev
4. Informetrics 5. Tague-Sutcliffe 6. Stonier 7. Alan Pritchard

بیش‌تر مواقع شیوه‌های خاصی از ارتباطات را شکل می‌دهند بویژه در جامعه دانشگاهی، و در این کتاب از این معنا استفاده شده است. از کتاب‌سنجی می‌توان به صورت کلی‌تر برای اشاره به هر نوع سندی استفاده کرد اما تحلیل سایر منابع غالباً منتهی به ابداع اصطلاحات جدیدی برای هر نوع منبع خاص می‌شود: عبارت «حق امتیازسنجی[1]» دلالت بر کاربرد روش‌شناسی‌های کتاب‌سنجی در حق امتیاز ثبت اختراع دارد(مورتنسن[2]، ۲۰۱۱)؛ و «دیسک‌سنجی» دلالت بر کاربرد روش شناسی‌های مشابه برای ضبط صدا و آثار شنیداری دارد (روریک[3]، ۱۹۸۷؛ شوبرت[4]، ۲۰۱۲). در ابعاد محدودتر، کتاب‌سنجی می‌تواند برای اشاره به تحلیل کتاب‌شناسی‌های درون کتاب‌ها و مقاله‌ها استفاده شود (برای مثال وایت و مک‌کلین[5]، ۱۹۸۹). البته چنین کاربردی محدودتر از کاربرد کلی آن بوده و آثار بسیاری را حذف می‌کند زیرا عبارت مناسب‌تری وجود ندارد.

مقبولیت سریع وب در اواسط دهه ۱۹۹۰ و نیز بازشناسی شباهت‌های بین پیوندزنی صفحه‌های وبی و استنادهای بین مقاله‌های نشریه‌ها منجر به این شد که تعدادی از پژوهشگران به توانایی وب و اینترنت به عنوان شالوده پژوهش‌های اطلاعات‌سنجی پی ببرند. نام های پیشنهادی برای این حوزه‌ها شامل «تعامل علمی بواسطه اینترنت[6]» (بوسی[7]، ۱۹۹۵)، «اینترنت‌سنجی» (آلماینـد و اینگورسن[8]، ۱۹۹۶)، «وب‌سنجی» (آلماینـد و اینگورسن، ۱۹۹۷) و «سایبرسنجی[9]» (آگویللو[10]، ۱۹۹۷) بوده است.

از بین این اصطلاحات «وب‌سنجی» به صورت گسترده‌ای در جامعه علم اطلاعات پذیرفته شده است، و بدواً از تعریف بیجورنبورن و اینگورسن (۲۰۰۴، ۱۲۱۷) استفاده شده است: «مطالعه ابعاد کمّی ساخت و کاربرد منابع، ساختار و فناوری‌ها اطلاعاتی در وب برپایه رویکردهای کتاب‌سنجی و اطلاعات‌سنجی». سایبرسنجی نیز در موارد کم‌تری بجای اینکه تنها برای اشاره به وب باشد به منظور

1. patentometrics
2. Mortensen
3. Rorick
4. Schubert
5. White and McCain
6. netometrics
7. Bossy
8. Almind and Ingwersen
9. cybermetrics
10 . Aguillo

بررسی کلی تر اینترنت مورد استفاده می‌گیرد. هرچند اخیرا وب سنجی تعریف خاص‌تری پیدا کرده است. تلوال (۲۰۰۹، ۶) مفهوم وب‌سنجی را از مبحث ابعاد کمّی دور کرده و بیش‌تر به سمت اهداف سوق می‌دهد: «مطالعه محتوای وب با روش های در درجه نخست کمّی برای اهداف تحقیقات علوم اجتماعی با بکارگیری روش‌هایی که تنها مختص یک زمینه مطالعاتی نیستند». این روش‌های کمّی در جهت اهداف پژوهشی تنها محدود به علوم اجتماعی نیستند بلکه می توانند به برای رشته‌های دیگر نیز استفاده شوند. در این کتاب از وب‌سنجی برای اشاره به مطالعه محتوای وب با روش های کمّی برای اهداف پژوهشی استفاده شده است. با توجه به اینکه سنجه‌های وبی در سراسر این کتاب بر اندازگیری کمّی ایجاد و کاربرد محتوای وب دلالت دارد، بنا بر این لازم است بین سنجه‌های وبی به صورت کلی‌تر و وب‌سنجی به صورت دقیق پژوهشی تمایز قائل شد. تغییر در معنای وب‌سنجی برای یک هدف خاص با انواع دیگر اطلاع‌سنجی هم‌راستا است، چراکه با هدف‌های ارزیابانه مانند «علم‌سنجی» و «دگرسنجی» همسو می‌شوند.

هدف: از علم‌سنجی تا دگرسنجه‌ها

آغاز علم سنجی معمولاً با اثر ارزنده اصلی پرایس[1] (۱۹۶۳) به نام *علم کوچک، علم بزرگ*[2] گره می‌خورد که در آن از یک روش کمّی برای شناخت ساختار علوم جدید استفاده شده است. خود واژه علم سنجی، یا حداقل معادل روسی آن، در سال ۱۹۶۹ توسط نالیموو[3] و مولچنکو[4] برای اشاره به مطالعاتی درباره «تمامی جنبه‌های کمّی علم علم، ارتباطات در علم، و سیاست علم» (هود[5] و ویلسون[6]، ۲۰۰۱، ۲۹۳) ابداع شد. در عمل، طی سالیان زیاد مطالعات کتاب‌سنجی اکثریت قریب به اتفاق مطالعات علم‌سنجی را تشکیل می‌داد و تعداد کم‌تری به کاربرد روش‌های کتاب‌سنجی برای حق ثبت اختراع می‌پرداختند (نارین[7]، ۱۹۹۴). کاربرد روش‌های کتاب‌سنجی برای حق ثبت اختراع می‌تواند حاکی از اهمیت تعامل علم و صنعت در فرایند نوآوری

1. Price
2. Little Science, Big Science
3. Nalimov
4. Mulchenko
5. Hood
6. Wilson
7. Narin

باشد(برای مثال لوندوال[1]، ۱۹۹۲؛ گیبونز و همکارانش[2]، ۱۹۹۴؛ اتزکوویتز[3] و لیدسدورف[4]، ۱۹۹۵). با این وجود، حتی با افزودن حق ثبت اختراع، انتشارات علمی قدیمی تصویر محدودی از گفتمان علمی را ارائه می‌دهند و اخیراً به ایجاد سنجه های جایگزین توجه بیش‌تری شده است.

وب شیوه‌های ارتباطی بسیاری، فراتر از چارچوب‌های قدیمی انتشارات فراهم کرده و روش‌های جدیدی را برای سنجش تأثیر در جامعه دانشگاهی و سایر بخش های جامعه عرضه کرده است. در صورتی که اطلاعات در یک قالب استاندار ساختاربندی شده باشد سنجش تأثیر ساده‌تر است و نفوذ تعداد سایت‌های شبکه‌های اجتماعی به این معناست که میلیون ها کاربر اطلاعات منسجمی را به شکل مشابهی ساختاردهی می‌کنند. واژه دگرسنجی که توسط پرایم[5] و دیگران (۲۰۱۰) ابداع شده بر استفاده از ماهیت ساختارمند فناوری‌های وب ۲/۰ در جهت پالایش و بررسی شاخص‌های تأثیر پژوهش دلالت دارد. امروزه با وجودی که دگرسنجه ها می‌توانند در حوزه ارتباطات علمی آگاهی‌های فراوانی ارائه دهد اما تمرکز اصلی آن بر روی بررسی خروجی های علمی قدیمی در اینترنت است. در این کتاب برای دگرسنجه ها تعریف کلی‌تری پذیرفته شده است که تأثیر شکل‌های جدید ارتباطی را پوشش داده و می‌توانند بخشی از خروجی یک پژوهشگر نیز باشند.

تحلیل‌های وبی

دانستن این نکته مهم است که کاربرد سنجه‌های وبی در مطالعات کتاب‌سنجی و علم‌سنجی علم کتابداری و اطلاع‌رسانی بویژه در خصوص تحلیل وبی با پیش‌زمینه آن در حوزه بازاریابی سابقه ندارد. بازاریابی به توانایی سنجه های وبی به منظور روشنگری در خصوص برخی نظریه‌های انتزاعی یا رشد علم علاقه‌ای ندارد، بلکه به دنبال کسب اطلاعات درباره مشارکت در تجارت موفق کالا و خدمات است. در اینجا واژه «تحلیل های وبی» برای اشاره به کاربرد سنجه‌های وبی برای درک و بهینه‌سازی کاربرد وب است.

1. Lundvall 2. Gibbons et al. 3. Etzkowitz 4. Leydesdorff 5. Priem

تعدادی ابزار و فناوری‌های بسیار جدید برای جامعه بازاریابی ایجاد شده‌اند، ولواین‌که گاهی ابزارهای تولید شده برای اهدافی خارج از اهداف توافقی و بهینه‌سازی کاربرد وب، استفاده شوند. برای مثال موضوعات داغ گوگل[1]، آلکسا[2] و اس‌ای‌او ماجستیک[3] نمونه ابزارهایی هستند که در آغاز برای اهداف بازاریابی راه‌اندازی شده‌اند ولی بعد از آن در جوامع دانشگاهی برای اهداف پژوهشی مورد استفاده قرار گرفته‌اند.

اگرچه این ابزارها در بازاریابی کاربرد عملی‌تری دارند، اما کتابداران از روش‌شناسی سنجه‌های وبی برای اهداف پژوهشی استفاده می‌کنند. برای مثال ممکن است یک کتابدار به تأثیر وب‌نوشت های شخصی خود علاقه‌مند باشد، یا ممکن است مدیر کتابخانه‌ای بخواهد بداند آیا با کاربران کتابخانه خود تعامل خوبی دارد یا خیر.

سنجه‌های ارزیابانه و رابطه‌ای

همان‌گونه که در بالا اشاره شد، سنجه‌های وبی دارای کاربردهای فراوانی هستند از مطالعه یک پژوهشگر در خصوص رشد حوزه‌های علمی گرفته تا شخصی که می‌خواهد بداند وب‌سایتش چند بازدیدکننده داشته است. به‌جای پرداختن به جزئیات انواع بررسی‌های احتمالی در هر یک از حوزه های علمی و نیز اصول نظری آن ها، بهتر است به دسته‌بندی کلی سنجه‌ها در قالب رابطه‌ای یا ارزیابانه پرداخته شود.

بورگمن و فورنر[4] (۲۰۰۲) در تحلیلی از کتاب‌سنجی ارتباطات علمی هفت بُعد را شناسایی کردند که کتاب‌سنجی می‌تواند در آن‌ها دسته‌بندی شود. تحلیل‌های آن‌ها نوعی از رفتار علمی را در بر می‌گرفت که نقطه تمرکز بررسی‌ها (پیوندزنی، نگارش، ارسال یا همکاری)، سطح تجمع (شخص، گروه، حوزه یا کشور) و هدف (توصیف، توضیح، پیش‌بینی یا ارزیابی) تلقی می‌شود. در کتاب حاضر سنجه‌های وبی در کل با توجه به رویکرد در نظر گرفته شده می‌توانند دسته‌بندی شده باشند: رابطه‌ای و ارزیابانه.

1. Google Trends(www.google.com/trends)
2. Alexa(www.alexa.com)
3. Majestic SEO(www.majesticseo.com)
4. Borgman and Furner

این دسته‌های کلی در نوع‌شناسی کتاب سنجی بورگمن و فورنر (۲۰۰۲) برای این استفاده شده‌اند تا سطح دقیق‌تری از سنجش را به بحث رفتار پیوندزنی اضافه نمایند. تحلیل رفتار پیوندزنی به شکل تحلیل استنادی موضوع اصلی تحلیل کتاب‌سنجی بوده است. همانگونه که در عصر نسخه الکترونیکی اسناد لازم نیست ارتباطات بین اسنادی[1] محدود به استنادها شوند، اما بورگمن و فورنر واژه «پیوندزنی» را به‌جای استناد[2] انتخاب کرده‌اند. تحلیل پیوند رابطه ای ارتباط بین کنشگران را بررسی می‌کند، حال می‌خواهد این کنشگران اشخاص باشند، یا سازمان‌ها، مجله‌ها و یا مقاله‌ها. مطالعه نحوه ارتباط پژوهشگران از طریق تحلیل هم-استنادی[3]، نمونه‌ای از تحلیل پیوند رابطه‌ای به شمار می‌رود (نویسنده‌هایی که به یک مقاله دانشگاهی استناد کرده‌اند). تحلیل پیوند ارزیابانه، در بیش‌تر موارد با استفاده از تحلیل استنادی، به مطالعاتی اشاره دارد که از پیوندها به منزله شاخص کیفیت یا تأثیر یک کنشگر استفاده می‌کنند، چه این کنشگر مجله باشد، چه پژوهشگر و چه یک گروه پژوهشی.

در این کتاب کاربرد دو واژه « تحلیل رابطه‌ای[4]» و «تحلیل ارزیابانه[5]» تنها محدود به استفاده آن ها در تحلیل پیوند نیست، بلکه در بررسی دو رویکرد مختلف بررسی هر موجودیت سنجه وبی بکار می‌روند: سنجه‌های وبی رابطه‌ای، در درجه نخست بر رابطه بین موجودیت‌ها تمرکز می کنند؛ اما تحلیل ارزیابانه در جایی استفاده می‌شود که از رابطه یک موجودیت با سایر موجودیت‌ها بتوان ارزشی را استنباط کرد. در برخی موارد تفاوت بین تحلیل رابطه‌ای و ارزیابانه تنها یک نگرش بوده و شالوده هر دو را یک روش‌شناسی واحد شکل می‌دهد.

سنجه‌های وب ارزیابانه

بطور کلی سنجه‌های ارزیابانه مهم‌ترین بخش سنجه‌های وبی و کتاب‌سنجی تلقی می‌شوند. این سنجه‌ها برای ارزیابی تأثیر پژوهش‌های دانشگاهی که کتاب‌سنجی به

1. inter-linking 2. citing 3. co-citation 4. relational analysis 5. evaluative analysis

آن توجه وافری دارد، و نیز پیگیری تأثیر محتوای وبی که علت اصلی استفاده از سنجه‌های وبی است، مورد استفاده قرار می‌گیرند.

روش‌های مختلفی وجود دارد که می‌توان از طریق آن‌ها تأثیر وبی موجودیت‌های خاص را ارزیابی کرد و موجودیت‌های فراوانی هستند که می‌توان تأثیر آن‌ها را اندازه گرفت. در حالی که می‌توان سنجه‌های وبی ارزیابانه برای اهداف تحلیل وبی استفاده شوند، برای اهداف وب‌سنجی، کتاب‌سنجی وبی و علم‌سنجی نیز مفید هستند. مانند مقاله مجله و استناد که مرکز توجه تحلیل‌های کتاب‌سنجی بوده‌اند، در هر بررسی ما به یک موجودیت وبی و یک ارزیابی تأثیر اشاره داریم. بررسی در قلمرو کتاب‌سنجی ارزیابانه کمک می‌کند تا برخی ظرفیت‌ها، مشکل‌ها و محدودیت‌های سنجه‌های وبی نشان داده شوند.

کتاب‌سنجی ارزیابانه

طی سال‌های متمادی کتاب سنجی ارزیابانه به منزله مجموعه ابزارهایی تلقی شده است که به کتابدارها کمک می‌کند تا مجموعه‌های خود را مدیریت کنند. از آنجایی که کتابدارها به ندرت بودجه‌های نامحدود دارند، لازم است که کتابداران روشی داشته باشند تا با آن بتوانند مجله‌های هسته برای کاربرهای خاص خود شناسایی نمایند. لذا ضریب تأثیر مجله[1] به مثابه روشی برای شناسایی مجله‌های هسته برای نمایه استنادی علوم[2] توسعه یافت، و با تقسیم تعداد استنادهای دریافتی یک سال مقاله‌های منتشره در دو سال قبلی، بر تعداد مقاله‌های منتشره در طی آن دو سال محاسبه می‌شود (گارفیلد، ۲۰۰۶). بین محاسبه تأثیر مجله‌ها و رتبه‌بندی مجله‌ها گام کوچکی است و ایجاد نمایه استنادی علوم در سال‌های ۱۹۶۰ فرآیند دسترسی پژوهشگران به داده‌های استنادی را تسهیل کرده است بویژه در مورد استنادهایی که خارج از حوزه خاص خود بودند. در حال حاضر واژه ضریب تأثیر[3] برای در برگیری تأثیر مجله و تأثیر مؤلف بسط یافته است.

1. journal impact factor 2. Science Citation Index 3. impact factor

کتاب‌سنجی‌ها بیش‌تر برای ارزیابی تاثیر خروجی‌های پژوهش افراد و گروه‌ها استفاده می‌شوند: دولت‌مردان وقتی که به دنبال شناسایی بهترین روش برای تخصیص منابع مالی پژوهشی هستند به توانایی نمایه‌های کتاب‌سنجی توجه نشان می‌دهند، در حالی که کتاب‌سنجی از منظر مؤسسه‌ها روشی است برای پیگیری تأثیر خروجی‌های دانشگاهی. بدترین وضعیت این است که از آن‌ها برای جایگزینی تأثیر یک مقاله استفاده شود در حالی که محاسبه ضریب تأثیر مجله آسان‌تر بوده و لازم نیست صبر کنند تا واقعاً به مقاله استناد شود. تعداد دفعاتی که مقاله‌های یک مجله خاص مورد استناد قرار می‌گیرند بسیار متفاوت است. در صورت اندک بودن مقاله‌های یک مجله باید ضریب تأثیر مجله بدست آید. فعالیت‌هایی در مقابل سوءاستفاده از ضریب تأثیر مجله انجام شده است (برای مثال بیانیه ارزیابی پژوهش سان فرانسیسکو[1]). حتی اگر از سنجش‌های نویسنده–محور[2] استفاده شود خروجی پژوهشی یک نویسنده امکان مقایسه ساده بین افراد در تنها یک شکل فراهم می‌کند و برای محاسبه چنین شکلی تعداد روش‌های متفاوت زیادی وجود دارد. هرش[3] (۲۰۰۵) در مقاله خود برای معرفی شاخص اچ[4] چهار مورد را لحاظ کرد: تعداد کل مقاله‌ها، تعداد کل استنادها، استنادهای هر مقاله، و تعداد مقاله‌های مهم (مقاله‌های مهم مقاله‌هایی هستند با بیش از یک تعداد خاص استنادهای اختیاری). ظاهراً هر یک از اینها مزایا و معایبی دارند: بهره‌وری از طریق تعداد مقاله‌های منتشره ارزیابی می‌شود، در حالی که تأثیر آن‌ها را ارزیابی نمی‌کند؛ تعداد استنادها تأثیر را ارزیابی می‌کند اما امکان دارد که با شمار اندکی مقاله پر استناد تجمیع شده باشد؛ استناد به ازای هر مقاله ممکن است بهره‌وری بالا را کاهش دهد؛ میزان مناسب مقالات مهم در دوره های کاری مختلف هر پژوهشگری می‌تواند متفاوت باشد، هرچند تعداد مقالات مهم بر اساس سطح تعریف شده می‌تواند بطور اتفاقی موجب حمایت یا ضرر یک پژوهشگر شود.

1. San Francisco Declaration on Research Assessment
2. author-centric
3. Hirsch
4. *h*-index

شاخص اچ هرش برای اشاره به معایب روش‌های پیشین ارزیابی خروجی یک پژوهشگر مطرح شد و اینگونه است که اگر پژوهشگری اچ مقاله منتشر کرده باشد و هر یک از آنها حداقل اچ بار مورد استناد قرار گرفته باشند آنگاه آن پژوهشگر دارای شاخص اچ از اچ است (هرش، ۲۰۰۵). برای نمونه، اگر نویسنده‌ای پنج مقاله منتشر کرده باشد و آن‌ها ۵، ۴، ۴، ۲ و ۱ تعداد استناد داشته باشند، از آنجایی که سه مقاله وی حداقل سه بار به آن‌ها استناد شده است، شاخص اچ آن نویسنده ۳ خواهد بود. این فکر که تأثیر یک پژوهشگر امکان دارد به یک معیار واحد تقلیل پیدا کند قطعاً برای اهداف ارزیابی زیانبار است و از زمان ابداع آن تغییرات شاخص اچ برای این بوده است که محدودیت‌های شناخته شده آن نشان داده شود. بورنمن، موتز و دنیل[1] (۲۰۰۸) هشت مورد را شناسایی کرده‌اند:

- خارج قسمت *م*[2] برای این منظور طراحی شده تا مشکل شاخص اچ یک شخص را که در وهله اول بازتاب طول دوره کاری پژوهشی آنها است را با تقسیم این عدد بر تعداد سال هایی که فعالیت پژوهشی داشته‌اند نشان دهد.
- شاخص جی[3] برای این منظور طراحی شده تا وزن مفروضی را به مقاله‌های پراستناد اضافه نماید، که در آن شاخص جی یک نویسنده بیش‌ترین عدد جی مقاله‌هایی باشد که با هم بیش از جی به توان دو استناد دریافت می‌کنند.
- شاخص اچ (۲)[4] نیز برای اختصاص وزن بیش‌تر به مقاله‌های پراستناد طراحی شده است، که در آن شاخص اچ (۲) یک پژوهشگر بالاترین مقداد اچ (۲) در بین مقاله‌هایی است که هر یک [اچ (۲)] به توان دو مورد استناد قرار گرفته‌اند.

1. Bornmann, Mutz and Daniel
2. m quotient
3. *g*-index
4. *h(2)*-index

- شاخص ای[1] تعداد میانگین استنادهای مقاله در هسته هرش است (آن مقاله‌هایی که به اندازه‌ای به آن‌ها استناد شده است که با شاخص اچ پژوهشگر کمک کنند).
- شاخص ام[2] عدد میانه استنادهای مقاله‌های هسته هرش برای اشاره به ماهیت مخدوش توزیع استنادها است.
- شاخص آر[3] ریشه دوم تعداد استنادها در هسته هرش را در بر می‌گیرد. این شاخص در پاسخ به این موضوع است که شاخص ای، پژوهشگر را به دلیل داشتن شاخص اچ بالا تنبیه کرده است.
- شاخص ای آر[4] شاخص آر را برای در نظر گرفتن سن مقاله‌های هسته هرش تعدیل می‌کند.
- شاخص اچ دبلیو[5] مطابقت شاخص اچ است که تأثیر استنادها را مدّ نظر قرار می‌دهد.

عجیب است که تقلیل خروجی یک پژوهشگر به یک عدد واحد موجب پایه‌ریزی چنین طیف گسترده‌ای از سنجه‌ها شده است و نشانه‌های کمی هست که توافق همگان برای یک سنجه کاملاً مناسب نشان دهد. مقاله بورنمن، موتز و دنیل (۲۰۰۸) مطرح می‌کند که بخشی از چرایی این موضوع به تقسیم انواع شاخص اچ به دو دسته، برمی‌گردد: یک نوع آن اندازه هسته مولد پژوهش یک دانشمند را ارزیابی می‌کند و دیگری تأثیر هسته مولد را اندازه می‌گیرد. سنجه‌های چندگانه شاخص بهتری ارائه کرده و توصیه می‌کنند که از هر دو سنجه که مکمل هم هستند استفاده شود و اگر قرار است تنها از یکی از آنها استفاده شود سنجه‌ای که تأثیر هسته مولد را اندازه می‌گیرد می‌تواند برای ارزیابی‌های همسان پیش‌بینی بهتری بکند. با این وجود، اگر تنها از یک ارزیابی استفاده شود، استفاده از یکی از انواع پیچیده‌تر شاخص اچ زیاد به صرفه نیست (شریبر، مالسیوس و پی ساراکیس[6]، ۲۰۱۲).

1. *a*-index
2. *m*-index
3. *r*-index
4. *ar*-index
5. *hw*- index
6. Schreiber, Malesios and Psarakis

با وجود تمایل به تقلیل کار هر پژوهشگر به یک عدد ثابت، بی شک چنین ارزیابی ارزش کار پژوهشگر را بیش از حد ساده ساخته و توجه فراوان به سنجه‌ها خطرهایی نیز به همراه دارد. اشخاصی که دست اندرکار توسعه سنجه‌ها هستند نیز این نکته را تکذیب نمی‌کنند. گارفیلد (۲۰۰۶) به خطرهای ذاتی[1] اشاره کرده است هرچند وی هم‌چنین قبول دارد که بیش‌تر مردم وقت مطالعه تمام مقاله‌های مرتبط را در اختیار ندارند. هرش(۲۰۰۵) بر این باور است «یک عدد ثابت در خصوص پروفایل چند بعدی یک شخص چیزی در حد یک گمان تقریبی است و برای ارزیابی یک شخص باید عوامل بسیار دیگری در نظر گرفته شوند». با این وجود، سادگی سنجه واحد بر تصور جامعه علمی غالب شده است و زمانی که گوگل اسکالر[2] طرح استنادهای من[3] را معرفی کرد[4] نه تنها شاخص اچ را در نظر گرفته بود بلکه شاخصی را نیز معرفی کردند که تعداد مقاله‌های مهم را ارزیابی می‌کرد یعنی شاخص آی ۱۰[5] که دلالت بر تعداد آثاری دارد که حداقل ده استناد داشته‌اند (برای بحث مفصل‌تر درباره استفاده سنجه‌های گوگل اسکالر فصل هفتم را ببینید). هر دو گزارش استنادی برای مجموعه منتخبی از مقاله‌ها در *وب دانش*[6] تامسون رویترز[7] و مرور استنادها برای مجموعه‌ای از مقاله‌ها در *اسکوپوس*[8] نیز شاخص اچ را برای تعداد خاصی از مقاله‌ها محاسبه کردند.

تعداد دفعاتی که به یک پژوهش استناد می‌شود بسیار متفاوت است. در یک سمت مقاله لوری و دیگران[9] (۱۹۵۱) قرار دارد که از زمان نگارش ۳۰۱/۰۳۹ بار مورد استناد قرار گرفته است در حالی که در سمت دیگر میلیون ها مقاله است که یا هرگز استنادی نداشتند یا تنها یک یا دو بار مورد استناد قرار گرفته‌اند. البته این امر دلیلی بر بدی آنها یا مشارکت کم آنها در مباحثه‌های علمی نیست بلکه ارزش برخی از این مقاله‌ها الزاماً در استنادها و جریان دیرینه دانشگاهی منعکس نشده است.

1. inherent dangers
2. Google Scholar
3. my citation
4. (http://scholar.google.co.uk/citations)
5. *i10*-index
6. *web of knowledge*
7. Thomson Reuter
8. *Scopus*
9. Lowry et al.

همانگونه که مک رابرتز و مک رابرتز[1] (۱۹۹۶) اشاره کرده‌اند، در خصوص کاربرد استنادها به عنوان یک شاخص ارزش دغدغه‌های زیادی وجود دارد: نویسنده‌ها به تمامی مقاله‌هایی که کار آنها متأثر از آنها بوده استناد نمی‌کنند و زمانی که به تمام مقاله‌های تأثیرگذار استناد نشود مقاله‌های مورد استناد می‌تواند موجب سوگیری به نفع دوستان، همکاران و خود–استنادی‌ها داشته باشد؛ منابع ثانویه از قبیل مقاله‌های مروری بر پژوهش اصلی ترجیح داده شوند؛ و نویسنده‌ها به دلایل زیادی مانند ردّ نتایج مقاله استناد شده به مقاله‌ها استناد می‌کنند. درحالی‌که ممکن است یک کتاب‌سنج مدعی شود بسیاری از مشکلات در نهایت و بطور میانگین تأثیر یکدیگر را خنثی می‌کنند اما هم‌چنان این دغدغه وجود دارد که پژوهش خوب شناخته نشود درحالی که پژوهش بد با استفاده از سنجه‌های استنادی مورد حمایت قرار گیرد. این احتمال مخصوصاً در بررسی میزان ارزیابانه بودن بروز می‌کند. برای نمونه، تأثیر انتشارات یک کشور یا یک دانشگاه ممکن است یکدیگر را خنثی کنند، اما امکان دارد جزء کوچکی از آن کشور یا یک شخص از مدل رفتاری استنادی خود متضرر شوند (یا برعکس سود بیش از حدی ببرند). چنین ترس‌هایی حتی اگر پایه و اساسی داشته یا نداشته باشند در هر حال جریان دارد. در پژوهشی درباره نحوه ادراک استنادهای دانشمندان، آکسنس و ریپ[2] (۲۰۰۹) به بررسی دیدگاه‌های دانشمندان نروژی که مقاله‌های پر استنادی منتشر کرده بودند پرداختند. آن‌ها متوجه شدند گرچه دانشمندان استنادها را با سهم کار خود مرتبط می‌دانند، اما به صورت کلی‌تر نسبت به استنادها بسیار بدبین بودند. بدون شک در مواردی به مقاله‌هایی استناد شده است که الزاماً کیفیت بالایی نداشته‌اند. به‌دنبال رسوایی شون[3] که در آن فیزیک‌دانی مدعی پیشرفت غیر منتظره متقلبانه‌ای را در حوزه نیمه هادی‌ها شده بود، چند مجله پیشرو هم مقاله‌هایی که منتشر کرده بودند را بازپس گرفتند. هفت مقاله‌ای که توسط مجله *طبیعت*[4] بازپس گرفته شده بودند در حال حاضر به ترتیب ۶۲، ۲۹، ۱۶۵، ۱۳۲، ۲۲۴،

1. MacRoberts and MacRoberts 2. Aksnes & Rip 3. Schon scandel 4. *Nature*

۱۷۱ و ۱۳۶ استناد دارند (با استناد به گوگل اسکالر). چنین گزارشی هشداری است برای کسی که به «کیفیت» استنادها بیش از حد بها می‌دهد.

بزرگترین نگرانی در رابطه با کتاب‌سنجی با اهداف علم‌سنجی این نیست که سنجه‌ها به قدر کافی تأثیر تحقیقات را منعکس نمی‌کنند بلکه توانایی ایجاد تأثیر منفی بر فرایندهای علمی بویژه با افزایش اهمیت سنجه‌ها است. به این موضوع می‌توان همانند معادل اطلاعات در قانون گریشام[1] توجه کرد که مردم می‌گویند «پول بد پول خوب را از بین می‌برد»(رولنیک و وبر[2]، ۱۹۸۶). ایده این است که اگر پول خوب و بد یک ارزش ظاهری داشته باشند (ارزش پول بد بیش از حد معمول باشد)، پول بد پول خوب را از سیستم بیرون می‌کند. در تحلیل کتاب‌سنجی به ویژگی‌های خاصی صرف نظر از داشتن ارزشی یکسان ارزش‌های مشابهی داده می شود. برای مثال، ممکن است تمامی مقاله‌های هم‌ترازخوانی شده ارزش یکسانی داشته یا استنادهای دریافتی هم ارزش تلقی شوند. این موضوع می‌تواند منجر به موقعیتی شود که در آن نویسنده ها تحریک شوند بیش از حد لازم منتشر کنند یا در راستای افزایش استنادهای خود به مبالغه و مباحثه بپردازند. این نکته در قانون دیگری از علم اقتصاد به صورت دقیقی جمع‌بندی شده است در قالب قانون گودهارت[3]: «وقتی که هدف ارزیابی است، دیگر ارزیابی خوبی نیست» (استراترن[4]، ۱۹۹۷، ۳۰۸). اشخاصی که می‌خواهند از چنین شیوه‌هایی اجتناب کرده و هم‌چنان از هنجارهای قدیمی علم استفاده کنند ممکن است احساس کنند که اگر بخواهند از کسانی که بر مبنای سنجه ها پژوهش‌های موفق‌تری دارند، عقب نمانند بناچار باید رفتار شغلی خود تعدیل نمایند.

با توجه به بحث برانگیز بودن تحلیل استنادی با اهداف ارزیابانه و تمایل به کاهش کار یک پژوهشگر به یک عدد ثابت، به نظر می‌رسد در بررسی‌های علم‌سنجی ارزیابانه جایی برای کتاب‌سنجی وب وجود نداشته باشد. با اینهمه،

1. Gresham's law 2. Rolnick & Weber 3. Goodhart's law 4. Strathern

درحالی‌که استنادها به دلایل زیادی ایجاد می‌شوند، ولی پژوهش های هم‌ترازخوانی شده دارای اعتباری هستند که بیش‌تر مطالب وب فاقد آن است. تقلیل پژوهش به یک سنجه واحد که کل کار پژوهشگر را جمع‌بندی کند، چیز ضروری نیست بلکه مجموعه‌ای از سنجه‌ها مورد نیاز است که از استحکام متفاوتی برخوردار بوده و بتوانند برای نشان دادن تصویر کامل‌تری از فرایند پژوهش ترکیب شود. به همین دلیل سنجه‌های وبی موارد زیادی را برای عرضه کردن در اختیار دارند.

کتاب‌سنجی وبی ارزیابانه و علم‌سنجی وبی

چند سالی است که توانایی سنجه‌های وبی برای کتاب‌سنجی ارزیابانه شناسایی شده است. بیش از یک دهه از زمانی گذشته است که هارناد[1] (۲۰۰۳) و دیگران پیشنهاد دادند که پروژه ارزیابی پژوهش[2] انگلیس، پیشگام نظام تعالی پژوهش ها[3] که از طریق آن میلیون‌ها پوند بودجه پژوهشی در انگلیس توزیع می‌شود، می‌تواند استفاده از رزومه‌های برخط برای پژوهشگران را منوط به ایجاد پیوندهایی برای مقاله‌های دسترسی آزاد بداند. این موضوع موجب می‌شود که تعداد شاخص های علم‌سنجی بیش‌تری مثل تعداد دفعات بارگیری یک مقاله بجای عرضه صرف در پایگاه اطلاعاتی استنادی، مطرح شوند.

سرعت، ضروری‌ترین مزیت سنجه‌های وبی برای اهداف ارزیابانه است. سنجه‌های وبی تأثیر یک اثر را سریع‌تر از پایگاه‌های اطلاعاتی استنادی قدیمی نشان دهند (البته جز در مواردی که از ضریب تأثیر مجله به عنوان یک جایگزین استفاده می‌شود). فرایند انتشار سنتی هم چنان یک فرایند طولانی باقی مانده است. احتمال دارد که یک اثر در توسعه پژوهش دیگری بسیار تأثیرگذار باشد، اما فرایندی که منجر به یک استناد می‌شود ممکن است طولانی و دشوار باشد. در ابتدا شاید لازم باشد پژوهشی که بر آن تأثیر گذاشته شده انجام شود. سپس مقاله را باید با در نظر داشتن مقاله اصلی به نحو امیدوارکننده ای نگارش کرد. آنگاه مقاله باید از فرایند

1. Harnad 2. Research Assessment Exercise (RAE) 3. Research Excellence Framework (REF)

هم‌ترازخوانی که شاید یک فرایند طولانی باشد، عبور کند بویژه اگر مقاله اصلی به مجله‌ای با ضریب تأثیر بالا ارسال شده باشد. در نهایت مقاله منتشر می‌شود، هرچند احتمال دارد که هم‌چنان زمان زیادی مسکوت بماند زیرا مقالات بسیاری برای پر کردن مطالب مجلات آتی وجود داشته باشند. در این زمان ممکن است نویسنده مؤثر اصلی ارزیابی شده، و از ارتقا چشم‌پوشی کرده یا به موقعیت جدیدی رسیده باشد. سنجه‌های وب توانایی عرضه شاخص‌های سریع را دارند، چه به صورت نشانه‌گذاری‌ها، چه بحث در وب‌نوشت‌ها و توییتر و یا بارگیری‌های مقاله نشریات.

این توانایی برای طیف گسترده‌تری از سنجه‌های روزآمدتر هست که به دگرسنجه رو بیاورند، و تعداد روزافزونی ابزار برای ارزیابی تأثیر پژوهش دانشگاهی در وب وجود دارد. پیشرفت فناوری‌های وب ۲/۰، مانند سکوهای وب‌نوشت و وبگاه‌های شبکه های اجتماعی، برای تمام اشخاصی که از حداقل مهارت فنی و اتصال به اینترنت برخوردار هستند یک سکوی شخصی نشر اطلاعات فراهم کرده و موجب افزایش انتشار داده‌های ساختاریافته شده است. این‌ها خدمات منسجمی را در بر می‌گیرد که ممکن است نیاز به اشتراک داشته باشند، برای نمونه، دگرسنجه دات کام[1] و یا رایگان باشند مانند جریان تأثیر[2]. تعداد زیادی از نشریات به تعداد دفعاتی که به یک مقاله استناد شده اشاره نمی‌کنند بلکه سنجه‌های جایگزین دیگری دارند. برای نمونه، پلوس وان[3] دیدگاه‌ها، استنادها و نشانه گذاری‌های دانشگاهی را منتشر می‌کند.

دگرسنجه‌ها ابزارهای جدیدی را برای رتبه‌بندی کنشگرها ایجاد می‌کند. برای مثال، در دسامبر ۲۰۱۲ مجله *طبیعت* فهرستی از مقاله‌هایی که عظیم ترین تأثیر را در رسانه‌های اجتماعی داشته‌اند منتشر کرده است (ون نوردن[4]، ۲۰۱۲). چنین رتبه‌بندی‌هایی مجبور نیستند که از ارزیابی‌های قدیمی پیروی کنند. برای مثال، مندلی[5] به عنوان یک شبکه های اجتماعی دانشگاهی و مدیر منابع در *گزارش پژوهش جهانی*[6] خود می‌تواند رتبه‌بندی مؤسسه‌ها را با توجه به زمان برخط بودن

1. (altmetric.com)
2. (http://impactstory.org)
3. PLOS ONE
4. Van Noorden
5. Mandeley
6. Global Research Report

پژوهشگران و تعداد مقاله‌های موجود در کتابخانه برخط آنها در اختیار قرار دهد (مندلی، ۲۰۱۲). در چند پژوهش ارتباط بین دگرسنجه‌ها با شاخص قدیمی تأثیر- استناد علمی نشان داده شده است. بار-ایلان[1] (۲۰۱۲) بین خواندن مقاله‌های مندلی و تعداد استنادهایی که مقاله‌ها دریافت کرده‌اند رابطه‌ای را نشان داد، درحالی که ایسنباچ (۲۰۱۱) نشان داد که اشاره‌های[2] توییتر برای استنادها شاخص اولیه محسوب می‌شود. ارزش حقیقی دگرسنجه‌ها انعکاس استنادها نیست بلکه جنبه متفاوتی از تأثیر کار با اعتباری مشابه است.

ارتباطات علمی برخط تنها به نسخه‌های الکترونیکی ارتباط سنتی محدود نمی‌شوند بلکه در قالب‌های زیادی ظاهر می‌شوند. ظاهراً هر فناوری جدیدی را از جهت سازگاری آن در کمک به بحث‌های علمی مورد امتحان قرار می‌دهند و فناوری‌های منتخب ظرفیت‌هایی را برای سنجه‌های جدید مطرح می‌کنند تا از مباحثه‌های علمی شناخت های جدیدی بدست آید. این سنجه‌ها می‌توانند آگاهی‌هایی از ارتباطات غیر رسمی در داخل یا خارج از جامعه علمی ارائه کنند:

- ویرایش‌های ویکی‌پدیا[3] می‌تواند برای تلاش یک پژوهشگر پزشکی در خصوص بهبود اطلاعات بهداشت عمومی شاخصی ارائه کند.
- تعداد دفعات بازدید در یک کانال یوتیوب[4] می‌تواند شاخصی برای مشارکت عمومی یک دانشمند در بر داشته باشد.
- اعتبار بدست آمده در یک وبگاه پرسش و پاسخ مانند استک آورفلو[5] یا وبگاه‌هایی که طراحی مشابه دارند در شبکه های تبادل قفسه کتاب[6] ممکن است سطح بالایی از عملیات را در جامعه پژوهشی نشان دهد.

هرچند مقاله پژوهشی هم چنان قالبی ارزشمندی برای به اشتراک‌گذاری یافته‌های پژوهشی شمرده می‌شود، در ارتباط با اینکه پژوهش بتواند به شکلی روشن در ۲۰ صفحه جا داده شود باید توافق های زیادی انجام شود. پژوهش علمی ندرتاً از

1. Bar-Ilan
2. mentions
3. Wikipedia
4. YouTube
5. Stack Overflow
6. Stack Exchange Network (http://stackexchange.com)

سازوکارهای مقاله‌های دانشگاهی پیروی می‌کند و روش یا ابزار استفاده شده به دنبال کاستی ابزارها و روش‌های دیگر اتخاذ می‌شود. تمامی دانش‌ها را نمی‌توان به سادگی در قالب کلمه‌ها آورد؛ همانگونه که پولانی[1] (۱۹۶۶، ۴) بیان کرده است «ما می‌توانیم بیش از آنچه می‌گوییم بدانیم» و او عبارت دانش تلویحی[2] را برای اشاره به دانش غیر قابل بیان ابداع کرد. مجموعه داده‌های عظیمی که اگر منتشر شوند بدون شک تا میلیون‌ها صفحه‌ای که با پژوهشگر حفظ شده، خواهند بود، درحالی که ممکن است برنامه‌های رایانه‌ای پیچیده‌ای که این داده‌ها را آزمایش می‌کنند وقتی که به صورت متن درآید معنی خود را از دست بدهند.

وب امکان رویکردهای جدیدی را در پژوهش و نشر مانند رویکردهای بازتری برای علم فراهم می‌کند. علم منابع باز[3] رویکرد منابع باز را در جهت توسعه نرم‌افزار، با گردآوری بسیاری از پژوهشگران بر روی پروژه‌های که شاید به شکل دیگری روی آنها سرمایه‌گذاری نشده است دنبال می‌کند. برای نمونه، جهش سیناپسی[4] پژوهشگران علوم پزشکی را برای بررسی بیماری‌هایی چون مالاریا[5] و شیستوزومیازیس[6] گرد هم می‌آورد

(کپلر[7] و همکارانش، ۲۰۰۶). علم دفتر باز[8] یک رویکرد شخص محور است که به منظور فراهم‌آوری علم و «سازماندهی تولیدات علمی بر مبنای نمایش همگانی دستآوردها و شکست‌ها، و داده‌ها و روندهای مرتبط آنها است به گونه‌ای که آشکارا برای پیشرفت آتی علم از طریق حل و تمرکز روی مشکلات خاص تحلیل و بحث شوند» (وارا[9]، ۲۰۰۹).

علم دفتر باز برای به اشتراک‌گذاری انواع یافته‌ها و داده‌هایی که به صورت سنتی نتوانسته‌اند به مجلات راه یابند راهی را نشان می‌دهد، هرچند که خود صنعت انتشارات نیز تغییر کرده است. مجلاتی ویدئو محور در راستای تلاش جهت ثبت و ضبط اطلاعاتی که هنگام تبدیل آنها به متن گم شده راه‌اندازی کرده‌اند (مانند *مجله*

1. Polanyi
2. tacit knowledge
3. open source
4. The Synaptic Leap (www.thesynapticleap.org)
5. Malaria
6. schistosomiasis
7. Kepler
8. open notebook
9. Vara

تجربه‌های بصری[1]. ناشران و کتابداران بیش از پیش به انتشارات غنی شده که در آنها مقالات نشریات هسته به منابع مرتبط پیوند داده شده‌اند، علاقمند می‌شوند. این منابع می‌توانند کد رایانه‌ای، داده، مدل‌ها یا سنجه‌های پس از نشر باشند. بعلاوه، امکان جابجایی یک ایده پایانی، تکمیل مقاله تمام شده توسط یک نویسنده نیز وجود دارد و پژوهش علمی می‌تواند با همکاری و کنترل نسخه‌هایی که جامعه برنامه‌نویسی آنها را تأیید کرده است چیزهای زیادی را بدست آورد (هریناسکیوویکز[2]، ۲۰۱۳). مهم‌ترین واقعیت این است که جامعه دانشگاهی و عاملان بودجه‌بندی پژوهش انواع جدید خروجی را بیش از پیش قانونی می‌دانند. در *مجله علوم انسانی دیجیتال*[3] ارسال به شیوه سنتی لازم نبوده و آثاری را شناسایی می‌کند که در وب باز منتشر شده‌اند، درحالی‌که در ژانویه ۲۰۱۳ پیوور[4] خاطر نشان کرد که مؤسسه علوم ملی ایالات متحده[5] درحال حاضر می‌خواهد که پژوهش‌گران «تولیدات» پژوهشی خود را به‌جای «انتشارات» ثبت کنند.

اگر چنین رویکردهای نوآورانه‌ای برای علم مورد حمایت قرار گیرند از اهمیت برخوردارند زیرا از محدودیت‌های شاخص‌های تک بعدی مانند شاخص اچ رها شده و امکان ارائه مجموعه‌ای از شاخص‌ها را برای تأثیر کلی‌تری از خروجی یک پژوهشگر فراهم می‌کنیم. مقایسه کتاب‌سنجی وب با کتاب‌سنجی گریزناپذیر است، زیرا در خصوص توان سنجه‌ای بر مبنای داده‌های به راحتی قابل دستکاری می‌شوند پرسش‌هایی پیش می‌آید. همانگونه که با نفوذ گوگل می‌توان دید هر صنعت نوظهوری سبب می‌شود که سازمان‌ها تأثیر برخط خود را به هر وسیله درست (به نام بهینه‌سازی موتور جستجوی کلاه سفید[6]) یا نادرست (بهینه‌سازی موتور جستجو کلاه سیاه[7]) بنا کنند. از آنجا که دوستان و همکاران بیش از حد به یکدیگر استناد می‌کنند غالباً سؤالاتی برای اعتمادپذیری استنادها پیش می‌آید. در این صورت می‌توان به ساخت آسان‌تر پیوندهای وبی (یا سایر شاخص‌های تأثیر) اندکی شک کرد. از

1. Journal of Visualized Experiments (www.jove.com)
2. Hrynaszkiewicz
3. Journal of Digital Humanities(http://journalofdigitalhumanities.org)
4. Piwowar
5. US National Science Foundation
6. white hat
7. black hat

آنجایی که کاربران تلاش دارند که با سیستم بازی کنند، فناوری‌های جدیدی نیز برای شناسایی ناهنجاری‌ها و سوءاستفاده‌های بالقوه ظهور خواهند کرد. حتی امکان دارد که کتابداران در ممیزی داخلی هر سنجه وبی نیز نقشی ایفا نمایند.

سنجه‌های وبی ارزیابانه صرفاً برای افزایش کیفیت پژوهش نیستند بلکه می‌توانند برای شناسایی منابع مفید مورد استفاده قرار گیرند. اگرچه ویراستاران مجلات و ناشران ممکن است از روش محاسبه جدول‌های رتبه‌بندی تأثیر مجله امتناع ورزند، اما می‌توانند وسیله مفیدی را برای کمک به پژوهشگران عرضه کنند تا مجلات هسته یک حوزه، یا مقاله‌های پراستنادی که شاید لازم است از آنها مطلع باشند، را شناسایی کنند. برای آینده علم مهم است که چنین ابزارهایی بسیار بازتر باشند.

در دوران کنونی جستجو بخشی لاینفک از زندگی مردم است و هر سؤالی ارزش جستجو در گوگل را داشته و خسته‌کننده نیست. برای اکثریت مردم موتور جستجو یک جعبه سیاه است، و درباره نحوه رتبه‌بندی نتایج یا نتایجی که در صفحات بعدی نتایج نمی‌بینند کم‌تر می‌کنند. درجستجوی مقاله‌های پژوهشی با موضوعی خاص در گوگل اسکالر یا جستجوی دانشگاهی مایکروسافت[1] نظام‌های رتبه‌بندی با داشتن تعداد استنادهایی که تأثیر معنی‌داری در رتبه‌بندی مقاله‌ها دارند واضح‌تر از موتورهای جستجو قدیمی به نظر می‌رسد، اما ارتباط و استنادها متفاوت هستند و الگوریتم‌های بنیادین می‌تواند تأثیر بالقوه‌ای بر مسیرپژوهش داشته باشند. بخصوص با توجه به اینکه خدمات جستجوی دانشگاهی دنباله‌روی خدماتی همچون آمازون[2] هستند که چنین می‌گویند «مشتریانی که این جنس را خریدند این جنس را هم خریدند...» و برای شروع پیشنهاد مطالعه مقاله‌های دانشگاهی از پایگاه اطلاعاتی آنها استفاده می‌کنند.

شکی نیست که موتورهای جستجوی دانشگاهی و رتبه‌بندی‌های مجله ابزارهای مفیدی هستند که به پژوهشگران کمک می‌کنند تا اطلاعات مورد نیاز خود را بیابند. با این وجود بسیار مهم است که پژوهشگران محدودیت‌های این ابزارها را نیز

1. (http://academic.research.microsoft.com) 2. Amazon

بشناسند. لازم نیست که کتابداران مسیر برخی دانشگاهیان را دنبال کرده و گوگل اسکالر را مهندسی معکوس کرده تا به شناسایی عواملی که به رتبه‌بندی گوگل اسکالر کمک می‌کنند برسند(برای نمونه، بیل و گیپ[1]، ۲۰۰۹) اما باید بدانند که ابزارها و رویکردهای دیگری هم هستند که ممکن است از یک حوزه پژوهشی چشم‌اندازی متفاوت بدهند، و باید به خاطر داشته باشند که هیچ کدام از ابزارها به همه چیز فکر نکرده‌اند.

وب‌سنجی ارزیابانه

سنجه‌های وبی علاوه بر ارائه اطلاعات درباره فرایند پژوهش علمی، می‌توانند برای فعالیت‌های جامعه و ماهیت انسان آگاهی‌های بیش‌تری نیز ارائه کنند و در این کتاب عبارت «وب‌سنجی» برای استفاده از سنجه‌های وبی برای اهداف پژوهشی استفاده شده است. گستره آگاهی‌هایی که از فعالیت‌های برخط کاربران بدست می‌آید، کاربردهایی را نشان می‌دهد که افراد برای آن‌ها از وب استفاده می‌کنند. همانگونه که در کتاب *کلیک*[2] بیل تنسر[3] (۲۰۰۹) نشان داده شده است که کلیدواژه‌ها[4] نه تنها می‌توانند برنده نمایش های استعدادیابی را پیش‌بینی کرده و این واقعیت را نشان دهند که بعد از عید سپاسگزاری اوج بازدیدها بطرف وبگاه‌های رژیم غذایی است، بلکه موتورهای جستجو می‌توانند از اعماق درونی افکار افراد نیز اطلاعاتی ارائه کنند. مردم با طرح سؤال‌هایی که دوست ندارند درباره آن‌ها با نزدیک‌ترین دوست و شریک خود گفتگو کنند، بیم و امیدهای خود را وارد موتورهای جستجو می‌کنند. بدین‌ترتیب از طریق جستجوی‌های برخط آنها می‌توانیم به نگرانی‌های سلامتی اشخاص، کاری که می‌خواهند انجام دهند، جاهایی که قصد رفتن دارند، و نحوه تغییرات در طی زمان و گزارش‌های رایج اخبار، پی ببریم. تنسر مدیر کل پژوهش جهانی در هیتوایز[5] که از ارائه‌کنندگان خدمات اینترنتی به مشتری‌هایی که برای کسب آگاهی از رفتار برخط کاربران وب هزینه پرداخت می‌کنند داده‌ها را جمع‌آوری

1. Beel & Gipp 2. Click 3. Bill Tancer 4. queries 5. Hitwise

می‌کند. وی به گستره وسیعی از داده هایی که اغلب کتابداران امکان دسترسی بـه آن را ندارند دسترسی دارد. هرچند ابزارهای فراوان دیگری هم برای کسب آگاهی از رفتـار کاربری در جامعه دانشگاهی وجود دارد که از کتابداران مـی‌خواهنـد ابزارهـایی پیـدا کنند که کاربرد رایگان داشته باشند.

در طول دو دهه گذشته طرز استفاده افراد از وب و نحوه تحلیل آنهـا از وب نیـز به همان اندازه تغییر کرده اسـت. مـا از عصـری کـه در آن وبگـاه‌هـا منحصـراً ایجـاد می‌شدند و اکثریت کـاربران مصـرف‌کننـده محتـوا بودنـد، بـه عصـر چیرگـی تعـداد کمی وبگاه تجاری با عده کثیری کاربر وبی که هـم تولیـدکننـده محتـوای وب و هـم مصرف‌کننده آن هسـتند حرکـت کـرده‌ایـم. پـژوهش وب‌سـنجی ارزیابانـه از تحلیـل صفحه‌های وب و شمارش پیوندهای وبی به سمت تحلیل طیف گسـترده‌ای از محتـوا و ارتباطات فراتر رفته است. با این وجود، صحبت ما درباره تحلیـل موجودیـت هـایی است که امکان دارد بنیاد تحقیقات ارتباطی و ارزیابانه را شکل دهند.

ضـریب تـأثیر وب اینگورسـون[1] (۱۹۹۸) از سـرجمع تعـداد پیونـدهای ورودی (ابرپیوندهای ورودی از خارج) و خود-پیوندها[2] (ابرپیوندهای درونی) بـا مجموعـه‌ای از صفحه‌های وبی با تعداد صفحه‌های درون مجموعـه محاسـبه شـده اسـت. ضـریب تأثیر وب می‌تواند با طراحی وبگاه و نظرات تعداد اندکی از وبگاه هـای بیرونـی تغییـر بسیار زیادی داشته باشد. برای نمونه، اگر محتوا یک وبگاه به‌جای ۱۰ صـفحه در ۱۰۰ صفحه توزیع شده باشد و با این وجود همان تعداد پیوند درونی دریافت کنـد در ایـن صورت ضریب تأثیر وبی می‌تواند با یک فاکتور ده تغییر کند. بـه صـورت دیگـر اگـر یک وبگاه بیرونی از سیستم مدیریت محتوایی که یک پیونـد را در تمـام صـفحه‌هـای وبگاه تکرار می‌کند استفاده کند (برای نمونه، پیوند وب‌نوشت‌های پیشـنهادی در نـوار حاشیه‌ای وب‌نوشت)، نظر شخصی یک نفر برای وارد کردن یک پیونـد ممکـن اسـت منجر به هزاران پیوند درونی شود. برای تقلیل تأثیر نظرات هر شخصی، آزمایش‌هـایی

1. Ingwersen 2. self-links

با ضریب تأثیر وب بسیار متفاوتی انجام شده است: مدل‌های اسنادی جایگزین پیوندهایی را شمرده‌اند که از صفحه‌های شخصی نبوده‌اند، بلکه از وبگاه‌ها یا وبگاه‌های راهنما بوده‌اند (تلوال، ۲۰۰۲)؛ و مخرج‌های جایگزین برای تعداد صفحه‌های یک وبگاه شامل تعداد کارمندان تمام‌وقت بوده‌اند (لی و همکارانش، ۲۰۰۳). مناسبت نوع خاصی از ضریب تأثیر وبی به عنوان مبنایی برای یک شاخص خاص (برای نمونه، ارتقای یک مؤسسه پژوهشی) عموماً با ترکیبی از همبستگی با داده‌های موجود (برای نمونه، رتبه‌بندی ارزیابی پژوهش یا نظام تعالی پژوهش‌های یک مؤسسه) و تحلیل محتوای دلایل ایجاد پیوندها اعتباریابی می‌شوند.

امروزه دیگر تحلیل‌های وب‌سنجی تنها محدود به صفحه‌های وب نبوده و تأثیر آن تنها مربوط به پیوندهای وبی نیست. واحدهای تحلیل اسامی و عبارت‌های سازمانی با سنجه‌هایی بر اساس حضور برخط آنها یا اینکه هر چند وقت یکبار جستجو می‌کنند را نیز پوشش می‌دهد. این واحدها شامل واحدهای مجزایی هستند که برخی وبگاه‌های شبکه های اجتماعی از پروفایل‌ها گرفته تا تصاویر، با سنجه‌هایی حول مسایلی همچون لایک، دنبال‌کننده‌ها و بازدیدکننده‌ها میزبان آن‌ها هستند. با این وجود، اصول بکار رفته همان اصولی هستند که برای تحلیل پیوندهای اولیه مورد استفاده قرار می‌گرفتند: پیوندها (یا سایر بدل هایی که در تأثیر منظور می‌شدند) باید مستقلاً و به تنهایی توسط انسان‌ها ایجاد شوند و از ارزش برابری برخوردار باشند. درحالی که اگر هر استنتاجی بدست آید باید بر مبنای همبستگی داده های موجود و تحلیل محتوای دلایلی ایجاد پیوند، باید برای اعتباربخشی یافته‌ها گام‌های مناسبی برداشته شود (تلوال، ۲۰۰۴a).

تحلیل وبی ارزیابانه

تحلیل وبی به معنای کاربرد سنجه‌های وبی برای ارزیابی و ارتقاء خدمات بوده و دلیل اصلی اکثریت مردم از جمله کتابداران برای کاربرد سنجه‌های وبی است. با

وجود اختلافات در تمرکز و عمل، اما موارد مشترک بسیاری بین تحلیل وبی و سنجه‌های وب ارزیابانه است. تمرکز تحلیل‌های وبی در درجه نخست بر یک موجودیت خاص است، و نتایج بررسی آنها بیش تر از آنکه محکی برای صحت خودشان باشد برای عمل هستند.

بیش‌تر افراد به کتاب‌سنجی و وب‌سنجی برای اهداف پژوهشی علاقه‌ای ندارند، بلکه به تعداد اشخاصی که از وبگاه‌هایشان بازدید می‌کنند، دنبال‌کننده‌های آنها در توییتر و دوستان آنها در فیسبوک علاقه‌مند هستند. حتی ممکن است این بحث پیش بیاید که برخی گروه‌ها به شدت به چنین سنجه‌هایی علاقه مندند. در تحقیقی درباره استفاده گروهی از دانشجوهای دانشگاه از فیسبوک، بورنو و بارخوس[1] (۲۰۱۱) دریافتند که دانش‌آموزان بطور متوسط ۸۴۱ دوست در فیسبوک داشتند. کم ترین دوست را شخصی با ۱۸۷ دوست و بیش‌ترین را شخصی با ۲۹۳۹ دوست بود. این تعداد نه تنها بسیار بیش‌تر از تعداد متوسط دوستان فیسبوکی معادل ۱۳۰ دوست است بلکه بسیار بیش‌تر از تعداد اشخاصی است که می‌توان با آن‌ها رابطه اجتماعی پایداری برقرار نمود که تعداد دانبار[2] نامیده می‌شوند و معمولاً حدود ۱۵۰ نفر در نظر گرفته می‌شود (دانبار، ۲۰۱۱). در این جا می‌توان این بحث را پیش کشید که لایک‌های فیسبوکی برای لایک‌های دنیای واقعی خیلی مناسب نیستند، اما به عنوان شاخصی برای محبوبیت و جامعه‌پذیری استفاده می‌شوند. تشخیص مواردی که ارزش به حساب آوردن را دارند و آن‌هایی که ارزش به حساب آوردن را ندارند، برای اشخاصی که از تحلیل وبی مثل همه سنجه‌های وبی استفاده می‌کنند چالش برانگیز است.

از آنجایی که کتابداران از ابزارها و خدمات برخط در مقیاس گسترده‌تری استفاده می‌کنند، طیف بسیار وسیعی از سنجه‌ها هستند که مد نظر قرار می‌گیرند. این طیف می‌تواند تعداد بازدیدها، علایق یا اظهار نظرهایی که بخش خاصی از محتوا در وبگاه‌های شبکه‌های اجتماعی مانند فلیکر دریافت کرده است یا اینکه دریافت چه تعداد بازدیدکننده یک صفحه یا وبگاه از طریق بازدید گوگل آنالیتیکز[3] بوده است، را پوشش دهد.

1. Bornoe and Barkhuus 2. Dunbar number 3. Google Analytics

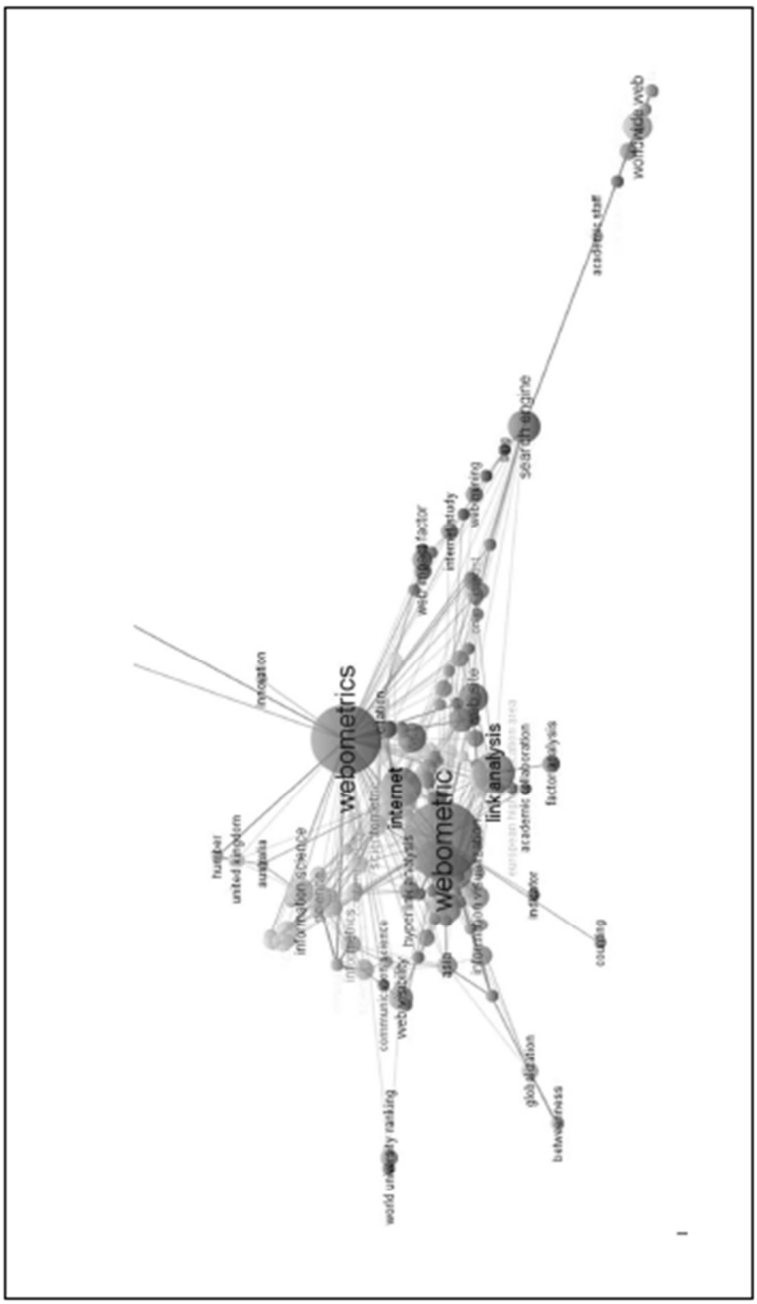
همانند تحلیل پیوندی، در تحلیل وبی لازم است کـه ایـن امـر را در نظـر داشـته باشیم که آنچه به حساب آورده می‌شود باید به تنهایی و مسـتقلاً ایجـاد شـده باشـند. کتابداران به سادگی می‌توانند با دنبال کردن رفتار تعداد زیادی ربات ایمیل ناخواسـته[1] دنبال‌کننده‌های توییتر خود را افزایش دهند. پیگیری خودکار حساب‌هـا بـه ایـن امیـد است که آن‌ها نیز متقابلاً شما را دنبال کنند. چنانچه آن‌ها بخواهند برای ایـن رویکـرد شناخته شده محتاط باشند، اغلب می‌توانند بعـد از چنـد روز دیگـر آن حسـاب‌هـا را دنبال نکنند. این کار هم باید بدون توصیه این رویکرد باشـد و کتابـداران بایـد رفتـار خود را تنظیم کنند.

سنجه‌های وب ارتباطی

عموماً سنجه‌های ارتباطی کم تر از سنجه‌های ارزیابانه مورد توجه هستند، بـا وجـودی که در کار کتابداران جایگـاه خـاص خـود را دارنـد. سـنجه‌هـای ارتبـاطی بـه دنبـال شناسایی بهترین یا مناسب‌ترین منابع نیستند بلکه یک بررسـی اجمـالی از رابطـه بـین کنشگران مختلف ارائه می‌کنند. ممکن است برای تهیه نقشه‌هـای علـم از سـنجه‌هـای وبـی ارتبـاطی اسـتفاده شـود. معمـولاً در ایـن مطالعـات از عناصـر موجـود در پایگاه‌داده‌های کتابشناختی در قالـب اغلـب از طریـق هـم–اسـتنادی[2]، تحلیـل لغـات مشترک[3] یا تحلیل هم– نویسندگی[4] استفاده می‌کنند.

فناوری رایانه‌ای انگیزه جدیدی برای کتاب‌سنجی‌های ارتباطی ایجاد کـرده اسـت همچنانکه که توسعه ابزارهای جدیدی را برای کاوش نقشـه‌هـا امکـان‌پـذیر مـی‌کنـد (نویـونس[5]، ۲۰۱۱). شـکل ۲–۲ تحلیـل لغـات مشـترکی را از کلیـدواژه‌هـایی نشـان می‌دهد که در نمایه مقالات وب‌سنجی وب آف ساینس اعمال شده و از طریق برنامـه رایانه‌ای VOSviewer به صورت رایگان در دسترس است[6].

1. spam bots
2. co-citation
3. co-word
4. co-author
5. Noyons
6. (www.vosviewer.com)



شکل ۲-۲: تحلیل لغات مشترک از کلیدواژه‌های مقالات وب‌سنجی

وب برای منابع کتابشناختی منابع اطلاعاتی جدیدی فراهم می‌کند. دیگر لازم نیست که تحلیل لغات مشترک منحصر به داده های داخل پایگاه اطلاعاتی کتاب‌سنجی باشد. اکنون پژوهشگران می‌توانند از سایر منابع برخطی که به انتشارات قدیمی اشاره می‌کنند استفاده کنند. این‌ها شامل استفاده از نشانه‌گذاری و مدیریت منابعی مانند وبگاه یولایک[1] و مندلی[2]، یا حتی عنوان موضوع‌ها و دسته‌بندی‌های موجود در فهرست‌های کتابخانه‌ای می‌شود.

بعلاوه، وب اطلاعاتی برای منابع غیر کتابشناختی فراهم می‌کند. از آنجا که آن‌ها به صورت برخط اتصال برقرار می‌کنند (استوآرت و تلوال، ۲۰۰۶) یا رابطه (یا کمبود آن را) بین رقیب های سیاسی در فضای وب نوشت ها، می‌توانند از رابطه بین سازمان‌ها اطلاعاتی ارائه نمایند (آدامیک و گلنس[3]، ۲۰۰۵).

در حالی‌که در برخی موارد اطلاعات این نقشه‌ها و نمودارها را می‌توان با تماشای آن‌ها جمع‌آوری کرد، برای تحلیل نقش گره‌های[4] منحصر بفرد موجود در داخل شبکه‌ها و ساختار شبکه های در کل، روش‌های آماری حوزه تحلیل شبکه های اجتماعی وجود دارد. درحالیکه در بسیاری از مطالعات وب‌سنجی اولیه تأثیر از طریق تعداد دریافت پیوندهای مستقیمی یک وبگاه اندازه‌گیری می‌شد که دسترسی به داده ها برای بیش از یک شبکه های بود اما معیارهای مرکزیّت دیگری را نیز در نظر گرفته‌اند، مانند بینیت[5] و نزدیکی[6] (برای نمونه، اورتگا[7] و آگویللو، ۲۰۰۸)، که موضوع مورد توجه این است که آیا یک گره در ژئودسیکس[8] (نزدیک‌ترین مسیر) بین گره‌های دیگر قرار دارد یا نه و نزدیکی گره به سایر گره‌ها چقدر است.

در برخی موارد اختلاف بین سنجه‌های وب ارتباطی و ارزیابانه بسیار کم و در حد یک نگرش است. یان، دینگ و ژو[9] (۲۰۱۰) در تحلیل شبکه‌های هم-نویسندگی[10] کتابداری و اطلاع‌رسانی در چین دریافت که ارزش مرکزیّت تحلیل

1. CiteULink (www.citeulike.org)
2. (www.mandeley.com)
3. Adamic & Glance
4. nodes
5. betweenness
6. closeness
7. Ortega
8. geodesics
9. Yan, Ding & Zhu
10. co-authorship

شبکه های اجتماعی در یک شبکه های با رتبه‌بندی‌های استناد همبستگی بسیاری دارد. از بعُد ارتباطی، مرکزیّت درباره ساختار شبکه های اطلاعات می‌دهد ولی از بعُد ارزیابانه امکان دارد مرکزیّت شاخصی برای تأثیر نویسنده در یک حوزه ارائه کند.

یک کتابدار ممکن است از چنین نقشه‌هایی استفاده کند تا بتواند برای یک پژوهشگر تازه وارد در یک حوزه دید کلی از پژوهش های مرتبط ایجاد کند یا به اشخاصی که درگیر سیاست هستند کمک کند تا زمینه‌های نو ظهور پژوهش در درون حوزه خاصی شناسایی شوند. با توجه به اینکه درباره فرایند علمی اطلاعات بیش تری به صورت برخط قابل دسترس است، دیگر لزومی ندارد که چنین بررسی‌های ارتباطی را محدود به بررسی‌هایی در انتشارات قدیمی کنیم.

اعتباریابی نتایج

این کتاب به اجمال از توصیه تلوال (۲۰۰۴a) برای تحلیل پیوندی پیروی می‌کند: جایی که نتایج احتمالی بررسی‌های سنجه‌های وبی ارزیابانه و ارتباطی باید از طریق ترکیبی از همبستگی و دسته‌بندی اعتباریابی شوند. گستره چنین اعتبارهایی به مقدار زیاد به هدف سنجه‌های وبی و هر اقدامی که می‌تواند بر مبنای آن‌ها باشد بستگی دارد.

همبستگی

یکی از رایج‌ترین آزمایش‌های آماری که در وب‌سنجی استفاده می‌شود بررسی این است که آیا مجموعه‌ای از نتایج با یک معیار بیرونی استنباطی ارزیابی می‌شود همبستگی[1] دارد یا خیر. برای نمونه، پژوهشگری ممکن است برای مجموعه‌ای از شرکت‌ها، بررسی کند که آیا تعداد پیوندهای درونی دریافت شده توسط پایگاه وب شرکت با برخی سایر شاخص‌های موفقیت مانند حجم معاملات[2] آن‌ها همبستگی دارد یا خیر، یا اینکه آیا پیوندهای درونی گروهی از وبگاه ها دانشگاه‌ها با برتری پژوهشی دانشگاهیان همبستگی دارد یا نه. در کل از آنجایی که توزیع پیوندها، رفت و آمد و

1. Correlation 2. Turnover

سایر سنجه‌های وبی اغلب در وب به شدت مخدوش می‌شوند، چنین همبستگی‌هایی به‌جای رتبه پیرسون[1]، با محاسبه ضریب رتبه اسپیرمن[2] آزموده می‌شوند.

نکته مهم قابل توجه دیگر اینکه در هنگام بررسی اینکه آیا یک مجموعه خاصی از داده ها با چیز دیگری همبستگی دارد یا نه، مقدار مطلق همبستگی‌ها است که می‌تواند مورد بررسی قرار گیرد. وب امکان دسترسی به مجموعه متنابهی از داده ها را در هر موضوع قابل تصور فراهم می‌کند به گونه‌ای که که توانایی احتمالی نتایج نادرست از داده‌ها را افزایش می‌دهد، بویژه از این نظر که توانایی تحلیل و آزمودن خودکار به این معنی است که همبستگی می‌تواند تنها با جستجوی زیاد دیده شود. برای نمونه، بسیاری از مطالعات اگر تنها ۱٪ یا ۵٪ احتمال وجود همبستگی به صورت تصادفی موجود باشد همبستگی معنی‌داری آماری را نشان می‌دهند. اگر تنها یک مجموعه از داده‌ها مورد بررسی قرار گرفته و از نظر آماری معنی‌دار باشد، در آن صورت پژوهشگر می‌تواند کمابیش به یافته‌های خود اعتماد داشته باشد. هرچند چنانچه ۱۰۰ مجموعه مختلف داده آزموده شود و در یکی از آنها در سطح آماری ۱٪ معنی‌دار باشد آنوقت پژوهشگر قطعاً به یافته‌هایش با اعتماد کم‌تری نگاه می‌کند. در این صورت، تصحیح بونفرونی[3] روشی است برای مقابله با افزایش احتمال تأثیر چندین مقایسه که در آن هر فرضیه ای در سطح ۱ معنی‌داری آماری تقسیم بر تعداد فرضیه های بررسی شده در نتیجه ای ضرب می‌شود که اگر تنها یک فرضیه آزمایش می‌شد. برای نمونه، اگر شخصی بخواهد از یافته‌ها در سطح ۱٪ معنی داری آماری مطمئن شود و در حال آزمودن ده فرضیه است، آنگاه هریک از این ۱۰ فرضیه باید برای معنی‌دار بودن در سطح ۰/۱٪ آزموده شوند. این موضوع موجب می‌شود تا یافته‌های معنی‌دار آماری بسختی شناسایی شوند، به همین دلیل است که انتخاب فرضیه‌هایی مورد بررسی برای همبستگی‌ها از ایجاد وب مهم‌تر است.

1. Pearson 2. Spearman 3. Bonferroni

همانگونه که اغلب خـاطر نشـان مـی‌شـود همبسـتگی و آزمـون علّـی[1] دو چیـز متفاوت هستند. تنها بخـاطر اینکـه تعـداد پیونـدهایی کـه مجموعـه‌ای از دانشـگاه‌هـا دریافت می‌کنند با برتری پژوهشی مؤسسه همبستگی دارد نمی‌توان نتیجه گرفـت کـه پیوندها بخاطر برتری پژوهش جای داده شده‌انـد. دلیـل آن مـی‌توانـد ایـن باشـد کـه پیوندها چیز دیگری را بازنمون می‌کنند که با برتری پژوهشی مؤسسـه نیـز همبسـتگی دارد، برای نمونه، اندازه مؤسسه یا اعتبار آموزشی آن. اشتباه گـرفتن یکـی بـا دیگـری می‌تواند به معنای این باشد که جریمه یا پاداش برخـی مؤسسـه‌هـا غیرمنصـفانه بـوده است. برای حمایت بیش تر از یک همبسـتگی، یـا بـرای بررسـی دلایـل مـواردی کـه همبستگی برای یک رخداد وبی قابل اجرا نیست می‌توان استفاده کرد.

دسته‌بندی

همانگونه که مکرراً در متون کتاب‌سنجی گفته شده است، استنادها بـه دلایـل مختلـف زیادی ایجاد شده و برای تحلیـل اسـتنادی طـرح‌هـای دسـته‌بنـدی مختلفـی پیشـنهاد شده‌اند (کرونین[2]، ۱۹۸۴). همین حالت نیز برای پیونـدهای وبـی، ارجاعـات وبـی یـا سایر رویدادهای وبی که ممکن است پایه و اساس بررسی‌های سـنجه وبـی را شـکل دهند صادق است. پیوندها در وبگاه ها قرار گرفته‌اند و سـازمان‌هـا و افـراد بـه دلایـل مختلف زیادی مورد اشاره قرار می‌گیرند اما همه آن‌ها قابل تعریـف نبـوده و شـاید از یافته‌های پژوهشگران حمایت نکنند. در واقع، به دلایل مختلف، محتـوای وب مربـوط به افراد بوده و بد تعریف شده‌اند و دلایل ذکر بسیار متفاوت تـر از دلایلـی اسـت کـه اسـتنادها در انتشـارات دانشـگاهی قـرار داده مـی‌شـوند. همینطـور دلایـل بازدیـد بازدیدکنندگان از یک وبگاه نیز ممکن است به طرز چشمگیری با هم متفاوت باشد.

همانگونه که مارک[3] (۲۰۱۱، ۶) اظهار کرده: «داده‌های خام تنهـا نیمـی از جریـان را فاش می‌کند». برای اینکه بتوان از داده ها نتایج معنی‌داری استخراج کـرد لازم اسـت از دلیل وجودی آن شناخت بیش تری در دست باشد. دلیل این موضوع می‌توانـد ایـن

1. causation 2. Cronin 3. Marek

باشد که تیم پروژه پژوهش می‌خواهد تأثیری را که پروژه آنها داشته را نشان دهد و تصمیم می‌گیرد از اشاره های وبی به عنوان یک شاخص استفاده کند. اگر اکثریت اشاره های وبی حول محور انتقادها از کارگزار تأمین‌کننده بودجه چنین پژوهشی باشد در آن صورت اشاره‌های وبی ضرورتاً شاخص خوبی از تأثیر پژوهش نیستند. از این رو لازم است که گروه پژوهشی نشان دهندکه بخش قابل توجهی از پیوندها به دلایل مثبت قرار داده شده‌اند. همینطور اگر یک مطالعه وب‌سنجی تلاش دارد پیوندهای وبی را برای ارزش پژوهش در یک مؤسسه شاخص به نمایش بگذارند، این مورد در صورتی با فشار بیش تری ایجاد می‌شود که بتوان نشان داد که بخش چشمگیری از پیوندها برای ارجاع به پژوهش‌هایی است که در حال اجرا هستند. همانگونه که استِرن (۲۰۱۰) در کار خود در حوزه سنجه‌های رسانه های اجتماعی گفته است «ارزش واقعی از دسته‌بندی بدست می‌آید» (۱۸۶). دسته‌بندی امکان ایجاد تمایز بین پیوندها یا اشاره های وبی که مثبت هستند و آن‌هایی که منفی هستند را فراهم می‌کند. آن‌هایی که برای ارجاع یک پژوهش دانشگاهی در نظر گرفته شده‌اند و آن‌هایی که برای برجسته کردن برنامه ای در یک اتحادیه دانشجویی در نظر گرفته شده‌اند نیز قابل تمایزند.

دسته‌بندی در پژوهش سنجه وبی می‌تواند هم از طریق تحلیل عاطفی یا از طریق محتوا بدست آید. تحلیل عاطفی، یا نظر کاوی، اغلب برای اشاره به تحلیل خودکار دیدگاه های مردم درباره موضوعات خاص، اشخاص، رخدادها، کالاها، خدمات و یا هرچیز دیگری که مردم می‌توانند نظراتی داشته باشند، استفاده می‌شود. این امر می‌تواند برای یک سند، یک جمله یا جنبه‌های خاصی از یک موضوع اعمال شود (فلدمن[1]، ۲۰۱۳). در طول دهه گذشته حجم وسیع داده‌های بی اساسی که به صورت برخط در دسترس هستند موجب شده که تحلیل عاطفی بسرعت مورد توجه قرار گرفته به طوری که افراد از طریق وبگاه‌های شبکه های اجتماعی و صفحات وب

1. Feldman

شخصی بیش از پیش می‌توانند دیدگاه های خود را درباره هر موضوع قابل تصوری بیان کنند(لیوو[1]، ۲۰۱۲). در مقایسه، تحلیل محتوا به متون محدود نمی‌شود، بلکه تحلیل نظام‌مند هر نوع ارتباطاتی (از جمله تصاویر و ویدئوها) است و طیف پرسش‌هایی را که ممکن است درباره یک موضوع پرسیده شوند فارغ از دیدگاه مثبت یا منفی (یا خنثی) گسترش می‌دهد. طیف گسترده‌تر محتوا و پرسش‌هایی که می‌تواند پرسیده شود به این دلیل اتفاق می‌افتد که تحلیل محتوا یک رویکرد انسان محوری برای طبقه‌بندی اسناد است.

یک روش به سادگی بهتر از روش دیگر نیست، اما هر دو تحلیل عاطفی و محتوا دارای مزایا و معایبی هستند، و انتخاب یکی از آنها بستگی به ویژگی‌های موضوع مورد بررسی دارد. دلیلی که ممکن است باعث ارجحیت یک روش به روش دیگر شود یا مربوط به ویژگی‌های محتوا است یا مربوط به ویژگی‌های پژوهش .

محتوا ممکن است به دلیل قالب محتوا، موضوع محتوا، اندازه نمونه یا ساختار منبع داده‌ها برای نوعی از تحلیل مناسب‌تر از یک نوع دیگر باشد. افراد در تعداد میلیونی برای نوشتن نقد و بررسی، به اشتراک‌گذاری دیدگاه ها، توییت کردن مشاهده‌ها، و بارگذاری تصاویر، ویدئوها و فایل‌های صوتی برخط می‌شوند. اگر شرکتی بخواهد بداند که در میان هزاران عکس موجود در فلیکر مثبت ارزیابی می‌شوند یا منفی، شاید بتوانند از اظهار نظرها یا برچسب‌هایی[2] که روی عکس‌ها گذاشته می‌شود اطلاعاتی بدست آورند، گرچه آن‌ها در درجه نخست مجبورند خود محتوا را تحلیل کنند. اگرچه مطالعات به صورت روزافزونی تحلیل‌های عاطفی خودکار شده‌ای را بدست می‌آورند که با تحلیل‌های عاطفی انسان یکسان باشند، کنایه[3] حوزه‌ای است که هم‌چنان به خوبی کار نمی‌کند (فلدمن[4]، ۲۰۱۳). در این مورد اگرچه ممکن است انسان‌ها بهتر کار کنند، با این وجود تحلیل عاطفی به دلیل نارسایی متن به عنوان یک رسانه های دشوار است. یک مزیت روشن تحلیل‌های

1. Lio 2. tags 3. sarcasm 4. Feldman

عاطفی این است که از فرایندی خودکار برخوردار است و یک برنامه تحلیل عاطفی در طی چند دقیقه می‌تواند به دسته‌بندی هزاران متن را که می‌توانستند وقت چند صد و چند هزار انسان را بگیرند، بپردازند. برای نمونه، در هنگام تحلیل توییت‌های ارسال شده در طول H1N1 همه‌گیر، اگرچه چو[1] و ایسنباچ (۲۰۱۰) بیش از ۲ میلیون توییت بدست آورند اما تحلیل محتوا دستی تنها شامل ۵۳۹۵ توییت شده بود. به نحو دیگر، اگر سازمانی بخواهد هر وقت که در معرض دیدگاه‌های منفی قرار می‌گیرد پاسخ دهد، اگر این‌ها بخش کوچکی از تعداد کل نظرات باشند آنوقت شاید برای شناسایی اظهار نظرها به یک حدی از تحلیل‌های خودکار نیاز پیدا کند. به همین ترتیب، اگر تنها تعداد اندکی از بخش‌ها نیاز به تحلیل داشته باشند، تحلیل کار به صورت دستی می‌تواند سریع‌تر از انجام یک تحلیل عاطفی باشد، بخصوص اگر نیاز به ایجاد دامنه واژگانی خاصی باشد.

ساختار نوع منبع اطلاعات نیز مهم است. توجه به تحلیل عاطفی تنها به دلیل رشد وب ایجاد نشده است بلکه به خاطر گسترش محتوای ساختار یافته‌ای است که فناوری‌های رسانه های اجتماعی فراهم کرده‌اند. برای نمونه، توییتر نه تنها امکان دسترسی به حجم عظیمی از داده را امکانپذیر می‌سازد بلکه امکان دسترسی به محتوایی را می‌دهد که به روشنی در واحدهای قابل درک پدیدار می‌شوند. در مقایسه، صفحات وب قدیمی به هم ریخته هستند و شناسایی قسمت‌های مرتبطی از صفحه را برای تحلیل نمی‌توان به سادگی خودکارسازی کرد. همچنین، در بسیاری از موارد شخصی که تحلیل را انجام می‌دهد به احساس مجموعه خاصی از داده توجه نداشته بلکه به برخی سؤال‌های دیگر توجه نماید. برای نمونه، در فرایند تحلیل وبی، پژوهشگران ممکن است بخواهند بدانند که آیا عبارات جستجوی مورد استفاده برای ورود به وبگاه آن‌ها با مأموریت اصلی وبگاه ارتباط داشته‌اند یا نه. آنچه را که نمی‌شود با بررسی تعیین کرد این است که عبارت‌های جستجو مثبت هستند یا منفی.

1. Chew

نتیجه‌گیری

در این فصل هم توانایی‌ها و هم محدودیت‌های سنجه‌های وبی مورد تأکید قرار گرفتند. آن‌ها قادرند اطلاعات جدیدی در خصوص فرایند علمی و سبک‌های نشر سنتی ارائه کنند و به عنوان یک ابزار پژوهشی ارزشمند مطرح شوند که به پژوهشگران کمک می‌کند تا از رفتار افراد و گروه‌ها شناختی بدست بیاورند. بعلاوه، برای افراد و سازمان‌ها روشی را عرضه می‌کند تا اعتبار و رفتار برخط خود را تحلیل کنند. هرچند محدودیت‌های شناخته شده‌ای نیز در ابزارها، روش‌ها و ماهیت خود وب وجود دارد.

حتی در جاهایی که یافته‌ها از نظر آماری معنی‌دار هستند لازم است که نتایج با احتیاط مطرح شوند. همانگونه که اغلب بیان می‌شود همبستگی با آزمون علَی یکی نیست و به دلیل اینکه داده‌ها به سهولت به صورت برخط ایجاد می‌شوند و به دلیل محدودیت‌های ابزارهای موجود، یافته‌ها باید «نشان‌دهنده و نه تعیین‌کننده[1]» باشند (تلوال، ۲۰۰۹، ۱۴).

با این وجود، همانگونه که می‌توان از این فصل نتیجه‌گیری کرد، هر دو سنجه‌های ارتباطی و ارزیابانه در ابعاد وسیعی مورد استفاده کتابداران هستند. برخی از کاربردهای ارزشمندتر آنها عبارتند از:

- ارزیابی تأثیر محتوای کتابخانه.
- ارزیابی تأثیر محتوای کاربر کتابخانه.
- شناسایی منابع و پژوهش‌های مهم درون یک حوزه.
- حمایت از پژوهشگران با بررسی‌های وب‌سنجی.
- ارائه دید کلی در خصوص حوزه پژوهشی به مشتریان.

1. indicative rather than definitive

۳

ابزارهای گردآوری داده‌ها

مقدمه

همانند کتابخانه در قوانین پنجگانه علوم کتابداری رانگاناتان (۱۹۳۱)، وب یک عضو در حال رشد بوده و یکی از موضوعات مورد بحث در سراسر این کتاب ماهیت متغیر آن است. همانگونه که در این فصل نشان داده می‌شود، فناوری‌های مورد استفاده، مانند وبگاه‌ها و خدمات موجود برای بررسی وب در حال تغییرند. اگر کتابداران در نظر دارند که از سنجه‌های وبی چه برای اهداف ارتباطی و چه برای اهداف ارزیابانه استفاده کنند، باید متوجه ماهیت در حال تغییر چنین وبگاه‌ها و خدماتی باشند.

در طی ۲۰ سال گذشته وب به بخشی همیشه حاضر در دنیای مدرن تبدیل شده و توسعه ابزارها و فناوری‌های وابسته به آن تأثیر فراوانی بر ماهیت بررسی‌های سنجه وبی داشته است. در طی این بازه زمانی، موتورهای جستجو بخش زیادی از وب را نمایه‌سازی کرده، و قابلیت جستجوی پیشرفته‌ای را عرضه می‌کنند که از رهگذر آن بررسی‌های پیچیده ای از محتوای وبی و شبکه‌های ابرپیوندی میسّر شده است. انقلاب وب ۲/۰ شاهد بود که موتورهای جستجو و خدمات رسانه های اجتماعی رابط‌های برنامه‌نویسی برنامه کاربردی[1] را فراهم کردند به گونه‌ای که

1. Application Programming Interfaces (API)

طراحان بتوانند به صورت خودکار با محتوای یک پایگاه وب تعامل داشته و همچنین امکان انجام بررسی‌هایی با حجم بسیار بالا فراهم شود. راه اندازی استانداردهای وب عاطفی نوید وبی را می‌دهد که بسیاری از کارهای محاسباتی روزانه افراد به صورت خودکار با برنامه‌های رایانه ای قابل انجام است (برنرز-لی[1]، هندلر و لاسیلا[2]، ۲۰۰۱) و توسعه کوکی‌ها[3] به سازمان‌ها امکان پیگیری کاربران در وبگاه‌ها را میسّر می‌کنند (تورو[4]، ۲۰۱۱). البته، پیشرفت فنی هموار و بدون وقفه نبوده است. موتور جستجو و قابلیت رابط‌های برنامه‌ریزی کاربردی در برخی موارد نادیده گرفته شده، و وب عاطفی با سرعت مورد انتظار رشد نکرده است و کمسیون اروپا[5] محدودیت‌هایی را برای کوکی‌های کاربران اعمال کرده است (دفتر کمیسیون اطلاعات[6]، ۲۰۱۲). در نتیجه تمام این تغییرها بر انواع سنجه‌های وبی تأثیر داشته است: آن سنجه‌هایی که دیروز در دسترس بوده‌اند ممکن است فردا در دسترس نباشند ولی سنجه‌های جدیدی ظهور کنند. درک نحوه تغییر ابزارها در گذشته می‌تواند به فهم ما در نحوه تغییرات آینده و شناسایی فرصت‌های احتمالی کمک کند.

جز در مواردی که پژوهشگران به صورت دستی از وب اطلاعات گردآوری می‌کنند، به ابزارهایی نیازمندند که بتواند از طرف آنها اطلاعات را جمع‌آوری کند، و در فاصله کم‌تر از دو دهه، از اولین مقاله‌ای که می‌تواند به عنوان وب‌سنجی (لارسون[7]، ۱۹۹۶) مدّ نظر گرفته شود، چهار بازه مختلف برای ابزارهای پژوهش وب‌سنجی وجود داشته است. پس از معرفی اولین واژگان مورد نیاز برای توصیف اسناد وبی و پیوندهای بین آن‌ها، این فصل در خصوص ماهیت در حال تغییر ابزارهایی که اساس پژوهش‌های وب‌سنجی را در جامعه علم اطلاعات شکل می‌دهند به بحث می‌پردازد. وب‌سنجی به بحث درباره ابزارهای خاص نمی‌پردازد

1. Berners-Lee
2. Hendler & Lassila
3. cookies
4. Turow
5. European Commission
6. Information Commission Office
7. Larson

بلکه به انگاره‌هایی[1] می‌پردازد که پژوهشگران از طریق آن‌ها وب را درک می‌کنند. اگرچه تأکید این کتاب تنها بر وب‌سنجی نبوده و بر کتاب‌سنجی وبی، علم‌سنجی وبی و تحلیل وبی نیز تمرکز دارد، بسیاری از ابزارهای مورد بحث در تمامی این شرایط استفاده می‌شوند. شرح حال جامعی از تمام ابزارهای وابسته به سنجه‌های وبی بیش‌تر کتاب را اشغال خواهند کرد.

آناتومی یک نشانی وبی، پیوندهای وبی و ساختار وب

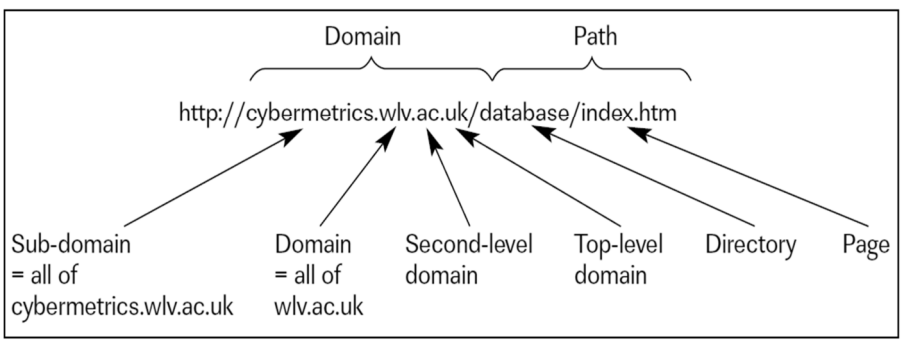
نیازی نیست پدیده‌هایی که پایه و اساس پژوهش‌های سنجه وبی را شکل می‌دهند ابرپیوند صریح و روشنی از اسناد باشند بلکه می‌توانند بر اساس ارتباطاتی باشند که از صراحت کم‌تری برخوردارند مانند اصطلاح‌ها، عبارت‌ها، زبان‌ها یا انواع اسناد مشترک. با این وجود، بیش‌تر بحث‌ها درباره ماهیت در حال تغییر ابزارهای تحلیل وب‌سنجی و از بین رفتن قابلیت‌ها به شناخت ساختار نشانی وبی و ساختار پیوندی خود وب ارتباط دارند.

نشانی وبی یکی از واحدهای پایه‌ای وبی و مطالعات وب‌سنجی است، و هنگام بحث در خصوص صفحه‌های وبی، زیر دامنه‌ها یا وبگاه‌ها، مفاهیم به طور کلی در آدرس‌های اینترنتی آنها عملیات‌سازی می‌شوند. این تنها راهی نیست که بتوان در آن مفاهیم را عملیات‌سازی کرد بلکه صفحه‌های وبی را می‌توان با توجه به محتوای صفحه‌ها یا پیوند درونی[2] بین آنها به صورت اسناد وبی یا وبگاه‌ها خوشه‌بندی کرد (کوتی، آگویللو و آرویو[3]، ۲۰۰۶). گرچه نشانی وبی می‌تواند مشهودترین و ساده‌ترین راه برای عملیاتی‌سازی این مفاهیم در نظر گرفته شود و پایه و اساس اغلب پژوهش‌های وب‌سنجی را شکل دهد.

ترکیب نشانی وبی خودش می‌تواند از چند بخش مختلف باشد: یک تفاهم‌نامه[4]، یک نام کاربری، یک رمز عبور، یک نام دامنه، یک درگاه[5]، یک کلیدواژه جستجو[6] و

1. paradigm
2. interlinking
3. Cothey, Aguillo and Arroyo
4. Protocol
5. port
6. query

یک پشتیبان[1]. با وجود اینکه نشانی‌های وبی گاهی شامل نام کاربری و رمز عبور هستند، و شاید لازم باشد که برای نحوه پرداختن محتوای تولیدی پویا (بخش کلیدواژه‌ای نشانی وبی) تصمیم‌هایی اتخاذ شود، مهم‌ترین واژگان آدرس‌های اینترنتی استفاده شده در سراسر این کتاب در شکل ۳-۱ نشان داده شده است.



شکل ۳-۱: نمونه‌ای از یک نشانی وبی علامت‌گذاری شده با واژگان تحلیل پیوندی

ساختارنشانی‌های وبی به پژوهش‌های وب‌سنجی امکان استفاده از پیوند درونی بین اسناد وب را می‌دهد که ممکن است در دامنه‌های سطح بالا، دامنه‌های سطح دوم، دامنه‌ها، زیر دامنه‌ها و همچنین راهنماها و صفحه‌ها عملیاتی‌سازی می‌شوند. دامنه‌های سطح بالا[2] به صورت سنتی محدود به تعداد اندکی دامنه سطح بالاهای عمومی[3] (برای نمونه، .com, .org, .gov و .edu) و دامنه سطح بالاهای کد کشور[4] هستند. همچنین برخی دامنه سطح بالاهای کد کشور نیز بیش‌تر به نام‌های دامنه سطح دوم شکسته شده‌اند (برای نمونه، .ac.uk و .co.uk)، اگرچه تمام کشورها از سیستم نام‌های دامنه سطح دوم استفاده نمی‌کنند (برای نمونه، .fr).

این ساختار در تحلیل پیوند درونی بین انواع مؤسسه‌ها (برای نمونه، کاربرد دامنه‌های سطح بالای عمومی) و بین کشورها (برای نمونه، کاربرد دامنه های سطح بالای کد کشورها) استفاده می‌شود. با توجه به اینکه در سال ۲۰۱۳ سازمان اینترنتی

1. anchor
2. top-level domains (TLD)
3. generic gTDL
4. country-code TDLcc

نام‌ها و شماره‌های تخصیصی[1] از سازمان‌ها خواسته است که از دامنه‌های سطح بالای جدید استفاده کنند، انتظار می‌رود تعداد دامنه‌های سطح بالای عمومی سیر صعودی داشته باشد. به زودی آدرس‌های اینترنتی مانند http://news.bbc یا http://pubs.london خواهیم داشت. بیش‌تر پژوهش‌های درون- اسنادی[2] در پیوندهای درونی بین دامنه‌های دانشگاه‌ها یا سازمان‌های تجاری در سطح دامنه اتفاق افتاده‌اند هرچند در برخی موارد امکان دارد که زیر دامنه، راهنما یا یک صفحه وب جهت عملیاتی‌سازی اسناد وب مناسب‌تر باشند. برای نمونه، وبلاگ‌نویسان در وبگاه بلاگر[3] برای وب نوشت‌های مختلف از زیر دامنه های گوناگونی استفاده می‌کنند (برای نمونه، http://googleblog.blogsport.co.uk) و از آنجایی که این‌ها غالباً دارای نویسنده و اهداف مشخصی هستند این طور فکر بوجود می‌آید که زیر دامنه‌های مختلف همانند وبگاه‌های مختلف عمل می‌کنند.

در پژوهش‌های وب‌سنجی نه تنها اسناد بلکه اتصال درون سندی بین اسناد وبی نیز بررسی می‌شوند. حتی هنگامی که ما خود را به جای هر نوع اتصال درون سندی به گمارش ابرپیوندها محدود می‌کنیم، از واژگان اتصالی متفاوتی استفاده کرده‌ایم. درحالیک-ه یک ابرپیوند ایجاد یک پیوند درونی زبان نشانه‌گذاری ابرمتنی[4] یک سند را توصیف می‌کند، صفحه مورد نظر از آن پیوند با عناوین متفاوتی به عنوان حاوی یک استناد ابرمتنی (چن و همکارانش[5]، ۱۹۹۸)، یک نقل قول[6] (روسیو[7]، ۱۹۹۷) یا یک پیوند به عقب[8] (هارتر و فورد[9]، ۲۰۰۰) یاد می‌کندکه در جامعه بهینه‌سازی موتور جستجو هم‌چنان رایج است. این کتاب از واژگان بیجورنبورن و اینگورسن (۲۰۰۴) پیروی می‌کند:

- یک *پیوند بیرونی*[10] جایی رخ می‌دهد که یک پیوند در داخل یک سند وبی به یک نشانی وبی خارج از آن سند وبی اشاره کند.

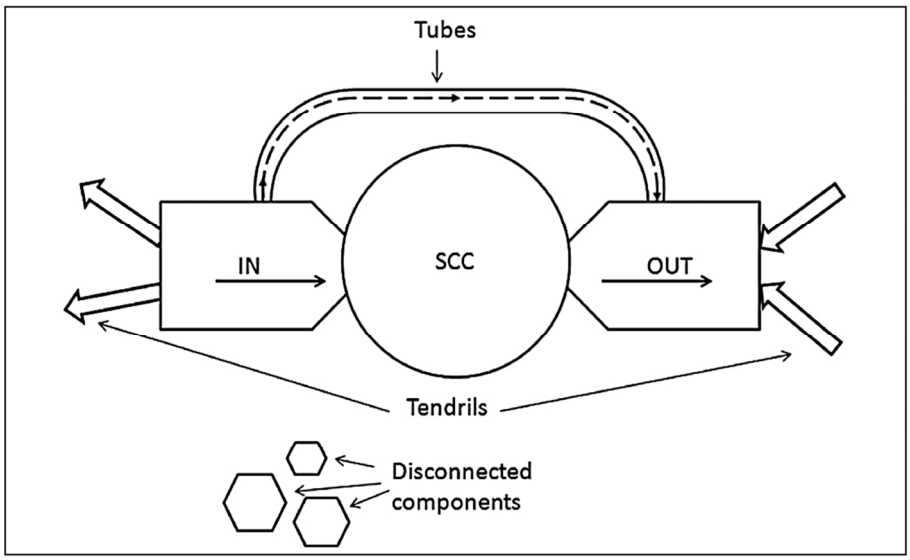
1. Internet Corporation for Assigned Names and Numbers (ICANN)
2. inter-document
3. Blogger
4. Hypertext Markup Language (HTML)
5. Chen et al.
6. Sitation
7. Rousseau
8. backlink
9. Harter & Ford
10. *outlink*

- یک *پیوند درونی*[1] جایی است که در آن یک ابرپیوند از خارج از یک سند وبی به جایی در داخل سند وبی اشاره کند.
- یک *خود-پیوند*[2] به یک پیوند در داخل یک سند وبی که به مکانی در داخل سند وبی اشاره می‌کند اطلاق می‌شود. برای مثال، اگر اسناد وبی در سطح دامنه عملیاتی می‌شود، یک صفحه در داخل پیوند دامنه ای به صفحه دیگر دامنه، به عنوان خود پیوند در نظر گرفته می‌شود.
- یک *پیوند-متقابل*[3] جایی اتفاق می‌افتد که دو سند وبی با یکدیگر پیوند برقرار کنند، هرچند پیوند همان دو صفحه به یکدیگر الزامی نیست.

برای استفاده ازنشانی‌های وبی علاوه بر آگاهی از نحوه عملیاتی‌سازی اسناد وبی و درک واژگان پیوندهای درون سندی مختلف، قرار دادن پژوهش‌های وب‌سنجی در ساختار وبی به صورت کامل الزامی است. شکل ۳-۲ مدل اتصال وب برودر و دیگران[4] (۲۰۰۰) را نمایش می‌دهد. این مدل علاوه بر اینکه کمک می‌کند تا متوجه نحوه اتصال وبگاه‌ها به یکدیگر شویم، به فهم پیامدهای ماهیت یک طرفه پیوندهای وبی روی هر چیزی که می‌تواند از وب دانسته شود نیز کمک می‌کند.

مدل پاپیونی[5] برودر و دیگران (۲۰۰۰) بر اساس دو پیماینده[6] وبی توسط موتور جستجو آلتاویستا[7] طراحی شده و از ۲۰۰ میلیون صفحه وب و ۱/۵ میلیارد پیوند وبی تشکیل می‌شود. صفحه‌های وبی می‌توانند بر اساس نحوه پیوند آنها در وب بصورت کامل در داخل یکی از گروه‌ها دسته‌بندی شوند. جزء کاملاً متصل[8] هسته وب را نشان می‌دهند، که مجموعه‌ای از صفحه‌های بسیار متصل است می‌توان از طریق یک صفحه دیگر در داخل جزء کاملاً متصل با پیگیری یک مسیر از پیوندها به آن دسترسی پیدا کرد.

1 . *inlink* 2. *self-link* 3. *reciprocal-link* 4. Broder et al.
5. bow tie 6. crawls 7. AltaVista
8. strongly connected component (SCC)



شکل۳-۲: ساختار پیوند وبی بر اساس مدل اتصال وب برودر و دیگران (۲۰۰۰)

جزء IN شامل صفحه‌هایی است که می‌توانند به جزء کاملاً متصل دسترسی پیدا کنند، ولی نمی‌توان از طریق آن در دسترس قرار گیرند، درحالی که جزء OUT شامل صفحه‌هایی است که می‌توانند از طریق جزء کاملاً متصل در دسترس قرار گیرند اما از طریق آن نمی‌توان به اجزای کاملاً متصل دسترسی پیدا کرد. همچنین صفحه‌هایی وجود دارد که متصل به جزء IN و/یا از جزء OUT هستند اما به جزء کاملاً متصل وصل نیستند، همینطور اجزای غیرمتصلی که به جزء کاملاً متصل، جزء IN یا جزء OUT، وصل نیستند.

هنگام انتخاب ابزاری برای بررسی وب، ساختار وب پیامدهای مهمی دارد. اگر نقاط شروع بررسی وبی تماماً از اجزا غیر متصل یا جزء OUT گرفته شده باشند در آنصورت در خصوص جزء IN، جزء کاملاً متصل، سایر اجزای غیر متصل، مسیرها[1] یا پیچک‌های[2] خاص نمی‌توان چیزی دانست. این ساختار تأکید می‌کند که برخورداری از چندین نقطه برای شروع پیمودن وب اهمیت زیادی دارد، و اگر چنانچه اجزای غیر

1. tubes 2. tendrils

متصل بسیاری پیدا شوند، مالکان وبگاه‌ها باید خواستار پیدا شدن آنها باشند. این ساختار بر اهمیت موتورهای جستجو در پژوهش وب‌سنجی تاکید می‌کند.

موتورهای جستجوی ۱/۰

موتورهای جستجو از زمان اولین مقاله‌های وب‌سنجی، یکی از مهم‌ترین منابع اطلاعاتی را در بررسی‌های وب‌سنجی عرضه کرده‌اند (مانند لارسون، ۱۹۹۶؛ روسوو، ۱۹۹۷؛ اینگورسن، ۱۹۹۸). آن‌ها یک منبع اطلاعاتی آماده‌ای را عرضه می‌کنند که می‌تواند یا با وارد کردن دستی کلیدواژه‌های جستجو و یا با اسکرپ[1] (خراش یا استخراج داده‌ها از وب) اطلاعات را بسرعت جمع‌آوری کند (استخراج خودکار داده‌های مورد نیاز از صفحه وب بارگیری‌شده). بخصوص آلتاویستا در بررسی‌های اولیه رایج بود زیرا قابلیت پیوند آلتاویستا امکان بررسی نحوه پیوندزنی وبگاه‌ها به یک صفحه یا دامنه خاص را فراهم می‌کرد. برخی پژوهش‌های مربوط به قابلیت پیوند آلتاویستا عبارتند از: بررسی ارزیابانه تأثیر وبگاه‌ها (مانند اسمیت[2]، ۱۹۹۹) و بررسی ارتباطی بین دانشگاه‌ها، صنعت و دولت (بودوریدز، سیگریست و آلویزوز[3]، ۱۹۹۹) و نام‌های دامنه سطح بالای عمومی (تلوال، ۲۰۰۱a).

موتورهای جستجوی کنونی (مثل بینگ[4] و گوگل) بخش بسیار بزرگی از وب را بیش از هر پژوهشگر یا گروه‌های پژوهشی به دلایل متعددی مانند: نیاز نمایه‌سازی وب به حجم بالایی از توان پردازش و پهنای باند؛ قابلیت اسناد برخط در آینده به چندین فرمت مختلف؛ نمایه می‌کنند و مالکان وبگاه‌ها می‌خواهند مطمئن شوند که محتوای آن‌ها را موتورهای جستجو نمایه‌سازی کرده است.

موتورهای جستجو از وب پیماها[5] که به نام‌های عنکبوت[6] یا ربات[7] نیز شناخته می‌شوند استفاده می‌کنند که به صورت خودکار صفحه‌ها را از وب بارگیری می‌کنند. با شروع از یک فهرست هسته[8] ازآدرس های اینترنتی، تمام صفحه‌های اشاره شده

1. scraping
2. Smith
3. Boudourides, Sigrist & Alevizos
4. Bing
5. web crawlers
6. spider
7. robot
8. seed

اساس نشانی‌های وبی بارگیری شده، و هر نشانی وبی در داخل آن صفحه‌های استخراجی جاسازی شده، و به نشانی‌های وبی اضافه شده تا بارگیری شوند. سپس این فرایند به صورت تکراری انجام می‌شود تا وب به میزان مورد نیاز بارگیری شود. وب‌پیماهای مختلف ممکن است از راهکارهای پیمایش متفاوت، اولویت‌گذاری بارگیری از صفحه‌های مشخصی از وب، یا نمایه‌سازی تنها بخشی از صفحه‌های بارگیری شده استفاده کنند. وب نه تنها وسیع نیست بلکه دائماً در حال رشد و تغییر است، و نمایه حتی بخش کوچکی از آن هم مشکل آفرین است. درحالی که پرسش درباره اندازه دقیق وب بی‌معنی است زیرا بسیاری از صفحه‌ها به صورت پویا ایجاد می‌شوند و تا حتی زمان فراخوانی از بین می‌روند، گوگل ادعا می‌کند که ۳۰ تریلیون نشانی وبی منحصر به فرد شناسایی کرده است و به طور متوسط روزانه ۲۰ میلیارد از آن‌ها را پیمایش می‌کند (سولیوان[1]، ۲۰۱۲).

این صفحه‌های وبی تنها به صورت فرمت استاندارد زبان نشانه‌گذاری فرامتن نیستند بلکه به صورت طیف گسترده‌ای از سایر فرمت‌ها نیز هستند، از فلش[2] تا اسناد مایکروسافت ورد[3]. گوگل علاوه بر زبان نشانه‌گذاری فرامتن طیف گسترده‌ای از مدارک را نیز نمایه‌سازی می‌کند (گوگل، ۲۰۱۳)، اگرچه چنین تولید انبوهی از وب‌پیماها احتمالاً فراتر از کار گروه‌های پژوهشی کوچک یا شخصی است.

در حالی‌که مالکان وبگاه‌ها می‌خواهند مطمئن شوند که محتوای وبگاه آن‌ها را موتورهای جستجو نمایه‌سازی کرده‌اند یا خیر، نه تنها علاقه‌مند نیستند که محتوای آنها را وب پیمای یک پژوهشگر نمایه‌سازی کند، بلکه حتی اقداماتی برای ممانعت از پیمودن وب‌پیمای پژوهشگر نیز انجام می‌دهند. یک صفحه وبی فقط وقتی می‌تواند مشاهده شود که از طریق صفحه وب دیگری به آن پیوند شده باشد، و ساختار وبی اینگونه است که صفحه‌های وب بسیاری وجود دارند که یا غیرمتصل هستند یا به هسته کاملاً متصل وبی پیوند دارند اما نمی‌شود از طریق هسته کاملاً

1. Sullivan 2. Flash 3. Microsoft Word

متصل با آن‌ها پیوند برقرار کرد (برودر و دیگران، ۲۰۰۰). اگرچه ممکن است مالکان وبگاه‌ها، موتورهای جستجو اصلی مانند گوگل را برای نمایه‌سازی پایگاه وبی یا صفحه خود انتخاب کنند، اما آن‌ها در واقع می‌خواهند از پیمودن وب توسط پژوهشگران ممانعت کنند و از استاندارد حذف ربات‌ها استفاده می‌کنند تا این واقعیت را نشان دهند که آن‌ها نمی‌خواهند از طریق وب پیماهای ناشناخته پیمایش شوند. از آنجایی که پیمایش زیاد ممکن است بدون هیچ نفع ملموس و قابل رویتی برای مالک بر یک پایگاه وبی کوچک تأثیر چشمگیری داشته باشد، ولی برخورداری از یک رویکرد اخلاقی در پیمایش وب برای پژوهشگران مهم است. این رویکرد این است که به استانداردهای حذف ربات‌ها احترام بگذارند و وبگاه‌هایی که استاندارد حذف ربات ندارند نیز باید اقدامات احتیاطی منطقی انجام دهند (تلوال و استوارت، ۲۰۰۶).

با این وجود، کاربرد داده های پیمایشی موتورهای جستجو در بررسی پیوند بین پایگاه‌های وب معایبی دارد. این مسائل پوشش، وضوح، قابلیت اطمینان، قابلیت پیوند و قابلیت بازیابی تمام نتایج را در بر می‌گیرد. اگرچه پوشش موتورهای جستجو از وب به نسبت آنچه که اغلب پژوهشگران می‌توانند برای خودشان جمع‌آوری کنند بسیار گسترده‌تر است، اما موتورهای جستجو تمام وب را پیمایش نمی‌کنند. این موضوع تا حدودی به محدودیت‌های فنی برمی‌گردد (مشکل شناسایی تمام اجزا غیر متصل برودر و دیگران۲۰۰۰) و تا حدودی هم ناشی از تصمیم‌های اولویت‌بندی پیمایش بخش‌های خاصی از وب است (مانند اولویت دادن به صفحه‌های انگلیسی در بالای وبگاه‌ها). پژوهشگر وب‌سنجی نمی‌داند سیاست پیمایش یک موتور جستجو چیست زیرا موتور جستجو برای اینکه در رقابت جلو بماند سیاست شفافی ندارد. لازم است که پژوهشگران وب‌سنجی بدانند چه چیزی وب‌پیمایی شده است و یک موتور جستجو چگونه نتایج را رتبه‌بندی کرده و تعداد نتایج بدست آمده را تخمین می‌زند. تعداد تخمینی نتایج یک کلیدواژه خاص اغلب بسیار متفاوت تر از عدد واقعی نتایجی است که یک موتور جستجو می‌تواند امکان

دسترسی به آن را برای خواننده فراهم کند. برای نمونه، جستجو عبارتی در گوگل برای کتابی با عنوان *تسهیل دسترسی به وب داده*[1] حدود ۱۲۳۰۰ نتیجه دارد. هرچند اگر شخصی در صفحه‌های نتایج کلیک کند در صفحه ۱۵ به توقف ناگهانی برمی‌خورد و پیام می‌رسد که تنها ۱۴۶ نتیجه پیدا شده که با این کلیدواژه هم‌خوانی دارد. موضوع قابلیت اعتماد می‌تواند پیامدهای حادی برای پژوهش وب‌سنجی داشته باشد زیرا موتورهای جستجو اغلب تعداد نتایج قابل مشاهده را محدود می‌کنند (برای نمونه، تا ۱۰۰۰ عدد). از این رو هیچ راهی برای تعیین تعداد واقعی نتایج برای اغلب جستجوها وجود نداشته و این اعداد صرفاً تخمین‌هایی ابتدایی هستند. تعداد محدود نتایج نه تنها دقت تعداد نتایج بدست آمده را محدود می‌کند، بلکه پیامدهایی برای اعتباریابی نتایج از طریق تحلیل محتوا دارد. تحلیل محتوا نمی‌تواند بر اساس نمونه تصادفی تمام نتایج مشابه باشد، بلکه الزاماً باید بر اساس ۱۰۰۰ نتیجه‌ای باشد که می‌تواند نتایج را مخدوش کند.

علاوه بر محدودیت نتایج، محدودیت‌هایی هم در کلیدواژه‌هایی است که می‌توان آن‌ها را به موتورهای جستجو فرستاد. با وجودی که موتورهای جستجویی مانند گوگل امکان استفاده از تعدادی قابلیت های جستجو پیشرفته همچون جستجوی عبارت و عملگرهای بولین[2] (برای نمونه، AND, OR, NOT) و محدودیت نتایج بر مبنای ناحیه، زبان، نوع سند و حتی سطح خواندن موارد و قابلیت‌های جستجو پیوند را فراهم می‌کنند، قابلیت‌های زیادی دارند که در دسترس نیستند. هنگامی که جستجوی پیوندی در دسترس بود در سطح دامنه یا در صفحه عملیاتی‌سازی می‌شد و امکان جستجوی پیوندهایی که به زیردامنه‌ها یا راهنماها اشاره می‌کرد وجود نداشت.

جستجوی کلیدواژه‌ای و عبارتی هم امکان جستجوهای پیچیده‌ای که از طریق عبارت‌های معمولی[3] می‌توانسته بیان شود را نمی‌داد. برای نمونه، هیچ راهی نیست تا از گوگل بخواهیم تمام صفحه‌ها را با یک آدرس الکترونیکی برگرداند چرا که هیچ

1. *Facilitating Access to the Web of Data* 2. Boolean 3. regular expressions (regex)

راهی برای بیان مفهوم یک آدرس الکترونیکی در جعبه جستجو گوگل نیست. هرچند ساختار یک آدرس الکترونیکی را می‌توان به عنوان یک عبارت معمولی بیان کرد، که آن‌هم امکان بیان الگوهای نوشتاری را می‌دهد.

بخشی از این محدودیت‌ها مربوط به حفاظت از مالکیت معنوی و بخشی دیگر محدودیت‌های برای نمایه‌سازی است. نمایه‌سازی وبی برای هر جستجوی عبارت معمولی موجود امکان‌پذیر نیست و درعوض سازگاری یک عبارت معمولی با مجموعه خاصی از داده ها الزامی است. به همین دلیل پژوهشگر باید برای خودش نسخه‌ای از داده‌ها را داشته باشد.

وب پیماها

اگرچه ممکن است پژوهشگران نتوانند کل وب را به تنهایی پیمایش کنند اما می‌توانند بخشی از آن را بپیمایند. تعدادی وب‌پیما به صورت رایگان در دسترس پژوهشگران هستند مانند بایگانی اینترنت هاریتیکس[1] و گروه پژوهش سایبرسنجی آماری ساکسایبوت[2]. طراحی ساکسایبوت به عنوان یک وب‌پیما بخصوص برای اهداف وب‌سنجی (تلوال، b۲۰۰۱)، امکان دسترسی به داده های بسیار موثقی را برای پیوندها و محتوای وبگاه‌ها میسّر می‌کند.

پژوهشگر وب‌سنجی می‌تواند با استفاده از یک وب‌پیما پارامترهای داده‌ای که می‌خواهد در یک پیمایش لحاظ شود را مشخص کند. همچنین انتظار می‌رود هرگاه مجموعه داده ها مشابه جستجو شود، همان نتایج بدست آید. این بدان معنی نیست که بگوییم هیچ ابهامی در داده‌های حاصل از یک وب‌پیما وجود ندارد: وب‌پیماهای مختلف ممکن است داده‌های مختلفی جمع‌آوری کنند چرا که آنها می‌توانند داده‌ها را از پیوندهای مختلفی استخراج کنند، و حتی اگر خود صفحه‌های وبی تغییر نکرده باشند، سرورهایی که بدون واکنش هستند چنین برداشت شود که نسخه‌های وبی

1. Internet Archive's Heritrix (http://crawler.archive.org)
2. Statistical Cybermetrics Research Group's SocSciBot (http://socscibot.wlv.ac.uk)

مختلفی قابل دستیابی هستند. ممکن است طیف گسترده‌تر از جستجوهایی که در موتورهای جستجوی معمولی امکان داشت ایجاد شود، از جمله جستجوکننده‌های پیوند. از منظر یک روش علمی خوب، برای اشتراک‌گذاری مجموعه داده‌های بررسی شده دو امکان وجود دارد، یا دسترسی برخط آن، و یا ارائه به اشخاصی که خواستار آن هستند.

مطالعه وب پیماهای زیادی متوجه فضای وبی دانشگاهی بریتانیا هستند (برای نمونه، تلوال و پرایس، ۲۰۰۳؛ تلوال، هاریس[1] و ویلکینسون[2]، ۲۰۰۳)، چرا که فضای وبی دانشگاهی بریتانیا آنقدر کوچک است که می‌توان آن را با یک گروه پژوهشی پیمود و به اندازه‌ای برزگ است که مورد توجه قرار گرفته و فرصتی را برای انجام آزمون‌های همبستگی بین تأثیر وب و RAE فراهم کند.

اگرچه وب‌پیماها امکان بررسی‌های موثق تری را برای تمرکز بر روی یک قسمت کوچک تر وبی می‌دهند، افزایش قابلیت موتورهای جستجو به شکل رابط‌های برنامه‌نویسی برنامه کاربردی، که دسترسی به حجم زیاد داده های خود را تسهیل می‌کنند، توجه وب‌سنجی را به سمت آن‌ها سوق داده است.

موتورهای جستجو ۲/۰

در سال ۲۰۰۲ گوگل شروع به طراحی یک رابط برنامه‌نویسی برنامه کاربردی برای دسترسی خودکار به موتور جستجوی خود کرد، و یاهو و لایو سرچ[3] (آنگونه که موتور جستجوی مایکروسافت در آن زمان نام‌گذاری کرده بود) خیلی زود این کار را انجام دادند. از آنجایی که در هر سه موتور جستجو نیز امکان شناسایی پیوندها به یک صفحه یا دامنه خاصی را تعبیه می‌کنند، آن‌ها این امکان را برای مطالعات وب‌سنجی فراهم می‌کنند که مجموعه بسیار بزرگ‌تری از وبگاه ها را که می‌توانستد با یک وب‌پیما شخصی پیموده شوند، در نظر بگیرند. اگرچه موتورهای جستجو هم‌چنان بسیاری از محدودیت‌هایی شناسایی شده قبلی را در خصوص قابلیت

1. Harries 2. Wiklkinson 3. Live Search

اعتماد، شفافیت، جستجو های ممکن و قابلیت بازیابی تمام نتایج دارا هستند، ظاهراً رابط های برنامه نویسی برنامه کاربردی برای پژوهش وب‌سنجی وضعیت را به نفع موتور جستجو تغییر داده است. رابط‌های برنامه‌نویسی برنامه کاربردی موتور جستجو اساس بسیاری از مطالعات وب‌سنجی را چه ارتباطی (مانند استوارت و تلوال، ۲۰۰۷) و چه ارزیابانه (مانند کوشا[1] و تلوال، ۲۰۰۸) تشکیل داده‌اند. هرچند، مزایای موتور جستجو عمر نسبتاًکوتاهی داشت. قابلیت پیوند نادیده گرفته شده است و رابط‌های برنامه‌نویسی برنامه کاربردی رو به زوال گذاشتند.

بررسی توانایی سایر انواع پیوندزنی الزامی است، چون سابقاً می‌توانستند با استفاده از قابلیت پیوند موتورهای جستجوی، پیوندهای بین وبگاه‌ها را مورد بررسی قرار دهند. هر چند که توسعه رابط‌های برنامه‌نویسی برنامه کاربردی توسط موتورهای جستجو برای پردازش خودکار محتوای موتور جستجو مزیت پژوهشی وب‌سنجی محسوب می‌شد، از ماه می ۲۰۱۱ رابط‌های برنامه‌نویسی برنامه کاربردی جستجو در بینگ ۲/۰ به عنوان تنها منبع پژوهش وب‌سنجی از یک موتور جستجوی اصلی مناسب برای پردازش برون‌خطی[2] شناسایی شد (تلوال و سود[3]، ۲۰۱۲). در زمان نگارش، بینگ هیچ قابلیت پیوند درون‌ساختی را در بر نمی‌گیرد و کاربران به ۵۰۰۰ جستجوی ماهانه محدود هستند.

اولویت‌های موتورهای جستجوی تجاری با اولویت‌های افرادی که از موتورهای جستجو به عنوان ابزارهای جستجو استفاده می‌کنند متفاوت است. نتایج متناقض، عملگرهای موتور جستجو متناقض و سیاست‌های رتبه‌بندی و وب‌پیمای پنهان از تجربه کاربر متوسط از یک موتور جستجو می‌کاهند، با وجود اینکه برای پژوهشگر دو خصیصه ثبات و باز بودن بسیار ارزشمند خواهد بود.

متأسفانه هیچ موتور جستجویی خصوصیات یک موتور جستجوی ایده‌آل را برای پژوهش وب‌سنجی بدست نمی‌آورد (برای نمونه، بار-ایلان، ۲۰۰۵). با این

1. Kousha 2. offline 3. Sud

وجود، آن‌ها هم‌چنان در پژوهش‌های وب‌سنجی نقش دارند چرا که پژوهشگران روش‌های خود را با توجه به محدودیت‌های ابزارهای موجود تنظیم می‌کنند. برای نمونه، مقاله‌های اخیر وب‌سنجی حاوی مطالعه ای بود درباره اینکه آیا تعداد صفحه‌های وبی که نام یک شرکت در آن‌ها ظاهر می‌شود با سود، دارایی و درآمد آن شرکت همبستگی دارد یا نه(وگان و رومرو-فاریاس[1]، ۲۰۱۲)، و برای کشف شباهت‌ها و تفاوت‌های مسیر استفاده وبی در دو کشور کره و ژاپن از تحلیل محتوای صفحه‌های وبی آنها (هاسو و پارک[2]، a۲۰۱۲) استفاده شد. هرچند، ما بیش از پیش به سمت خرد کردن[3] ابزارهای پژوهش حرکت کرده‌ایم.

پسا موتور جستجو ۲/۰: خرد کردن

در حال حاضر ما بدون هیچ ابزارهای وب‌سنجی غالبی به سمت دوره پسا موتور جستجو ۲/۰[4] حرکت کرده‌ایم. در عوض پژوهشگران وب‌سنجی درحال جستجو برای جایگزین های ابرپیوند به عنوان شاخص تأثیر یا ارتباط، همواره به دنبال منابع جدید داده‌ای و نیز بازگرداندن ابزارهای پیشین هستند.

موتورهای جستجو و پیوندهای جایگزین: استنادهای نشانی وبی و اشاره‌های وبی

ابرپیوندها تنها شاخص تأثیر وب نیستند، و کاهش قابلیت‌های تحلیل پیوند از موتورهای جستجو باعث شده که به روش‌های جایگزینی که در آن برای محاسبه تأثیر از وب‌پیماهای[5] گسترده موتورهای جستجوی عمومی استفاده می‌شود، توجه روزافزونی شود.

یک نوع شاخص تعداد استنادهای نشانی وبی است، به دلیل شباهتش با ابرپیوندها. استنادهای نشانی وبی یک صفحه وبی به عنوان «اشاره‌هایی از نشانی وبی آن در آزمون سایر صفحه‌های وبی، چه ابرپیوند باشد و چه نباشد» تعریف شده‌اند (کوشا و تلوال، ۲۰۰۶). تمام استنادهای نشانی وبی به یک وبگاه می‌تواند با انجام

1. Vaughan and Romero-Frias
2. Hsu and Park
3. fragmentation
4. Post search engine 2.0
5. crawl

جستجوی یک عبارت برای یک نشانی وبی و نادیده گرفتن آن استنادها در همان وبگاه محاسبه شود. برای مثال:

www.bbc.co.uk-site:bbc.co.uk

استنادهای نشانی وبی ممکن است برای سنجه‌های وبی ارتباطی نیز استفاده شوند. برای نمونه، استوارت و تلوال (۲۰۰۶) بررسی کردند که آیا همکاری[1] بین دانشگاه‌ها، دولت و سازمان‌های تجاری در وست میدلند[2] در بریتانیا از طریق استنادهای نشانی وبی بین وبگاه‌ها قابل رویت است یا خیر. برای نمونه، اشاره‌های نشانی وبی وبگاه انجمن ولورهامپتون[3] در وبگاه دانشگاه ولورهامپتون:

www.wolverhampton.gov.uk-site:wlve.ac.uk

با این وجود، مشکلاتی در استنادهای نشانی وبی وجود دارد، چرا که تنها شاخص در تعداد ابرپیوندها است و نتایج موتورهای جستجو غیر قابل اعتماد محسوب می‌شوند (تلوال و سود، ۲۰۱۱). با این وجود، آن‌ها هم‌چنان به عنوان منبع بالقوه تأثیر مورد بررسی قرار می‌گیرند (مانند اوردنا–مالیا و رگاززی[4]، ۲۰۱۳).

ممکن است اشاره‌های وبی یک عبارت خاص نیز مورد بررسی قرار بگیرند. یک شیوه ارزیابی تأثیر وب، ارزیابی تأثیر منابع یا ایده است، از طریق ارزیابی که هرچند وقت یک‌بار به آن‌ها اشاره برخط می‌شود (تلوال، ۲۰۰۹)؛ این ایده به صورت مفصل‌تر در فصل هفتم بررسی خواهد شد که در آنجا این روش به منابع کتاب‌سنجی اضافه شده است. هم- اشاره ها[5] نیز ممکن است مورد بررسی قرار بگیرند. برای نمونه، چونگ و پارک[6] (۲۰۱۲) رویت‌پذیری شبکه‌ای خود را با بررسی اینکه آیا نامشان با هم آورده شده است یا خیر بررسی کردند.

رابط‌های برنامه‌نویسی برنامه کاربردی و وب‌اسکرپرها

لازم نیست پژوهش وب‌سنجی تنها حول محور داده‌هایی باشد که موتورهای

1. co-operation
2. West Midlands
3. Wolverhampton
4. Orduna-Malea and Regazzi
5. co-mentions
6. Chung and Park

جستجو در دسترس قرار می‌دهند. بسیاری از وبگاه‌های مهم در معرض مطالعات وب‌سنجی بوده‌اند، چه از طریق استفاده مجاز از داده های وبگاه بواسطه رابط‌های برنامه نویسی یا از طریق دسترسی غیرمجاز به داده ها به وسیله یک وب‌اسکرپر[1]. دو اصل کلیدی انقلاب وب ۲/۰: اهمیت داده ها و مهار هوش جمعی (اورلی[2]، ۲۰۰۵). بسیاری از وبگاه‌ها و خدمات وبی ۲/۰ امکان جمع‌آوری خودکار داده‌ها را از خدمات خود میسّر کرده‌اند، بطوری که بعدها پایه و اساس پژوهش‌های دانشگاهی را شکل داده است: توییتر (مانند استوارت، ۲۰۱۰؛ چیو[3]، پارک و پارک۲۰۱۲؛ ویلکینسون[4] و تلوال، ۲۰۱۲)، فلیکر (مانند آنگس[5]، تلوال و استوارت، ۲۰۰۸)، یوتیوب (مانند بنونتو[6] و دیگران، ۲۰۰۸) و دیگ[7] (پارتوگلو[8]، تلوال و بوکلی[9]، ۲۰۱۰).

البته، بیش‌تر وبگاه‌ها استخراج داده‌ها را سخت‌تر می‌دانند، و برای استخراج اطلاعات مورد نیاز از وبگاه‌های خاص ایجاد ابزارهای سفارشی مانند مای اسپیس[10] (تلوال، ۲۰۰۸) و اظهارنظرهای روی پایگاه وبی بی.بی.سی (پارتوگلو، تلوال و بوکلی، ۲۰۱۰) را ضروری می‌دانند. جایی که رابط برنامه کاربردی در دسترس نباشد، ممکن است از وب‌اسکرپرها استفاده شود. وب‌اسکرپر مشابه وب‌پیما است، هرچند به‌جای بارگیری محتوای ساختارنیافته صفحه وبی، از آنها برای استخراج محتوای خاص در وب استفاده می‌شود. برای نمونه، امکان دارد یک وب‌اسکرپر به منظور استخراج اظهار نظرها از وبگاه یک روزنامه خاص یا نوشته‌های یک وبگاه شبکه های اجتماعی طراحی شده باشد. همچنین تعدادی اسکرپرها در اسکرپرویکی[11] موجود است. این وبگاه بستری است که برنامه‌نویس‌ها می‌توانند در آن کدی ایجاد کنند که به صورت خودکار اجرا شده و داده ها ذخیره می‌شود. حتی اگر کسی بفهمد که داده‌ها هنوز اسکرپ نشده و مهارت لازم را برای نگارش برنامه اسکرپ نداشته باشد، گزینه درخواست داده ها هم در دسترس است: می‌توانید به کسی پول پرداخت کرده تا داده‌ها را برای شما اسکرپ کند.

1. web scrapers
2. O'Reilly
3. Choi, Park & Park
4. Wilkinson
5. Angus
6. Beenevenuto et al.
7. Digg
8. Paltoglou
9. Buckley
10. MySpace
11. ScraperWiki (http://scraperwiki.com)

توسعه نرم‌افزار چه برای استخراج خودکار داده‌ها از رابط برنامه کاربردی در وبگاه و چه برای اسکرپ کردن محتوا از خود وبگاه احتمالاً از قابلیت‌های بسیاری از پژوهشگران بیش‌تر بوده، همانگونه که گیلس[1] (۲۰۱۲) اشاره کرده است، در حالی‌که دانشمندان علوم اجتماعی پرسش‌های جالب بسیاری در ذهن دارند برای جمع‌آوری و بررسی داده‌های فراوان وبی از مهارت‌های لازم رایانه‌ای برخوردار نیستند.

عرضه داده‌ها از طریق یک رابط برنامه کاربردی امکان کنترل بیش‌تری را بر روی محتوا به وبگاه می‌دهد: آن‌ها سطح دسترسی‌ها را محدود کرده و محتوا را از طریق رابط برنامه کاربردی به اشتراک می‌گذارند و مجبور نیستند تمام محتوای خود را نیز در اختیار همگان قرار دهند. در بسیاری از موارد چنین حالتی به نفع کاربر نهایی است. برای نمونه، حتی اگر یک موتور جستجو در صدد باشد تمام محتوای یک فایل بزرگ را به اشتراک بگذارد، برای بسیاری از کاربران بالقوه بلااستفاده خواهد بود!

با اینکه رابط‌های برنامه‌نویسی برنامه کاربردی می‌توانند در خصوص کاربردهای وب نگرش جالبی به ما بدهند، اما چشم‌انداز گسسته و بی‌ربطی از وب را به ما ارائه می‌کنند. چنانچه رابط‌های برنامه‌نویسی برنامه کاربردی محتوای خود را به این روش و با توجه به استانداردهای گسترده، و با مشخصات منحصر به فرد برای اشاره به پیشینه افراد ساختاردهی کنند، بسیار مفید خواهد بود. چنین رویکردی چشم‌اندازی از وب معنایی است. وب معنایی، و گسترش ابزارهای وب‌سنجی هنوز در سوابق پژوهشی وب‌سنجی جایی ندارند. هرچند توانایی ارائه الگویی جدید برای پژوهش وب‌سنجی را در آینده خواهند داشت، آنچنان که اجزای پراکنده کنونی را به هم متصل می‌کنند، و در فصل هشتم توضیح داده خواهد شد.

جمع‌کننده‌های داده

در حال حاضر ابزارهایی چون جمع‌کننده‌های داده[2] در خصوص نوع خاصی از داده ها وجود دارند. هرچند به نظر می‌رسد موتورهای جستجوی کنونی فاصله زیادی

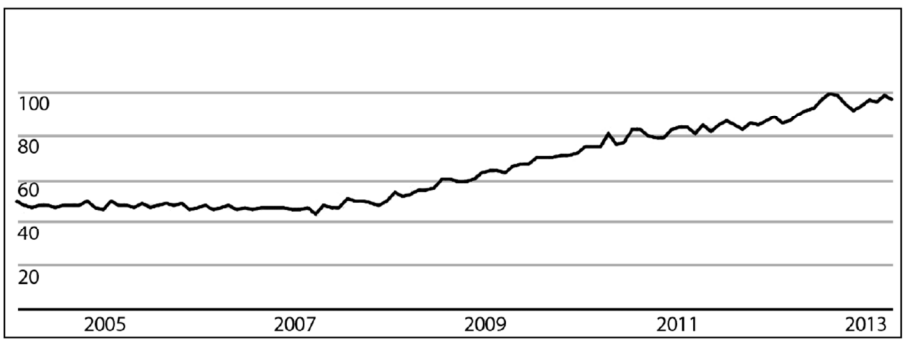
1. Giles 2. Data aggregators

با موتورهای جستجوی ایده‌آل وب‌سنجی دارند همانطور که تاکنون نبوده‌اند حتی در آینده نیز به آنها نمی‌رسند، اما ابزارهای جدیدی بویژه توسط گوگل ایجاد شده است. ماندگاری خدمات در گوگل تنها چهار سال است (آرتور[1]، ۲۰۱۲) و برخی ابزارهایی که ارزشمندی آنها فقط یکبار در پژوهش وب‌سنجی تأیید شده کنار گذاشته شده‌اند. اینها رابط برنامه کاربردی نموداری اجتماعی گوگل[2] هستند، که ارتباطات اجتماعی ابراز شده را بر مبنای دو استاندارد مرسوم نمایه می‌کرد، و در آوریل سال ۲۰۱۲ کنار گذاشته شد. آگاهی‌های گوگل برای جستجو (گوگل اینسایت فور سرچ[3])، نیز اطلاعاتی برای عبارت‌های مورد جستجو درگوگل می‌دهد، و در سپتامبر ۲۰۱۲ با گوگل ترندز[4] ادغام شد.

گوگل به عنوان رایج‌ترین موتور جستجو در آگوست سال ۲۰۱۲ اعلام کرد که ماهانه ۱۰۰ میلیارد جستجو را مدیریت کرده (سولیوان، ۲۰۱۲)، و می‌تواند در خصوص محتوای برخط موجود آگاهی بدهد، و تعدادی ابزار که اطلاعات چیزهای مورد جستجوی عموم مردم را می‌دهد. گوگل ترندز روش ساده‌ای برای مشاهده تغییر مقدار جستجوی انجام شده درباره یک عبارت خاص را در طی زمان نشان می شود. برای نمونه، شکل ۳-۳ مقدار جستجوهایی که با کلیدواژه «چگونه[5]» بوده را نشان می‌دهد. این شکل نشان می‌دهد که تعداد جستجوهای «چگونه» از ژانویه ۲۰۰۴ تا آخر سال ۲۰۰۷ نسبتاً ثابت بوده و از آن زمان به بعد افزایش ممتدی داشته است. این فرضیه اولیه منطقی بنظر می‌رسد که با بحران مالی ۲۰۰۷-۲۰۰۸، و رکود جهانی پس از آن، تعداد اشخاصی را که ترجیح می‌دادند کارها را خودشان انجام دهند تا اینکه از متخصص کمک بگیرند افزایش یافت. هرچند گوگل نمی‌تواند بگوید چرا افراد در این زمان خاص شروع به جستجو درباره چنین عبارت‌هایی کرده‌اند، اما می‌تواند بر مبنای جستجوهای منطقه‌ای، که ضمیمه اطلاعات بیش‌تری از اقتصادهای فردی است اطلاعات مفصل‌تری بدهد، شواهد حمایتی ممکن است ارائه شود یا

1. Arthur
2. Google's Social Graph
3. Google Insights for Search
4. Google Trends
5. how to

نشود. بهمین طریق، پریس[1] و همکارانش (۲۰۱۲) از گوگل ترندز استفاده کرده‌اند تا نشان بدهند که کاربران اینترنتی کشورهای دارای تولید ناخالص ملی بیش‌تر، در آینده ممکن است بیش از گذشته به جستجوی اطلاعات بپردازند.



شکل ۳-۳: سهم جستجوهای گوگل بر اساس داده‌های گوگل ترندز از عبارت «چگونه» بین سال‌های ۲۰۱۳ تا ۲۰۱۵
[گوگل و آرم گوگل به عنوان نماد تجاری برای شرکت گوگل ثبت شده‌اند، استفاده با اجازه]

معروف‌ترین رفتار جستجو که می‌تواند فعالیت های دنیای واقعی را منعکس کند آنفولانزا گوگل[2] است که در آن گوگل از داده‌های جستجو برای تخمین شیوع آنفولانزا استفاده می‌کند. برخی بر این باورند که چنین بررسی‌هایی می‌تواند به کشف اولیه شیوع بیماری، و در نتیجه کاهش ابتلا کمک کنند. به دنبال موفقیت آنفولانزا گوگل، از همبسته گوگل (گوگل کورلیت[3]) بمنزله ابزاری برای کمک به شناسایی عبارت‌هایی که با پدیده خاصی همبستگی دارند رونمایی کرد.

نتیجه‌گیری

قطعاً وب به تغییرات خود در آینده ادامه خواهد داد. وب مانند کتابخانه یک عضو در حال رشد بوده و کتابدارانی که می‌خواهند از سنجه‌های‌وبی استفاده کنند باید بپذیرند که تغییر امر اجتناب‌ناپذیری است. وبگاه‌ها و فناوری‌های ارائه شده در این کتاب

1. Pries
2. Google Flu Trends(www.google.org/flutrends)
3. Google Correlate (www.google.com/trends/correlate)

تغییر می‌کنند ولی اصول همواره ثابت هستند. در حال حاضر بخش بزرگی از وبگاه‌های شبکه‌های اجتماعی را فیسبوک، توییتر و لینکداین تسخیر کرده‌اند و دور از انتظار نیست که مورد توجه کتابدارانی باشد که می‌خواهند تأثیر خود را اندازه بگیرند. هرچند، همانگونه که تاکنون دیده شده است وبگاه‌های رسانه‌های اجتماعی به سرعت می‌آیند و می‌روند، و چیزی که قابل ذکر نیست اینکه فناوری‌های کنونی از شبکه‌های اجتماعی قبلی و از رده خارج شده‌ای مانند فرندزتر[1] و مای اسپیس (بوید و الیسون[2]، ۲۰۰۷) پیروی نخواهند کرد، حتی اگر همان کارآیی را داشته باشند. کتابداران باید انواع اطلاعات ارزشمند، انواع ابزارهایی که احتمالاً موجود هستند، و برخی محدودیت‌های بالقوه سنجه‌های وبی را شناسایی کنند.

امروزه قطعاً پژوهش‌های وب‌سنجی و در حالتی کلی‌تر از سنجه‌های وبی با کمبود یک ابزار واحد مواجهند. درحالی که مسلماً در چنین شرایطی زمینه انواع گسترده‌تر تحقیقات به نسبت روش‌های سنجه‌های وبی قدیمی‌تر فراهم می‌شود، روش شناختی‌های فعلی کم‌تر به توصیف زیاد توجه می‌کنند. در این صورت ضروری است که پژوهشگران از مجموعه ابزارهای بیش‌تری استفاده کنند. هرچند این ابزارها بحث اصلی پنج فصل پیش‌رو ما هستند اما بسیار حائز اهمیت است که ماهیت تغییر ابزارها و فناوری‌ها را در خاطر داشته باشیم.

1. Friendster 2. Ellison

۴

ارزیابی تأثیر در وب

مقدمه

وب فرصتی را برای کاربرد سنجه‌ها در طیف وسیعی از پژوهش‌ها ذیل دو گروه ارزیابانه یا ارتباطی فراهم می‌آورد. این فصل بطور اجمالی به سنجه‌های ارزیابانه برای وبگاه ها، وب‌نوشت‌ها و سایر محتواهای خود- میزبان می‌پردازد. فصل پنجم سنجه‌های ارزیابانه را برای خدمات رسانه های اجتماعی گروه- ثالث که مورد استفاده سازمان‌ها و افراد زیادی است مدّ نظر قرار داده و فصل ششم به سنجه‌های وبی ارتباطی می‌پردازد. از جنبه‌های مختلف، شکاف بین محتوای خود- میزبان و گروه- ثالث ساختگی است. برای مثال، وبگاه مانند بلاگر میزبان میلیون‌ها وب‌نوشت وبگاه‌های گروه-ثالث است و لازم نیست افراد و سازمان‌ها بر یک ریز وب‌نویسی[1] مانند توییتر تکیه کنند، بلکه می‌توانند با استفاده از نرم‌افزار متن باز[2] مانند استتوس دات نت[3] خدمات ریز وب‌نویسی خود را ارائه کنند. این شکاف حاکی از تفاوت چشمگیری است که کاربر بر کنترل داده‌های موجود دارد و نیز عدم اطمینان در داده‌های جمع‌آوری شده است.

1. microblogging 2. Opne-source 3. http://status.net

زمانی که افراد یا سازمان‌ها میزبان محتوای خود در سرورهای خود هستند در خصوص استفاده از محتوای وبگاه خود مشکلات فنی تنها محدودیتی است که در گردآوری اطلاعات با آن مواجه می‌شوند. هرچند، زمانی که یک وبگاه در بستر زیست بومی وسیع‌تری ادغام می‌شود عدم اطمینان بیش‌تری وجود دارد: همانگونه که درباره تعداد صفحات وب هیچ پاسخ ساده‌ای وجود ندارد، هیچ پاسخ ساده‌ای هم برای اینکه چند پیوند به یک صفحه اشاره می‌کنند یا چند بار به آن اشاره شده است وجود ندارد. مسئله‌ای که بیش‌تر مشکل ایجاد می‌کند این است که وب الزاماً از طریق چشم‌انداز محدود یک موتور جستجوی خاص یا سایر ابزارها دیده می‌شود، و محدودیت‌های ابزارها همیشه روشن نیست. در مقابل، بسیاری از وبگاه‌های شبکه های اجتماعی محیطی محدود ارائه می‌کنند، با فقط رفتار خاصی (به‌دلیل محدودیت‌های فنی و ویرایشی) مجاز است و تنها می‌توان به داده‌های خاصی دسترسی داشت.

این محدودیت‌ها ابهام در داده های موجود از خدمات را نیز کاهش می‌دهند. یک حساب توییتر تعداد مشخصی دنبال‌کننده دارد، یک کاربر فیسبوک تعداد مشخصی دوست دارد، یک ویدئو یوتیوب تعداد مشخصی بازدیدکننده دارد. امکان دارد که ما با مفهوم لایک در فیسبوک یا آنچه به عنوان نظر در یوتیوب تلقی می‌شود مخالف باشیم، اما این‌ها بخش‌های عینی و قابل درکی محسوب می‌شوند.

سنجه‌های وب ارتباطی همواره برای ارزیابی داده ها از وبگاه‌ها و خدمات گروه- ثالث الزامی هستند. هرچند تحلیل می‌تواند متفاوت باشند، اما برای جمع‌آوری داده‌ها می‌توان از ابزارهای مشابه بسیاری استفاده کرد. این فصل به انواع محتوای میزبانی که ممکن است در وب یافت شوند (وبگاه‌ها، وب‌نوشت‌ها و ویکی‌ها) و انواع اطلاعاتی که می‌توانند در جمع‌آوری اهداف ارزیابانه مورد توجه کتابداران باشند، می‌پردازد. سپس نگاه دقیق‌تری به روش اصلی تحلیل محتوای طیف وسیعی از وبگاه‌ها یعنی تحلیل محتوا خواهد داشت.

وبگاه‌ها

اکثر دارندگان وبگاه ها برای ارزیابی تأثیر برخط خود به روشی، شکلی یا نوعی گام‌هایی بر می‌دارند. در حال حاضر ممکن است رتبه بالا در نتایج جستجو بسیار کم‌تر از قبل مورد توجه باشند، چرا که صفحات خانگی با طراحی خوب در شبکه های اجتماعی وجود دارند. با این وجود اکثر افراد و سازمان ها که وبگاه دارند برای اینکه بتوانند در خصوص تأثیر محتوای آنها اطلاعاتی بدست آورند گام هایی برداشته‌اند. ممکن است آن‌ها به اندازه‌ای که باید به این داده‌ها نپردازند و در عوض روشی که مستلزم کم ترین تعهد موجود است را برگزینند و سپس نتوانند داده‌ها را بخوانند. اما اغلب دارندگان وبگاه‌ها توجه زیادی به ارزش بالقوه این داده‌ها در تحلیل وب دارند. تا زمانی که مالک یک وبگاه برای ارزیابی چگونگی تعامل مردم با وبگاه خود گام‌هایی بر ندارد، نمی‌توانند مطمئن شود که اهداف وبگاه محقق شده است یا خیر. در واقع آن‌ها نمی‌توانند مطمئن باشند آیا شخصی از وبگاه آنها بازدید می‌کند یا نه! این اهداف بطور چشمگیری با نوع وبگاه تغییر می‌کنند.

ساده‌ترین وبگاه می‌تواند یک صفحه کاملاً متنی باشد: بدون هیچ پیوندی، بدون هیچ جایی برای کلیک، فقط کلماتی روی یک صفحه. با این وجود، حتی برای چنین صفحه ساده‌ای مدیر وبگاه می‌خواهد بداند چند نفر این محتوا را می‌خوانند. از آن‌جایی که آنها نمی‌توانند بررسی کنند که هر بازدیدکننده واقعاً محتوا را می‌خواند، یا در واقع آیا در یک زمان بیش از یک نفر محتوا را می‌خواند، اغلب تعداد دفعات درخواست صفحه‌های وب را از سرور به عنوان شاخصی برای تعداد بازدیدکننده‌ها می‌پذیرند، اگرچه این امر به هیچ وجه محدودیتی برای سنجه‌ها تلقی می‌شود. مدیر می‌تواند از نحوه دستیابی افراد به وبگاه، از طریق موتور جستجو و یا از طریق کلیک بر روی پیوندی در یک وبگاه خارجی، داده‌هایی جمع‌آوری کند. آن‌ها ممکن است به تعداد بازدیدکننده‌هایی که مجددا بازمی‌گردند، یا سیستم اجرایی یا مرورگر مورد استفاده علاقه داشته باشند چرا که می‌تواند اطلاعات نوع کاربر را ارائه کند.

حتی در ساده‌ترین وبگاه‌ها مجموعه سنجه‌هایی وجود دارد که مدیر یک وبگاه می‌تواند به وضعیت تأثیر (یا عدم آن) در وبگاه خود بپردازد: این‌که بازدیدکننده‌های وبگاه به اندازه مورد انتظار نیست می‌تواند توجیه‌کننده این واقعیت باشد که آن بازدیدکننده‌هایی که بازدید داشته‌اند ابتدا کاربران خاصی بوده‌اند، همانگونه که سیستم عملیاتی اپل[1] نشان داد. تعداد بازدیدکننده‌های بیش از حد انتظار می‌تواند این واقعیت را توجیه کند که آن‌ها به صفحه وب با عبارت‌های موتور جستجویی دست یافته باشند، در حالی که در متن صفحه وارد شدند، ولی بطور تصادفی به محتوای اولیه صفحه رسیدند شدند. هرچه وبگاه پیچیده‌تر شود، ظرفیتی که برای سنجه‌ها در نظر گرفته می‌شوند نیز پیچیده‌تر می‌شود. صفحه های متعدد موجب می‌شود که سؤالاتی در خصوص اینکه کاربران به کدام صفحه مراجعه می‌کنند، از چند صفحه بازدید می‌کنند و میزان زمانی که در هر صفحه صرف می‌کنند، را به ذهن می‌آورد. هرجا که فضایی برای تعامل وجود داشته باشد، فضا برای سنجه‌های بیش‌تر نیز وجود دارد: چند بار به پرسشنامه برخط پاسخ داده شده است؟ چند ایمیل دریافت شده است؟ غالباً تعامل برای رسانه های اجتماعی در درجه نخست اهمیت و در وبگاه قدیمی در درجه دوم اهمیت قرار دارد.

وب‌نوشت‌ها

رسانه های اجتماعی به عنوان «برنامه های کاربردی اینترنت- محور که بر پایه‌های فنی و فکری وب ۲/۰ بنا نهاده شده‌اند و امکان ایجاد و تغییر محتوای تولید کاربر[2] را می‌دهند» (کاپلان و هانلین[3]، ۲۰۱۰، ۶۱) تعریف شده‌اند. هرچند این تعریف شامل وبگاه‌های شبکه‌های اجتماعی غالب همچون فلیکر، توییتر و فیسبوک و محتوای ایجاد شده مشترک مانند ویکی‌پدیا نیست، ممکن است موضوعات شخصی مانند وب‌نوشت‌ها یا حتی ویکی‌های میزبان را نیز بپوشاند. در وبگاه های قدیمی می‌توان از خواندن یا بازدید کاربران از محتوا اطلاعاتی بدست آورد، اما در رسانه های

1. Apple 2. User Generated Content 3. Kaplan and Haenlein

اجتماعی تنها انتشار محتوا مطرح نیست، بلکه بازخورد[1] تعامل با بازدیدکننده‌ها و دریافت نیز مطرح است.

وب‌نوشت‌ها یا وبگاه هایی که مرتباً به روز می‌شوند و صفحه‌ها را به ترتیب زمانی معکوس نشان می‌دهند، یکی از قدیمی‌ترین فناوری‌های رسانه های اجتماعی راه‌اندازی شده هستند که اولین بار در سال ۱۹۹۷ اجرا شدند و به دنبال اجرای نرم‌افزار وب‌نویسی وب-محور رایگان در ۱۹۹۹ انفجاری از ظهور وب‌نوشت‌ها رخ داد (بلوود[2]، ۲۰۰۰). آنها در حال حاضر توسط هر نوع سازمانی که می‌توان فکر کرد و توسط هر شخصی که می‌خواهد هر اطلاعات قابل تصوری از زندگی خود را به اشتراک بگذارد استفاده می‌شوند. کتابداران از فریب وب‌نویسی مصون نبوده‌اند و می‌توان وب‌نوشت‌هایی را مشاهده کرد که از هر بخشی ضمیمه خدمات اطلاعاتی و کتابداری شده اند. کتابداران بسیاری نیز میزبان وب‌نوشت‌های خود هستند. هر چند در سال‌های اخیر کاهشی در وب‌نویسی دیده شده است، شاید به دلیل افزایش وبگاه‌های شبکه‌های اجتماعی مانند توییتر و فیسبوک باشد، بهر حال مجموعه اصلی وب‌نوشت‌های کتابداران هم‌چنان پابرجاست (تورس-سالیناس، کبزاس-کلاویجو و روییز-پرز[3]، ۲۰۱۱).

تعریف گسترده از یک وب‌نوشت که در بالا استفاده شده یکی از چشمگیرترین ویژگی‌های متمایز وب‌نوشت را پنهان می‌کند: اظهارنظر. بیش‌تر افراد هنوز یک وب‌نوشت را به عنوان یک وب‌نوشت صرف نظر از اینکه آن وب‌نوشت اجازه اظهار نظر را می‌دهد یا خیر در نظر می‌گیرند. هر چند اظهار نظرهایی هستند که وب‌نوشت را فقط از یک سیستم مدیریت محتوای دیگر متمایز می‌کنند، نرم‌افزار وب‌نویسی بدون اظهار نظر تنها به تولیدکننده فایل‌های وبی یک ویرایش متن تبدیل می‌شود (ولو با تلاش بسیار قابل توجه). اظهار نظرها فرصتی را برای مباحثه با بازدیدکنندگان یک وبگاه فراهم می‌کنند، تا دریابند که رجوع آن ها بخاطر توجه زیاد به آن وبگاه

1. feedback 2. Blood 3. Torres-Salinas, Cabezas-Clavijo and Ruiz-Perez

است چرا که در جای خوبی قرار گرفته اند یا اینکه صرفاً به‌دلیل پذیرش همگان است. این امر می‌تواند بخاطر این باشد که ارزش اظهار نظرها برای کتابداران بسیار بیش‌تر از تعداد انتزاعی کاربران است: ارزش ۱۰ بازدیدکننده‌ای که اظهار نظر می‌گذارند و به تقاضای بازخورد پاسخ می‌دهند بیش از داشتن ۱۰۰ بازدیدکننده روزانه از یک وبگاه است بخصوص اگر اظهار نظرها با توجه به نوع اظهار نظرها دسته‌بندی شوند مانند «بازخورد مثبت برای منابع جدید»، «پیشنهاد اصلاحات» یا حتی «انتقادها». اگرچه در بیش‌تر موارد از انتقاد کم‌تر استقبال می‌شود، و در موارد بسیاری ممکن است بی‌پایه و اساس باشند، با این وجود گزینه اظهار نظر کانال مفیدی برای کاربران است تا بتوانند شکایت خود را عرضه کنند. در دراز مدت اظهار نظرها نیز می‌توانند به عنوان شاخصی برای بهبود خدمات باشند چرا که نسبت انتقادها با کاهش درصد کل اظهار نظرهای وب‌نوشت کاهش می‌یابد.

ویکی‌ها

ویکی، رسانه اجتماعی دیگری است که یک مؤسسه یا یک شخص ممکن است برای خودش میزبانی کند و بسیار کم‌تر از وب نوشت، حد اقل در وب باز، مورد استفاده قرار گرفته است. ویکی‌ها به عنوان «نرم افزارهای وب–محوری که که به تمام بازدیدکننده‌های یک صفحه اجازه تغییر محتوا را با ویرایش برخط صفحه در یک مرورگر می‌دهد» تعریف شده‌اند (ایبرساباچ، گلاسر و هگل[1]، ۲۰۰۸). معروف‌ترین نمونه ویکی‌ها ویکی‌پدیا است[2] : دانش‌نامه رایگانی که هر کسی می‌تواند آن را ویرایش کند. رویکرد تجمع منبع[3] به ایجاد اسناد وبی دارای مزیت و معایبی است. به عبارت دیگر اسناد ممکن است – به صورت بالقوه رایگان – در مقیاسی که به صورتی دیگر غیر ممکن باشند ایجاد شوند: ۴/۱۶۴/۸۷۱ مقاله انگلیسی که ویکی‌پدیا ادعا می‌کند اندازه دانش‌نامه قدیمی مانند *دانش‌نامه بریتانیکا*[4] را کم جلوه می‌دهد.

1. Ebersbach, Glaser and Heigl
2. http://en.wikipedia.org
3. crowd-sourced
4. *Encyclopaedia Britannica*

ویکی‌ها همچنین فرصتی را فراهم می‌کنند تا در خصوص یک موضوع دیدگاه‌های بسیار گسترده‌ای بدست آید. هرچند، مقاله‌ها برای حوزه‌هایی که مورد توجه کاربران است نگارش شده‌اند، نه حوزه‌هایی که امکان دارد از نظر یک کمیته تحریریه ارزشمند تلقی می‌شوند. این بدان معنی است که در مورد ویکی‌پدیا برخی حوزه‌ها گسترده‌تر از سایر حوزه‌ها هستند. برای مثال، سریال استار ترک[1] و کهکشان استارترک بسیار گسترده و دقیق هستند به همراه مقاله هایی برای هر قسمت، مسابقه، شخصیت و حتی سفینه فضایی!

ویکی‌ها می‌توانند در معرض خرابکاری و اطلاعات غلط قرار بگیرند، هرچند در کل به نظر می‌رسد یک ویکی خوب می‌تواند قانون لینوس[2] را از جامعه نرم‌افزاری متن باز رعایت کند: «حتی با چشمانی بزرگ، تمام مشکل‌ها سایه هستند[3]» (ریموند[4]، ۲۰۰۱، ۳۰)؛ چشمان بزرگ درباره خطاهای اطلاعاتی عمر کوتاهی ندارند. با این وجود، از آنجایی که احتمالاً تمامی مقاله ها بطور یکسان مورد توجه نیستند، و احتمال خرابکاری برخی مقالات بیش از سایرین است، تمام خرابکاری‌ها به سرعت کشف نمی‌شوند. احتمالاً معروف‌ترین مورد خرابکاری ویکی در مقاله جان سیگنتالر[5] روزنامه‌نگار تغییر یافت زیرا او در دو ترور جان اف. کندی[6] و رابرت کندی[7] مضنون به مشارکت بوده است و این خرابکاری برای چند ماه بدون تغییر باقی ماند[8]. با این وجود، علیرغم تخمین ۱۴/۰۰۰ ویرایش‌های خرابکاری همه روزه (مولاولاسکو[9] و روسو[10]، ۲۰۱۱) در مطالعاتی که چند سال پیش برای مقایسه انجام شد مقایسه مطالب علمی ویکی‌پدیا با *دانش‌نامه بریتانیکا* نتیجه خوبی داشته است (گیلس[11]، ۲۰۰۵) و در مقایسه‌ای با منابع مواد مخدر مداسکیپ[12] هیچ مشکل واقعی شناسایی نشد، هرچند ویکی‌پدیا به عنوان پایگاه داده ای ویرایش شده قدیمی کاربرد نداشته

1. Star Trek
2. Linus's Law
3. Given enough eyeballs all errors in information will be short lived
4. Raymond
5. John Seigenthaler
6. John F. Kennedy
7. Robbert Kennedy
8. (Journalism.org, 2005)
9. Mola-Velasco
10. Rosso
11. Giles
12. Medscape Drug Reference

است (کلاوسون و دیگران[1]، ۲۰۰۸). در مقایسه‌ای که از مقاله‌های تاریخی با سایر منابع دیگر انجام شد، نتیجه خیلی خوب نبود (رکتور[2]، ۲۰۰۸). گرچه، چنین مطالعاتی تأکید می‌کنند که انواع خاصی از مقالات در برابر اطلاعات غلط به نسبت انواع دیگر آسیب‌پذیری بیش‌تری دارند، اما به‌جای تعطیل کردن ظرفیت مشارکتی پروژه‌هایی چون ویکی‌پدیا، آنها نیاز به تعداد زیادی ویرایشگر مسئولیت‌پذیر را نشان می‌دهند. در سال‌های اخیر یک پیشقدمی برای بهبود رابطه بین ویکی‌مدیا[3]، سازمانی که ویکی‌پدیا را اداره می‌کند، و کتابداران کارگاه‌های ویکی و ویرایشگری در ۲۰۱۱ و ۲۰۱۲ بوده است (ویکی پدیا، ۲۰۱۳b).

تعداد ویکی‌های فعال در درون جامعه کتابداری، در مقایسه با سایر فناوری‌های رسانه‌های اجتماعی نسبتاً کم است. این مسئله می‌تواند ناشی از مشکلات نگهداری یک ویکی فعال باشد. موارد زیر برخی از آن‌هایی هستند که به صورت عمومی در دسترس بوده و چندین سال فعال بوده‌اند:

- *لایبرری ساکسس*[4] – راهنمای ایده‌ها و اطلاعات برای جامعه کتابداری. در زمان نگارش (تابستان ۲۰۱۳) صفحه اصلی ۱ میلیون بازدیدکننده داشته است.
- *ال‌آی اس ویکی*[5] – یک ویکی همگانی برای علوم کتابداری و اطلاع‌رسانی. صفحه ابتدایی آن بیش از ۳۰۰/۰۰۰ بازدیدکننده داشته است.
- *راهنماهای موضوعی اس‌جی‌سی‌پی‌ال*[6] – یک ویکی توسط کتابخانه عمومی کشور جوزف اس‌تی[7] که بجای کتابداران، جامعه عمومی را هدف قرار داده است. از صفحه ابتدایی تقریباً نیم میلیون بار بازدید شده است.

1. Clauson et al.
2. Rector
3. Wikimedia
4. Library Success (www.libsuccess.org)
5. *LISwiki* (http://liswiki.org)
6. *SJCPL subject guides* (www.libraryforlife.org/subjectguides)
7. St Joseph Country Public Library (www.mediawiki.org)

هر یک از این ویکی‌ها از نرم‌افزار ویکی‌مدیا[1] استفاده می‌کنند که از سال ۲۰۰۵ تاکنون فعال بوده است، سالی که رایتلی[2] (واژه پرداز[3] وب – محوری که بعداً توسط گوگل تصاحب شد و به گوگل داکز[4] پیوست) اجازه ویرایش مشترک اسناد را می‌دهد. این تحولات ویکی‌ها را تا حدودی تبدیل به یک محصول موفق کرد، چرا که بسیاری از محصولات اشتراکی که زمانی می‌توانند ارزشمند بوده باشند حالا می‌توانند با واژه پرداز مشترک ساده‌تر تحقق یابند. هم چنان جاهای بسیاری وجود دارد که نرم‌افزار ویکی، مانند نرم‌افزار اختصاصی کانفلوینس[5]، برای اهداف درون سازمانی استفاده می‌شود، و ویکی‌ها برای علم کتاب باز[6] ابزار ارزشمندی را عرضه می‌کنند (برای نمونه، آزمایشگاه بردلی[7] در پروژه علوم کتاب باز در رشته شیمی دانشگاه درکسل[8]). از این رو شاید لازم باشد کتابداران بسیاری سنجه‌های ویکی را برای اهداف تحلیل وبی یا حتی اهداف علم‌سنجی مدّ نظر داشته باشند.

مانند سنجه‌های وب‌نوشت‌ها، ویکی‌ها تعامل و همکاری مناسب بوده و راه‌اندازی سنجه‌ای برای ارزیابی آنها دارای اهمیت است. سنجه‌های خاص متناسب با هدف ویکی یا مقاله ویکی تغییرات چشمگیری دارند و تمرکز برخی مقاله‌ها و ویکی‌ها ممکن است بیش تر بر شخص باشد در حالی که سایرین شاید بیش‌تر مبتنی بر محتوا باشند. کتابداران کتابخانه‌های عمومی که مایلند بازخورد نحوه بهبود خدمات خود را بدانند ممکن است تمایل داشته باشند که بازخوردها را از تعداد هرچه بیش‌تر اعضای کتابخانه خود بگیرند، و تعداد مشارکت‌کننده‌ها بسیار مهم است. اگر هدف یک ویکی ساخت منبع ارزشمندی باشد، هرچند، در آن صورت تعداد صفحه‌ها (یا حداقل صفحه‌هایی با بیش از یک اندازه) می‌تواند مهم باشد. در سایر موارد مخصوصاً تعداد ویرایش‌هایی که یک صفحه دریافت می‌کند می‌تواند مهم باشد؛ ویلکینسون و هوبرمن[9] (۲۰۰۷) همبستگی بین کیفیت مقاله و تعداد ویرایش‌ها را شناسایی کردند.

1. MediaWiki
2. Writely
3. Word processor
4. Google Docs
5. Confluence
6. open notebook
7. Bradley Laboratory
8. Drexel University(http://usefulchem.wikispaces.com)
9. Wilkinson & Huberman

باید بین کیفیت مقاله و کیفیت هدفی که توصیف می‌شود تفاوت قائل شد: در زمان نگارش رمان *کد داوینچی*[1] در سال ۲۰۰۳، صفحه ویکی‌پدیا ۵۶۴۷ بار ویراش شده در حالی‌که برنده جایزه نویسنده مرد[2] همان سال، *ورنون خدای کوچک*[3] تنها ۱۷۳ بار ویرایش شده است. همانگونه که در فصل هفتم خواهیم دید، رفتار کاربر در ویکی‌پدیا می‌تواند در خصوص رفتار در دنیای واقعی اطلاعاتی ارائه کند.

این‌ها همه سنجه‌های داخلی بوده، و اطلاعات یا در خود وبگاه (مانند تعداد اظهارنظرها در وب‌نوشت یا تعداد ویرایش‌ها در ویکی) یا نرم‌افزاری راه‌اندازی کرده‌اند که اطلاعات خاصی را که می‌تواند تنها در اختیار دارندگان وبگاه باشد جمع‌آوری نماید(مانند بازدیدهای صفحه و منابع رفت و آمدی). سنجه‌های بیرونی بالقوه‌ای نیز وجود دارند که به تصمیم‌های وبگاه افراد در جمع‌آوری یا به اشتراک‌گذاری داده‌ها بستگی ندارند، ولی ممکن است در وبگاه‌ها در سراسر وب دیده شوند. این سنجه‌ها نه تنها اطلاعات بیش تری را برای دارندگان وب در اهداف تحلیل وبی ارائه می‌کنند، بلکه همچنین امکان بررسی‌های علم‌سنجی، وب‌سنجی و کتاب‌سنجی را هم می‌دهند.

اهمیت اولیه سنجه‌های بیرونی برای تحلیل وبی بخاطر فراهم کردن زمینه اطلاعاتی است که یک شخص در حال حاضر می‌تواند برای خودش جمع‌آوری کند. کتابداران ممکن است در حال حاضر بدانند که وب‌نوشت‌های آنها ۵۰۰ بازدید در روز دارد، اما اگر آن‌ها از تعداد بازدیدهای سایر مؤسسه‌های مشابه اطلاعاتی نداشته باشند نمی‌دانند این عدد خوب است یا بد. آن‌ها ممکن است درباره اینکه چرا آنها در حال دریافت رفت و آمد هستند نیز اطلاعات کمی داشته باشند: آیا محتوای مورد بحث دارای ویژگی‌های مثبت است یا منفی؟ آیا یک وبگاه یا محتوا به یک گزارش خبری جاری پیوند می‌خورد یا با یک وبگاه فوق‌العاده با رفت و آمدهای سطح بالا ارتقا یافته است؟ هرچند برخی از این اطلاعات از طریق سنجه‌های داخلی موجود

1. *The Da Vinci Code* 2. Man Booker Prize 3. *Vernon God Little*

هستند، مرورگرها بیش از پیش گام‌هایی برداشته‌اند تا کاربران بتوانند رفتار برخط خود را پنهان کنند (سولیوان، ۲۰۱۳).

سنجه‌های بیرونی نیز فرصتی را برای بررسی‌های وب‌سنجی و علم‌سنجی فراهم می‌کنند، که به همان اندازه اهمیت دارد. در بیش تر موقعیت‌ها وبگاه ها، وب‌نوشت‌ها و ویکی‌ها که کتابداران میزبان آنها هستند نمی‌توانند مجموعه داده های خیلی بزرگی را برای انجام محاسبات وب‌سنجی یا علم‌سنجی عرضه کنند. درعوض لازم است الگوهای زیادتری از داده‌ها ترسیم شود.

سنجه‌های داخلی

سنجه‌های داخلی تمرکز اصلی این کتاب نیستند زیرا این سنجه‌ها تاکنون بطور مفصل مورد بحث بسیاری از کتاب‌های دیگر بوده‌اند، از جمله در کتاب «استفاده از تحلیل وبی در کتابخانه» از کیت مارک (۲۰۱۱). آنچنان که در عنوان این کتاب مطرح می‌شود جامعه کتابداران بویژه مورد هدف قرار داده شده است. همچنین احتمالاً این موضوع می‌تواند برای بسیاری از متخصصان کتابداری مطرح باشد که برای جمع‌آوری داده‌ها گزینه کمی وجود دارد چرا که شاید داده‌ها مربوط به سیستم‌های مدیریت محتوایی باشند که متعلق به سازمان بزرگ‌تری است، و حتی ممکن است به اطلاعاتی محدود باشند که واحد خدمات فناوری اطلاعات بخواهند آنها را به به اشتراک بگذارند.

در اینجا تنها گذری کوتاه بر سنجه‌های داخلی خواهیم داشت. ابتداً، به منظور ایده دهی به کتابداران در خصوص نوع داده‌ای که ممکن است جمع‌آوری شود، نحوه جمع‌آوری آن و بعضی مسائلی که در خصوص مجموعه چنین داده‌هایی وجود دارد مروری کوتاه بر گوگل آنالیتیکز و تحلیل لاگ[1] داریم. اثر حاضر برای ارائه مشاوره در اجرای یک برنامه تحلیل وبی گسترده طراحی نشده است بلکه می‌خواهد به کتابداران ایده بدهد که چه چیزی را می‌شود انجام داد، و چنانچه کتابداران به

1. Log

داده‌هـایی دسترسـی ندارنـد مـی‌تواننـد منطقـاً از خـدمات فنـاوری اطلاعـات خـود درخواست کنند تا به آن‌ها دسترسی یابند.

این موضوع با بحث مختصری از سنجه‌های ویکی دنبال می‌شـود. داده‌هـای مـورد دسترسی به نرم‌افزار مورد استفاده بسیار وابسته هستند، هرچند ابزارهـایی سـاخته شـده است که سنجه‌های وبی (مانند بازدیدهای صفحه) مقرر را برای محتوای مبتنی بر ویکی و مفاهیم ویکی- محور بیش‌تری عرضه می‌کند (مانند تعداد ویرایش‌ها یا ویرایشگرها).

گوگل آنالیتیکز و تحلیل لاگ

تحلیل وبی عموماً بر تحلیل لاگ یا برچسب‌زنی[1] صـفحه تمرکـز دارد. تحلیـل لاگ از لاگ‌های سروری که میزبان داده‌ها هستند استفاده می‌کند تا در خصـوص دسـتیابی بـه فایل‌هایی که توسـط یـک مرورگـر درخواسـت شـده‌انـد اطلاعـاتی بدسـت آورنـد. برچسب زنی صفحه شامل وارد کردن یک قطعه کد جاوا اسکریپت[2] در داخل هر یـک از صفحه‌های وبی است، به نحوی که یک برنامه تحلیل وب بـا داده‌هـا عرضـه شـود. برنامه‌های تحلیل‌های برچسب‌زنی صفحه دارای واسط‌های کاربر پسند هستند کـه بـر بازدیدهای صفحه و ردگیری رفتار مشتری‌ها تمرکز دارنـد، و در حـال حاضـر بسـیار مورد استفاده قرار می‌گیرند. گوگل آنـالیتیکز بـه تنهـایی، بـزرگ‌تـرین نمونـه تحلیـل برچسب‌زنی صفحه است و در حاضر بر ۵۷/۷ درصد تمـام وبگـاه‌هـا موجـود اسـت (W3Techs، ۲۰۱۳). برچسب زنی صفحه‌ها اساساً امکان بررسـی تحلیـل‌هـای وبـی را حتی در مواردی که مالک وبگاه نمی‌تواند به لاگ‌هـای سـرور دسترسـی داشـته باشـد ارائه می‌دهـد، بـرای نمونـه در میزبـانی بیرونـی یـک وب‌نوشـت یـا وبگـاه. هرچنـد همانگونه که در سراسر این کتاب گفته شـده اسـت هـیچ ابـزار وبـی عـالی نیسـت و برچسب زنی صفحه‌ها محـدودیت‌هـای خـودش را دارد. مشـهودتـرین نکتـه دربـاره برچسب‌زنی صفحه این است که اگر چنانچه کاربران جـاوا اسـکریپت را در مرورگـر خود غیر فعال کرده باشند، این امر باعث می‌شـود کـه بازدیدکننـده در برنامـه تحلیـل

1. tagging 2. JavaScript

رؤیت نشود، یا همینطور وقتی که شناسه شناسایی کاربران را غیر فعال کنند، باعث می‌شود بازدیدکننده‌هایی که بر می‌گردند بازدیدکننده جدید به نظر بیایند. گوگل همچنین پلاگ-این‌هایی[1] دارد که به افراد اجازه می‌دهند تا در مورد اینکه آیا گوگل آنالیتیکز می‌تواند داده‌های آنها را جمع‌آوری کند[2]. تصمیم بگیرند. تحلیل‌های برچسب‌زنی صفحه قادر نیست که به یک مؤسسه برای پیوند درون برنامه‌ای تصویرهای خود هشدار دهد، عملی که کد برنامه‌نویسی در یک وبگاه قرار می‌دهد تا یک تصویر را از وبگاه دیگری نشان دهد.

از آنجایی که گوگل آنالیتیکز رایج‌ترین نوع تحلیل‌های وبی داخلی است، یک بازنگری اجمالی ارائه می‌کند تا درباره نوع داده‌های موجود نشانه‌ای بدهد، هرچند باید به خاطر داشت که گزینه‌های رایگان دیگری مانند پی‌ویک[3] و تحلیل‌های وب باز[4] در دسترس هستند. مؤسسات بزرگی احتمالاً به این‌ها توجه خاصی دارند چرا که در غیر این صورت مجبور خواهند شد که به خدمات فوق‌العاده گوگل آنالیتیکز بپیوندند. البته هرچند احتمالاً ۱۰ میلیون صفحه بدون محدودیت ماهانه برای بیش‌تر مؤسسه‌های کوچک‌تر کافی خواهند بود.

پس از آن‌که حساب گوگل آنالیتیکز راه‌اندازی شد و جاوا اسکریپت در صفحه های مربوطه درج شد، ممکن است کتابداران مشتاقانه بخواهند که اطلاعات ارزشمندی از فعالیت کاربر خود بدست آورند. هرچند کاربرانی که برای اولین بار از گوگل آنالیتیکز استفاده می‌کنند ممکن است بلافاصله از حجم داده موجود و روشی که می‌توان آن‌ها را به خوبی به هم متصل کرد احساس شکست کنند. کوتاه سخن اینکه برای هر بازدیدی، داده‌های حول و حوش بازدیدکننده، محتوایی که از آن بازدید کرده‌اند، و منبع رفت و آمد آنها بدست می‌آید. داده‌های بازدیدکننده نه تنها شامل زمان درخواست صفحه آن بازدیدکننده است بلکه شامل اینکه از کدام نقطه دنیا، مرورگر، سیستم عملیاتی، حتی کیفیت صفحه نمایش و نسخه فلش موجود نیز

1. Plug-ins
2. (http://tools.google.com/dlpage/gaoptout)
3. Piwik(http://piwik.org)
4. Open Web Analytics(www.openwebanalytics.com)

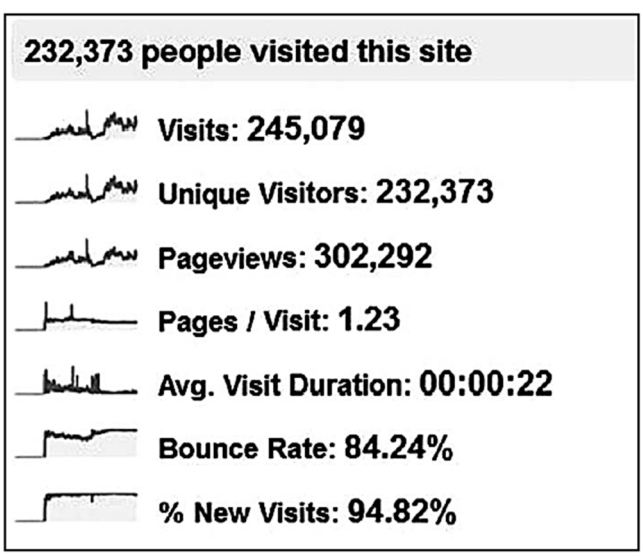
می شود. این موضوع، بیش از آنچه که برای اغلب افـراد و سـازمان‌هـا مـورد اسـتفاده است، برای گزارش‌ها و تصورات بسیار خاص امکان‌پذیر است. بـرای نمونـه، شـکل ۴-۱ تفکیک ۴۳۴ بازدیدکننـده‌ای کـه وب‌نوشـت مـن را بـر اسـاس کیفیـت صـفحه نمایشی که افراد ساکن قاهره[1] داشته است نشان می‌دهد. این حساب‌های کاربری تنهـا برای ۱۸٪ درصد از کل بازدیدهایی است که وب‌نوشت من داشـته اسـت، و مخاطـب محتوای من بطور اخص جامعه مصر نبوده است. مـن احتمـالاً زمـان زیـادی را بـرای تحلیل دقیق این جزئیات خاص نمی‌گـذارم، امـا از آنجـایی کـه هـزاران عـدد از ایـن جدول‌ها با چند کلیک در دسترس خواهند بود سنجه‌ها به سرعت می‌توانند بـه عامـل بازدارنده‌ای تبدیل شوند.

	Town/City	Screen Resolution	Visits ↓	Visits
1.	Cairo	1024x768	150	34.56%
2.	Cairo	320x480	53	12.21%
3.	Cairo	1280x800	47	10.83%
4.	Cairo	1366x768	43	9.91%
5.	Cairo	800x600	20	4.61%
6.	Cairo	1440x900	16	3.69%
7.	Cairo	1152x864	12	2.76%
8.	Cairo	1280x1024	11	2.53%
9.	Cairo	768x1024	8	1.84%
10.	Cairo	1600x900	5	1.15%

شکل ۴-۱: تعداد بازدید کنندگان blog.webometrics.org.uk از قاهره بر اساس کیفیت صفحه
[گوگل و آرم گوگل که به عنوان نماد تجاری برای شرکت گوگل ثبت شده‌اند، استفاده با اجازه]

1. Cairo

در بیش‌تر موارد سطح بررسی اجمالی از پنج گزارش استاندارد برای کاربرها کافی است. این‌ها نقطه‌های شروع درکندوکاو داده‌ها هستند که از پنج دیدگاه مختلف اطلاعاتی در خصوص داده‌ها عرضه می‌کنند: مخاطبان، منابع رفت و آمد (رفت و آمد)، محتوا، مذاکره‌ها و فعالیت‌های هم‌زمان[1]. در صورت گزینش یکی از این گزارش‌ها، کاربر می‌تواند یک چشم‌انداز کلی از سنجه‌های وابسته اصلی و ابرپیوندها برای کاوش بیش‌تر در داده‌ها بدست آورد. برای نمونه، گزینش گزارش مخاطبان در خصوص تعداد بازدیدها، بازدیدکننده‌های منحصر به فرد، بازدیدهای صفحه، صفحه‌های بازدید شده به ازای هر بازدید، مدت متوسط هر بازدید، نرخ پرش[2] و درصد بازدیدهای جدید یک چشم‌انداز کلی ارائه می‌کند. هم‌چنین گزینه‌ای برای کندوکاو بیش‌تر داده‌ها بر اساس ویژگی‌های جمعیت‌شناسی‌های مخاطبان می‌دهد (شکل ۴-۲).



شکل ۴-۲: بررسی اجمالی مخاطبان در گوگل آنالیتیکز

[گوگل و آرم گوگل به عنوان نماد تجاری برای شرکت گوگل ثبت شده‌اند، استفاده با اجازه]

1. real-time 2. bounce rate

مذاکره‌ها این امکان را به کاربر تحلیل وب می‌دهد تا اهداف خاصی را که در نظر دارند تنظیم کند. برای نمونه، تعداد خاصی کاربر که می‌خواهند به یک نشانی وبی خاص (یا مجموعه‌ای ازآدرس های اینترنتی) برسند، تخصیص زمان خاصی برای هر بازدید، بازدید از تعداد خاصی صفحات در یک بازدید، یا بازدید از یک رویداد (که در آن کد اضافی برای دنبال کردن تعامل‌ها با عناصر وبگاه وارد شده است، مانند فلش یا عناصر AJAX).

علاوه بر گزارش‌های استاندارد، گوگل آنالیتیکز سه روش اضافی برای دسترسی آسان به داده‌ها فراهم می‌کند:

- *داشبوردها*[1] – امکان به نمایش درآمدن خلاصه گزارش‌های مختلف در یک صفحه را می‌دهد.
- *میان‌برها*[2]– معادل یادآورها[3] در گوگل آنالیتیکز، که به کاربران امکان دسترسی سریع به خاص‌ترین یافته‌ها را می‌دهد.
- *رویدادهای هوشمند*[4] – امکان هشدارهای خودکار برای موقعیت‌های خاص را می‌دهد، مانند یک سقوط یا افزایش تعداد مشاهده‌های صفحه یا بازدیدها.

با این همه داده قابل رویت، باید کتابداران فقط روی چشمگیرترین ابعاد تمرکز کرده و داشبوردها، میان‌برها و رویدادهای هوشمند مناسبی را انتخاب کنند. اهداف با توجه به نوع کتابخانه و محتوای خاصی که برچسب‌های صفحه درون آن قرار داده شده‌اند تغییر می‌کنند. کتابداران کتابخانه‌های عمومی مجبورند به بازدیدهای محدوده بومی خود توجه ویژه‌ای داشته باشند، در حالی‌که توجه ویژه کتابداران کتابخانه‌های دانشگاهی ممکن است به داده‌هایی باشد که از طریق شبکه های دانشگاه به وبگاه دسترسی پیدا می‌کنند. در مواردی که یک وب‌نوشت تنظیم شده باشد ممکن است به این موضوع بپردازند که آیا بازدیدکننده‌ها از صفحه فرود[5] (نشست) خودشان به

1. *dashboards* 2. *shortcuts* 3. bookmarks 4. *intelligence events* 5. landing

صفحه اصلی وب‌نوشت رسیده‌اند یا خیر. جایی که برنامه آموزشی به صورت برخط قرار داده شده است، احتمال دارد به این موضوع بپردازند که آیا همه صفحه‌ها به ترتیب دیده شده‌اند یا اینکه افراد در طول مسیر با آنها برخورد کرده‌اند.

زمانی که سنجه‌های مناسب انتخاب، و بخش‌های بهینه‌سازی شناسایی شدند، شاید لازم باشد که اقدامات مناسب زیر انجام شود: مواردی که بازدیدهای یک صفحه بسیار کم است، لازم است آن صفحه از سایر صفحه‌ها برجسته‌تر شود؛ مواردی که بازدیدهای صفحه برای کل وبگاه کم است ممکن است ارتقا وبگاه از طریق رسانه‌های اجتماعی یا مکان دیگری الزامی باشد؛ جایی که بازدیدکننده‌ها مرتباً از مجموعه‌ای از صفحه‌های پشت سر هم جدا می‌شوند، ممکن است لازم به بررسی است که آیا ساختار یا محتوا صفحه می‌توانند ارتقا یابند تا افراد از طریق مجموعه صفحه‌ها تعقیب شوند. در بسیاری از موارد مناسب‌ترین اقدام این است که هیچ اقدامی نشود. کتابداران ممکن است یک روز وقت بگذارند و تلاش کنند تا دریابند که چرا اعداد در یک بخش خاص قرار گرفته‌اند، قبل از آنکه احتمال تغییر همه چیز در روز بعد وجود داشته باشد. ویژگی‌هایی مانند گزارش بهنگام گوگل آنالیتیکز، که از کاربران حاضر در وبگاه اطلاعات زنده‌ای عرضه می‌کند، ترغیب یک مشغله ذهنی با حال، در صورتی که در بیش‌تر موارد به کتابداران توصیه می‌شود به روندهای طولانی‌تر توجه کنند.

به عنوان قانونی کلی انسان ها، چه کارکنان کتابخانه باشند و چه کاربران کتابخانه، دوست ندارند مورد ارزیابی قرار بگیرند، لذا برای نشان دادن مزایای هر نوع ارزیابی باید گام‌هایی برداشته شود. این موضوع در گوگل آنالیتیکز بسیار با اهمیت است، چرا که از داده‌های موجود به سادگی استفاده نمی‌کند، بلکه داده‌های جدیدی جمع‌آوری می‌کند. کاربران همواره می‌توانند کناره‌گیری کنند، بطوریکه اخیراً به موضوع بخشنامه اروپایی حفظ حریم الکترونیکی تبدیل شده است، که معمولاً بیش‌تر به عنوان قانون کوکی[1] شناخته می‌شود. در زمان نگارش کتاب حاضر (تابستان

1. Cookie law

۲۰۱۳) موافقت ضمنی دفتر اطلاعات عضو کمیسیون[1] در بریتانیا کافی شناخته شد (۲۰۱۲) چراکه با سیاست روشن درباره حفظ حریم خصوصی و کوکی در یک وبگاه قابل دستیابی می‌شود. هرچند، متخصصان کتابداری شاید مایل باشند یک گام جلوتر رفته و قبل از جمع‌آوری داده‌ها کوکی‌ها را پذیرا باشند، گرچه این موضوع سوگیری داشته و داده‌ها را در جهت نفع افراد موافق پذیرش کوکی سوق می‌دهد.

اگرچه برچسب زنی صفحه هدف اصلی تحلیل وبی از طریق خدماتی مانند گوگل آنالیتیکز است، اما آگاهی از نوع اطلاعاتی که از طریق چنین خدماتی در دسترس قرار می‌گیرد در برخی مطالعات وب‌سنجی نیز می‌تواند ارزشمند باشد. اینکه داده‌های جمع‌آوری شده از یک وبگاه واحد احتمالاً در مطالعات وب‌سنجی بسیاری مورد استفاده نیستند می‌تواند به این علت باشد که در یک مطالعه وب‌سنجی از میزبانی وبگاه‌ها بخاطر اینکه می‌توانند درخواست‌های اطلاعاتی داشته باشند استفاده می‌کند.

ویکی‌سنجی

ویکی‌ها، مانند وب‌نوشت‌ها و سایر محتواهای میزبان، ممکن است از گوگل آنالیتیکز برای ارزیابی تأثیر یک وبگاه استفاده کنند. اگر کسی از یک ویکی که هدفش عموم مردم است و برای روزآمدی آن کارکنان باید وقت زیادی صرف کند بازدید نکند، قطعاً سؤال‌هایی پرسیده خواهد شد درباره اینکه آیا استفاده خوبی از آن می‌شود یا خیر. همچنین نحوه همکاری افراد با یک ویکی موضوع مورد توجه است، و اینکه آیا کاربران فقط از محتوا بازدید می‌کنند یا برای مشارکت در صفحه نیز تلاش می‌کنند. قابلیت دسترسی به این اطلاعات بر اساس نرم‌افزار، و همچنین پلاگ‌این‌های جانبی، اضافه شده‌ها[2] و سایر نرم‌افزارهایی که توسط جامعه کاربر ساخته شده‌اند تفاوت دارد. برای نمونه، استت‌مدیاویکی[3] امکان ویرایش‌های مختلف برای یک صفحه را داده و کاربران را از سراسر یک ویکی منسجم می‌کند. گرچه ماهیت کامل این داده‌ها بر اساس نرم‌افزار متغیر است، اما از دیدگاه کاربران و صفحه‌های

1. Information Commissioner's Office 2. extensions
3. StatMediaWiki (http://statmediawiki.forja.rediris.es)

ویکی می‌تواند مورد توجه قرار گیرند. مانند همیشه سنجه‌های دقیق که مهم هستند به هدف ویکی بستگی دارند.

سنجه‌های صفحه–محور[1] عموماً اطلاعات تعداد ویرایش‌هایی که یک صفحه داشته است، تعداد ویرایشگرهای مختلفی که یک صفحه را ویرایش کرده‌اند، و نحوه آخرین ویرایش صفحه را نشان می‌دهد. تعداد ویرایش‌ها معمولاً شاخصی است که نشان می‌دهد یک موضوع مورد توجه است، و تعداد بالای ویرایش همراه با ارتقا کیفیت مقاله است. در مورد یک موضوع بحث برانگیز، در راستای تلاش برای ایجاد توازن، می‌تواند ناشی از این باشد که تعداد زیادی از ویرایشگران مختلف آن را معتبر تلقی می‌کنند.

سنجه‌های کاربر– محور[2] می تواند اطلاعات تعداد ویرایش‌هایی که هر کاربر انجام داده و تعداد صفحه‌هایی که ویرایش کرده‌اند را نشان بدهد. در مواردی که همکاری با یک ویکی به اندازه پیش‌بینی شده نیست، باید بررسی شود که چه کسانی همکاری می‌کنند. برای مثال، با وجودی‌که ماهیانه ۲۵۰/۰۰۰ حساب کاربری جدید برای ویکی‌پدیا ایجاد می‌شود، تنها ۳۰۰/۰۰۰ کاربر بیش از ده ویرایش انجام داده‌اند (ویکی‌پدیا، c۲۰۱۳).

با اجرای گوگل آنالیتیکز در یک وبگاه، کتابداران شاید بخواهند اهداف و اقدامات را تعیین کنند اما بجای افزایش ارتقای محتوای وب و افزایش رفت و آمد آن، به دنبال راه‌هایی برای افزایش همکاری ویرایشگرها باشند. گرچه، بهترین راه افزایش مشارکت کتابداران هدایت و رهبری باشد، اما برای روزآمدسازی یک ویکی بزرگ شاید تلاش قابل توجهی لازم باشد. موضوع ویکی‌سنجی در فصل‌های پنجم و هفتم مجدداً بازبینی شده، و به ویکی‌سنجی‌ها برای پژوهش‌های وب‌سنجی، کتاب‌سنجی و علم‌سنجی با رجوع به مورد خاص ویکی‌پدیا می‌پردازیم.

1. Page-centric 2. user-centric

سنجه‌های بیرونی

این بخش نگاهی دارد به دو نوع منبع اطلاعاتی اصلی که ممکن است برای سنجه‌های ارزیابانه بیرونی استفاده شوند: ابزارهایی که اطلاعات رفتار کاربران را در مرورگرهای اطلاعاتی می‌دهند، و ابزارهایی که اطلاعات باقیمانده از جای پای کاربر را در محیط برخط می‌دهند. در برخی موارد از یک ابزار برای هر دو مورد استفاده می‌شود مانند آلکسا، به عنوان خدمات اطلاعاتی وب است که اطلاعات رفت و آمد کاربران و تعداد پیوندهای داخلی وبگاه را نشان می‌دهد.

رفتار کاربر

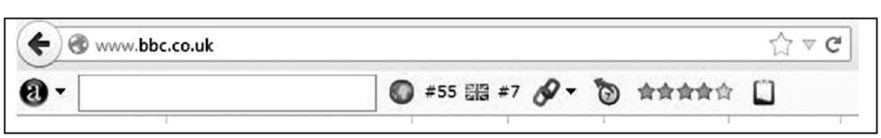
رفتارهای کاربری فراوانی وجود دارند که ردهای برخط روشنی از خود بر جای می‌گذارند. اگر کسی ۵۰ روز از تعطیلات خودش را در فلیکر بارگذاری کند، در آن صورت (اگر تنظیم‌های حریم خصوصی خاصی را اعمال نکرده باشند) ۵۰ عکس بیش‌تر وجود خواهد داشت که مردم می‌توانند آن‌ها را به صورت برخط ببینند. البته بیش‌تر رفتارهای کاربران پنهان باقی می‌ماند. بجز چند وبگاه شبکه‌های اجتماعی که صریحاً تعداد بازدیدها یا لایک‌هایی که برخی چیزها داشته اند را تبلیغ می‌کنند، بیش‌تر فعالیت کاربران برخط ما از دید عموم پنهان می‌مانند. با این وجود، ابزارهایی وجود دارند که می‌توانند اطلاعات منسجمی ارائه بدهند مانند آن‌هایی که اطلاعاتی در خصوص رفت و آمد وبی و رفتار ارائه می‌دهند.

آلکسا

چندین خدمت برخط وجود دارد که اطلاعات رفت و آمد وبی را، نه تنها از یک وبگاه بلکه از سرتاسر وب تأمین می‌کنند: کامپت[1]، کوانتکست[2]، و آلکسا. گرچه تمام این وبگاه‌ها حساب‌های پولی دارند اما تا یک حدی خاصی هم اطلاعات را رایگان می‌دهند مانند ارائه اطلاعات رفت و آمدی فقط ایالات متحده. با این وجود، چنین وبگاه‌هایی می‌توانند اطلاعات ارزشمندی برای تحلیل وبی و اهداف وب‌سنجی ارائه کنند.

1. Compete (www.compete.com) 2. Quantcast(www.quantcast.com)

روش گردآوری داده‌ها برای هر یک از این وبگاه‌ها متفاوت است. بـرای نمونـه، کوانتکست یک خدمت تحلیل وبی فراهم می‌کند، و داده‌های مبهمی را از کسـانی کـه از این خدمات استفاده مـی‌کننـد جمـع‌آوری مـی‌کنـد (کوانکسـت، ۲۰۱۳). داده‌هـای کامپـت از مجموعـه‌ای از ۲ میلیـون کـاربر اینترنتـی بـرخط جمـع‌آوری شـده اسـت (کامپت، ۲۰۱۳). داده‌های آلکسا از کاربران نـوار ابـزار[1] آلکسـا جمـع‌آوری مـی‌شـود، سپس داده ها، برای اصلاح سـوگیری‌هـا در داده‌هـای نمونـه هنجارسـازی مـی‌شـوند (آلکسا، ۲۰۱۳a). هر کدام از روش‌هـای نمونـه‌بـرداری محـدودیت‌هـای خودشـان را دارند: داده های کوانکست بـرای آن وب‌گـاه‌هـایی کـه از تحلیـل کوانکسـت اسـتفاده نمی‌کنند کم‌تر قابل اعتماد است، کامپت خیلی زیاد به ایـالات متحـده مبتنـی اسـت، و آلکسا به نفع آنهایی که بیش‌تر حتمال دارد که نوار ابزار آلکسا نصـب کننـد سـوگیری می‌کند (تحلیل‌گرهای وبی و بهینه‌سازهای موتور جستجو) (شـکل ۴-۳ را ببینیـد). بـا این وجود، مانند سنجه‌های وبی هرچند شاید آنها پاسخ‌هـای دقیقـی ندهنـد، امـا ایـن وبگاه‌ها می‌تواننـد شـاخص‌هـای ارزشـمندی عرضـه کننـد، بـویژه بـرای رایـج‌تـرین وبگاه‌هایی که احتمالاً داده های آن درست تر است. آلکسا امکان دسترسی رایگان را بـه جامع‌ترین طیف داده، با برخی داده‌های محدود اضافی برای اشخاصی کـه نـوار ابـزار آلکسا را نصب می‌کنند فراهم می‌کند. مناسب‌تـرین گزینـه بـرای کتـابی در خصـوص سنجه‌های وبی با مخاطبان کتابدار، آلکساندریا[2] آلکسا نـام دارد (کوییـنت[3]، ۱۹۹۸). در سال ۲۰۱۳ دسترسی خودکار به داده ها در ازای هر استفاده یک پرداخـت بـوده اسـت یعنی ۰/۱۵ دلار برای هر ۱۰۰۰ درخواست.



شکل ۴-۳: نوارابزار آلکسا

1. toolbar 2. Alexandria 3. Quint

آلکسا بجای اینکه تعداد مطلق بازدیدکننده را نشان بدهد، اطلاعات مختلفی ارائه می کند مانند میزان رفت و آمد – بر اساس تعداد بازدیدهای صفحه و تعداد کاربران؛ درصد کامیابی[1] – درصد کاربران جهانی اینترنت که از یک وبگاه بازدید می‌کنند؛ درصد بازدید صفحه– درصد تخمینی بازدیدهای جهانی صفحه در یک وبگاه مشخص؛ بازدیدهای صفحه به ازای هر کاربر – تعداد تخمینی بازدیدهای صفحه‌ای خاص به ازای هر کاربر برای یک وبگاه بخصوص؛ پرش – نسبت بازدیدهایی که شامل بازدید یک صفحه است؛ زمان صرف شده در وبگاه – زمان تخمینی صرف شده در یک صفحه؛ و درصد جستجو – درصد تخمینی بازدیدهای روزانه که از یک موتور جستجو می‌آیند. نکته مهم اینکه، برخلاف تحلیل لاگ و برچسب زنی صفحه، و با وجود محدودیت‌ها، امکان مقایسه بین وبگاه‌ها وجود دارد. برای نمونه، آلکسا تنها داده‌های سطح دامنه را میسّر می‌سازد، و داده‌ها به‌جای مقدار مطلق رتبه‌بندی می‌شود. بدین ترتیب، نوسانات کم‌تری وجود داشته و نمی‌توان داده‌ها را از چندین دامنه اضافه کرد. این محدودیت‌ها را می‌توان در هنگام مقایسه داده ها از کتابخانه‌های ملی پنج کشور بزرگ اتحادیه اروپا دید.

فکر کردن راجع به این موضوع که کتابدارانی علاقه‌مندند که کتابخانه‌های خود را با استفاده از آلکسا با مؤسسه‌های مشابه و با برای اهداف تحلیل وبی مقایسه کنند، آسان است. جدول ۴–۱ رتبه رفت و آمد آلکسا و درصد دسترسی کتابخانه‌های ملی پنج کشور بزرگ اتحادیه اروپا را نشان می‌دهد. درحالی که این جدول به وضوح نشان می‌دهد که بیش‌ترین رفت و آمد را کتابخانه ملی فرانسه[2] دارد و بعد از آن کتابخانه بریتانیا[3]، وضعیت کتابخانه‌های ملی آلمان، اسپانیا و ایتالیا به روشنی مشخص نیست. کتابخانه ملی آلمان[4] در رتبه بالاتری نسبت به کتابخانه ملی اسپانیا[5] قرار دارد اما موقعیت کتابخانه ملی ایتالیا خیلی روشن نیست.

1. reach-percentage
2. National Library of France
3. British Library
4. German National Library
5. National Library of Spain

جدول ۴-۱. رتبه رفت و آمد آلکسا برای کتابخانه‌های ملی پنج کشور بزرگ عضو اتحادیه اروپا در ماه مارس ۲۰۱۳

کتابخانه ملی	رتبه رفت و آمد آلکسا	درصد تحقق
کتابخانه ملی آلمان (www.dnb.de)	۹۴/۲۵۸	٪۰/۰۰۱۵۴
کتابخانه ملی فرانسه (www.bnf.fr)	۱۵/۲۷۲	٪۰/۰۰۷۹
کتابخانه بریتانیا (www.bl.uk)	۲۵/۵۵۰	٪۰/۰۰۵۶۶
کتابخانه مرکزی ملی رُم (www.bncrm.librari.beniculturali.it)	۴۵/۰۸۴	٪۰/۰۰۳۴۰
کتابخانه مرکزی ملی فلورانس (www.bncf.firenze.sbn.it)	۹۱/۷۸۸	٪۰/۰۰۱۴۲
کتابخانه ملی اسپانیا (www.bne.es)	۹۷/۹۸۰	٪۰/۰۰۱۳۷

در مورد اول، ایتالیا نه یک کتابخانه ملی بلکه دو کتابخانه ملی یکی در فلورانس و یکی در رم دارد. در نتیجه برای تجمیع یا عدم تجمیع داده‌های آنها باید تصمیم‌گیری شود. گرچه رتبه‌بندی رفت و آمد دو وبگاه نمی‌توانند ترکیب شوند، اما درصد کامیابی تا حدودی امکان‌پذیر است، چرا که بخشی از بازدیدکننده‌هایی که از هر دو وبگاه بازدید کرده‌اند دو بار شمرده می‌شوند. هرچند، مشکل بزرگ‌تر در شناسایی رفت و آمد کتابخانه‌های ملی ایتالیا این است که نه کتابخانه ملی مرکزی فلورانس[1] و نه کتابخانه ملی مرکزی رم[2] نام دامنه مخصوص به خود ندارند: کتابخانه ملی مرکزی فلورانس از یک زیر دامنه خدمات کتابخانه ملی[3] و کتابخانه ملی مرکزی رم از یک زیر دامنه وزارت فرهنگ ایتالیا[4] استفاده می‌کند. آلکسا جزئیات اطلاعات اینکه کاربران از کدام زیر دامنه بازدید کرده‌اند را نشان می‌دهد، اما از هیچ یک از کتابخانه‌های ملی ایتالیا ارزیابی مناسبی ارائه نمی‌دهد. اگرچه می‌توان متوجه شد که ۱۷/۲۱٪ بازدیدکننده‌های sbn.it از زیر دامنه Firenze.sbn.it بازدید می‌کنند، اطلاعات دامنه bncf.firenze.sbn.it وجود ندارد. به همین صورت در حالی‌که می‌توان فهمید که ۶/۳۳٪ از بازدیدکننده‌های beniculturali.it از زیردامنه library.beniculturali.it بازدید می‌کنند، اطلاعات bncrm.librari.beniculturali.it موجود نیست.

محدودیت‌های زیر دامنه و محدودیت‌های روش‌شناختی جمع‌آوری داده ها می‌تواند ناشی از اطلاعاتی باشد که آلکسا می‌تواند برای بسیاری از کتابخانه‌ها فراهم

1. National Central Library of Florence
2. National Central Library of Rome
3. National Library Service(www.sbn.it)
4. Italian Ministry of Culture(www.beniculturali.it)

کند. کتابخانه‌ها اغلب بخشی از سازمان‌ها یا موجودیت‌های بزرگ‌تر هستند و رفت و آمدی که دریافت می‌کنند معمولاً آنقدر چشمگیر نیستند، تا آلکسا نتایج موثقی ارائه کند. در هر صورت وبگاه‌هایی که میزان قابل توجهی رفت و آمد دارند، هم چنین نشانی وبی مناسبی دارند، آلکسا می‌تواند اطلاعاتی مشابه با تجربه‌های دنیای واقعی ما از صعودها و ریزش‌های رفت و آمدی عرضه و اطلاعات جمعیت شناختی کاربران را نیز ارائه کند.

در بین وبگاه‌های دارای رتبه‌های بالا می‌توانیم ببینیم به رغم اینکه Google.com همیشه در کل از بالاترین رتبه برخوردار است، اما در تعطیلات آخر هفته فیسبوک از گوگل سبقت می‌گیرد. شبکه های اجتماعی بخش قابل توجهی از زندگی شخصی مردم را به خود اختصاص می‌دهند. در مقایسه، لینکداین به عنوان یک وبگاه شبکه‌های اجتماعی بر مبنای کسب و کار در آخر هفته‌ها رفت و آمد آن افت می‌کند، و رتبه آن مرتباً از ۱۰ به ۱۶ سیر نزولی دارد. اطلاعات جمعیت‌شناختی این دو وبگاه این موضوع را تأیید می‌کند، آنچنان که از پاسخ‌های مردم به پرسشنامه‌ای کوتاه موقع نصب نوار ابزار آلکسا مشخص می‌شود. در کل، مردم بیش تر در محل کار از لینکداین استفاده می‌کنند تا در خانه. هم چنین مشاهده می‌شود که مخاطبان لینکداین مسن‌تر و تحصیل‌کرده‌تر هستند، البته نه به اندازه مسنی، تحصیل‌کردگی یا کم تجربگی مخاطب‌های یک کتابخانه ملی. آلکسا علاوه بر اینکه می‌تواند اطلاعاتی در خصوص سن، تحصیلات، درآمد و محل جستجوی اینترنتی (برای نمونه، خانه یا محل کار)، ارائه کند می‌تواند اطلاعات جنسیت مخاطبان وبگاهی، درآمد، قومیت و فرزند داشتن یا نداشتن و اینکه بازدیدکننده‌ها از کدام کشور وبگاه را بازدید می‌کنند را بدهد. در داده‌های وبگاه‌های کتابخانه‌های ملی، زبان و ارتباط‌های تاریخی منعکس شده است: کتابخانه بریتانیا از ایالات متحده و هند رفت و آمد بیش‌تری دارد؛ کتابخانه ملی فرانسه از الجزیره رفت و آمد بیش‌تر و کتابخانه ملی آلمان از اتریش رفت و آمد بیش‌تری دارد. از آنجایی که ما به سطح دقیق‌تر ارزیابانه توجه داریم

احتمال خطای بیش‌تری وجود دارد چرا که مبنای داده‌ها افراد کم‌تری را تحت پوشش قرار می‌دهد. اگرچه می‌توانیم دلایل منطقی بیاوریم که چرا کتابخانه بریتانیا از ایالات متحده و هند رفت و آمد بیش‌تری دارد، اما اینکه چرا مقدار رفت و آمد تایلند مشابه رفت و آمد هند است، و چرا کانادا نسبت به هلند رفت و آمد کم‌تری دارد، می‌تواند تعجب انگیز باشد.

آلکسا برای شناسایی وبگاه‌ها چند روش عرضه می‌کند. آلکسا علاوه بر تسهیلات جستجو که اجازه ورود کلیدواژه‌ها (یانشانی‌های وبی در صورتی که وبگاه مورد توجهی را بشناسیم) را می‌دهد، امکان مرور ۵۰۰ وبگاه برتر جهانی و مخصوص هر کشور، و نیز مرور وبگاه‌ها بر اساس دسته‌بندی را میسّر می‌کند (برای نمونه، وبگاه‌ها در کتابخانه‌های برتر[1] از آنجایی که راهنما به صورت خودکار ایجاد می‌شود باید برای استفاده از هر دسته‌بندی جانب احتیاط رعایت شود، چرا که بنیان بررسی‌های سنجه وبی است. آلکسا همچنین یک فایل سی.وی.اس[2] از یک میلیون وبگاه برتر دنیا ارائه می‌کند، هم‌چنین برای اشخاصی که بخواهند درباره رتبه‌بندی‌ها دید جامع‌تری داشته باشند روزانه روزآمدسازی می‌شود.

استفاده آلکسا از زیردامنه‌ها و تعداد محدود کاربرانی که نوار ابزار نصب کرده‌اند استفاده آن را می‌تواند به تحلیل وبگاه‌های بزرگ‌تر محدود کند. اما در تعدادی مطالعه و تحقیق، از جمله مقایسه وبگاه‌های استخدامی چین (زان و یان[3]، ۲۰۱۱)، وبگاه‌های شبکه‌های اجتماعی ایالات متحده و چین (لی[4]، ۲۰۱۱)، و دانشگاه‌های مالزی (دیدگاه و عرفان منش[5]، ۲۰۱۰) استفاده شده است. هم‌چنین برای بررسی بیش‌تر از آلکسا به عنوان یک روش نمونه‌برداری در رایج‌ترین وبگاه‌های یک حوزه خاص استفاده کرده‌اند(مثل پرایس و گران[6]، ۲۰۱۲). در ابتکاری بیش‌ترآلکسا را به عنوان اساس تحقیقات ویژگی‌های زبانی نام‌های دامنه (زیانگ[7]، ۲۰۱۲) استفاده کرده‌اند، و نیز تأثیری که پیوندزنی بر رفت و آمد یک وبگاه دارد از طریق مقایسه اطلاعات

1. www.alexa.com/topsites/category/Reference/Libraries
2. comma-separated values
3. Zhan and Yan
4. Li
5. Didegah and Erfanmanesh
6. Price and Grann
7. Xiang

رفت و آمد آلکسا با اطلاعات پیوندی(اینو[1] و دیگران ۲۰۰۵)، موضوعی که در ادامه مطلب تحت عنوان «ردهای کاربر» به آن می‌پردازیم.

یک محدودیت بسیار مهم‌تر آلکسا در ارائه اطلاعات رفت و آمدی وب این است که به شدت با نوار ابزارهای مرورگر پیوند دارد، و چون کاربران بیش از پیش از ابزارهای موبایل برای دسترسی به محتوا استفاده می‌کنند می‌تواند کم‌تر وابسته باشد، بخصوص در مقایسه با ابزاری مانند گوگل که سکوی خنثی[2] هستند.

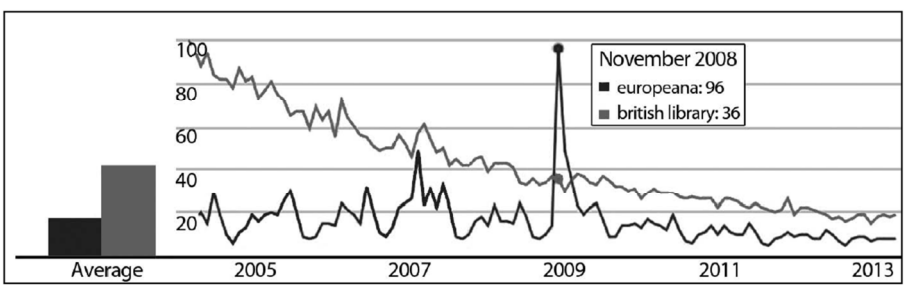
گوگل ترندز

به گفته آلکسا گوگل در مقصد رفت و آمدی دارای رتبه یک است (حداقل در پنج روز هفته)، و میلیون‌ها نفری که جستجوهای خود را در آن انجام می‌دهند منبع بالقوه عظیمی هستند برای کسانی که به رفتار برخط کاربران علاقه‌مند. علاقه بالقوه جستجوهای برخط کاربران به بازارها و پژوهشگران سال‌هاست که شناخته شده است، هرچند امروزه موتورهای جستجو با دقت بیش‌تری اقدامات احتیاطی را برای حفاظت از حریم خصوصی کاربران انجام می‌دهند. در سال ۲۰۰۶ پژوهش AOL داده های لاگ جستجوی نمونه تصادفی ۶۵۸/۰۰۰ کاربری که بجای نام کاربری یا آدرس آی پی[3] (پروتکل اینترنت) از اعداد شناسایی استفاده کرده و «ناشناخته[4]» شده بود را منتشر کرد. متأسفانه جستجوهای بسیاری از افراد حاوی اطلاعات قابل شناسایی است: ما برای خودمان، شهرهای محل سکونت خودمان و جاهایی که در آن کار می‌کنیم را جستجو می‌کنیم. این بدان معنا است که اگر چنانچه شخصی چندین جستجو را با هم جمع کند، داده‌های «ناشناخته» باعث شناسایی افراد می‌شود. از تجربه‌های قبلی درس‌هایی گرفته شده و سرویس‌ها در حال حاضر برای حفاظت از حریم خصوصی کاربران و تجمیع افراد جانب احتیاط بیش‌تری را می‌گیرند.

گوگل ترندز[5] اطلاعات عبارت‌های جستجو شده را در یکی از پنج محصول گوگل ارائه می‌کند: جستجوی وبی، جستجوی تصویری، جستجوی خبری،

1. Ennew
2. platform neutral
3. Interner protocol
4. anonymized
5. Google Trends (www.google.com/trends)

جستجوی محصول و جستجوی یوتیوب. در گوگل ترندز کاربر ترغیب می‌شود برای جستجوی یک عبارت یا تعداد بیش‌تری که به وسیله کاما از هم جدا شده‌اند. سپس قرار است عبارت‌های جستجو در نموداری که جستجوها را از ژانویه ۲۰۰۴ تاکنون نشان می‌دهد بطور جهانی نمایش داده شود(شکل ۴-۴ را ببینید). تعداد کل جستجوها در نمودارها نمایش داده نمی‌شود، بلکه تعداد دفعاتی است که یک جستجو درخواست و بر اساس تعداد کل جستجوها بهنجارسازی می‌شود، و سپس در نموداری از ۰ تا ۱۰۰ ارائه می‌شود. این اطلاعات همچنین در فرمت سی.اس.وی نیز بارگیری می‌شوند که آن را هم می‌توان در یک صفحه گسترده یا ویرایش‌گر متن بارگذاری کرد.



شکل ۴-۴. گوگل ترندز برای جستجوهای عبارت یوروپینا۱ و کتابخانه بریتانیا
[گوگل و آرم گوگل به عنوان نماد تجاری برای شرکت گوگل ثبت شده‌اند، استفاده با اجازه]

داده‌ها را نه تنها می‌توان به مکان‌ها و بازه‌های زمانی مختلف محدود کرد، بلکه زمان و مکان نیز می‌توانند نقطه شروع بررسی داده‌های جستجو باشند. برای مثال، یافتن عبارت‌های جستجوی طراز اول و افزایش در زمانی خاص یا در یک کشوری خاص.

شاید گفته شود که نتایج جستجو جالب هستند چرا که جستجوها بجای کلمات، اعمال را منعکس می‌کنند (دیزیلینسکی[2]، ۲۰۱۲) و این دو همیشه یکی نیستند. شاید کسی برای اینکه پی ببرد جامعه عمومی بریتانیا متوجه ارزشمندی کتابخانه بریتانیا و نیز شایستگی سرمایه گذاری عمومی برای آن است، انتظار داشته باشد که زمان

1. Europeana 2. Dzielinski

زیادی صرف و تحقیقات پرهزینه‌ای کند، اما شکل ۴-۴ مطرح می‌کند که توجه به کتابخانه بریتانیا سیر نزولی دارد. برای کاهش تعداد جستجوهای کتابخانه بریتانیا بین سال های ۲۰۰۴ و ۲۰۱۳ توضیح‌های متفاوتی وجود دارد. از جمله، تعداد روزافزونی از افراد بجای جستجو مستقیماً از خود وبگاه بازدید می‌کنند؛ محتوای کتابخانه بریتانیا در سایر وبگاه‌ها بینهایت منسجم به اشتراک گذاشته شده بطوری که لازم نیست کسی برود و برای آن جستجویی انجام دهد، یا اینکه اعمال محدودیت برای خود ابزارها باشد(برای مثال، هوبارد[1] (۲۰۱۱) درباره «نرخ زوال» برای تعدادی از لغات رایج انگلیسی اطلاعاتی ارائه می‌کند). با اینهمه، افت تعداد بازدید افرادی از وبگاه بدون داشتن یک دلیل شفاف و مشخص چیزی است که مسلماً ارزش بررسی‌های بیش‌تری را دارد.

از آنجایی که داده‌های جستجو در ابتدا به عنوان شاخصی برای سطح سرماخوردگی (ایسنباچ، ۲۰۰۶) و نرخ بیکاری بوده است (اترج، گردس و کاروگا[2]، ۲۰۰۵) اما در مطالعات فراوانی از داده‌های جستجو برای بررسی‌های اقتصادی و سلامت عمومی استفاده می‌کنند و در بسیاری از آن‌ها از داده‌هایی استفاده شده که از طریق گوگل ترندز براحتی سادگی قابل دسترسی بوده‌اند.

در مطالعات اقتصادی اخیر موارد زیر را دریافته‌اند: افزایش مطالعات اقتصادی حاکی از افزایش ناپایداری اقتصادی است (دیزیلینسکی، ۲۰۱۲)؛ کاربران اینترنتی کشورهایی با درآمد ناخالص بالاتر به احتمال زیاد اطلاعات آینده را بیش از گذشته جستجو می‌کنند (پریس و همکارانش، ۲۰۱۲)؛ و اینکه از داده‌های جستجو می‌توان برای عرضه طیف عظیمی از شاخص‌های اقتصادی در خصوص مصرف (وسن و اسشمیدت[3]، ۲۰۱۲) و اعتماد مشتری (چیو و واریان[4]، ۲۰۱۲) استفاده کرد.

درحالی‌که ایسنباچ (۲۰۰۶) توان جستجوها را در بررسی‌های بیماری‌های واگیردار نشان داد، به دنبال انتشار روندهای سرماخوردگی گوگل[5] و انتشار مقاله

1. Hubbard
2. Ettredge, Gerdes and Karuga
3. Vosen and Schmidt
4. Choi and Varian
5. (www.google.org/flutrends)

مشترکی در مجله طبیعت (گینسبرگ[1] و دیگران ۲۰۰۹) توجه‌ها به این سمت سوق داده شد. روندهای سرماخوردگی گوگل عبارت‌های جستجویی که با چرخه‌های آنفولانزا همبستگی داشته‌اند را تجمیع می‌کند تا جایی ادامه می‌دهد که از طریق ذخیره اطلاعاتی متخصصان مراقبت‌های سلامتی بتوانند شاخص‌های قدیمی‌تر واگیرهای سرماخوردگی نشان داده شوند. همانگونه که در مطالعات بعدی نشان داده شد، روندهای سرماخورگی گوگل الزاماً سرماخوردگی را نشان نمی‌دهد، بلکه انتشار علایم سرماخوردگی را نشان می‌دهند (اورتیز[2] و دیگران ۲۰۱۰). بین واگیری‌های سرماخوردگی فصلی اخیر و سطوح نشان داده شده روندهای سرماخوردگی گوگل اختلاف وجود دارد (بوتلر[3]، ۲۰۱۳) بطوری که می‌تواند ناشی از بازخورد در سیستم باشد (وقتی که گزارش سرماخوردگی گوگل جستجوهای زیادی را درباره سرماخوردگی می‌دهد، سازمان‌های خبری آن را گزارش کرده و مردم جستجوی عبارت‌های مربوط به سرماخوردگی را شروع می‌کنند). با این وجود توان عبارت‌های جستجو برای ارائه شاخص‌های بیماری‌های مسری بیش از پیش مورد توجه است. سیفتر[4] و دیگران (۲۰۱۰) مطرح کرده‌اند که از این روش می‌تواند به صورت گسترده‌تری استفاده شود و نشان می‌دهند که چگونه جستجوها با بیماری‌های فصلی از قبیل بیماری لایم[5] همبستگی مناسبی دارند. البته وبگاه روندهای سرماخوردگی گوگل گسترش یافته تا بتواند اطلاعات فعالیت تب دانگ[6] را نشان بدهد، بطوری که با استفاده از داده‌های جستجوی می‌توانند خیلی زودتر از منابع قدیمی شناسایی شود (چان و همکارانش، ۲۰۱۱).

استفاده از گوگل ترندز برای دسترسی به اطلاعاتی در خصوص سلامت عمومی تنها به ارائه اطلاعات بیماری‌های مسری محدود نمی‌شود: والکات[7] و دیگران (۲۰۱۱) نشان داده‌اند که داده های جستجو با شیوع سکته مغزی در ۵۰ استان ایالات متحده همبستگی دارد؛ شوستر، روگر و مک ماهون[8] (۲۰۱۰) دریافتند که جستجو های

1. Ginsberg 2. Ortiz 3. Butler 4. Seifter 5. Lyme (یک بیماری مسری و عفونی)
6. dengue fever 7. Walcott 8. Schuster, Rogers and McMahon

اطلاعات پزشکی موتورهای جستجو با درآمد دارویی و بهره‌گیری از مراقبت سلامتی کلی در یک جامعه همبستگی دارد. داده های جستجو همچنین راهی برای بررسی تأثیر کمپین‌های اطلاعات سلامت، که شاید وبگاه مشخص واحدی نداشته باشند عرضه می‌کنند. برای نمونه، گلین و دیگران (۲۰۱۱) فعالیت جستجوی فزاینده‌ای را در رابطه با سرطان سینه شناسایی کردند که با کمپین آگاهی از سرطان سینه همبستگی دارد.

ظرفیت داده‌های جستجو تنها به شاخص‌های اقتصاد کلان یا اطلاعات سلامت و بهداشت عمومی محدود نمی‌شود، بلکه استفاده از آن در سایر مطالعات ظرفیت داده های جستجو برای پیش‌بینی پذیرش‌های سینما (هند و جادج[1]، ۲۰۱۲) و فرایندهای خرید اتومبیل (دو و کاماکورا[2]، ۲۰۱۲) شناسایی شده است. کتابداران می‌توانند از داده‌های جستجو برای ارائه اطلاعاتی در خصوص حضور برخط سازمان‌های خود استفاده کنند، و هر تغییری را در میزان توجه به موضوعات مختلف، فناوری‌های در حال ظهور و حتی توجه مستمر به کتاب‌ها را در طی زمان ارزیابی کنند (موضوعی که در فصل هفتم مجدداً درباره آن بحث خواهد شد). هنوز روزهای نخستین است، و ظرفیت داده های جستجو نسبتاً تازه شناخته شده است. هنوز کارهای زیادی است که باید انجام شود برای کشف موارد قابل پیش‌بینی و موارد غیر قابل پیش‌بینی، و درک تناسب عبارت‌های خاص. عبارات جستجویی که انتظار می‌رود با فعالیت خاصی همبستگی داشته باشد همیشه این همبستگی را نشان نمی‌دهد. از جمله در حالی‌که انتظار همبستگی بین تعداد جستجوهایی با موضوع خودکشی و تعداد واقعی خودکشی‌ها می‌رود، در مقایسه این آمار با آمار دولتی رسمی ژاپن هیچ همبستگی دیده نشد (سویکی[3]، ۲۰۱۱). قابلیت کنکاش موفقیت‌آمیز داده های جستجو در آینده نیز با ابزارهای موجود وابستگی قوی دارد و در سال ۲۰۱۱ گوگل از ابزارهای جدیدی رونمایی کرد به نام گوگل کورلیت[4] که توانایی شناسایی جستجوهایی را دارد که دارای الگویی مشابه با مجموعه داده‌های خاصی هستند.

1. Hand and Judge 2. Du and Kamakura 3. Sueki
4. Google Correlate(www.google.com/trends/correlate)

ردگیری‌های کاربر

درحالی‌که ابزاری مانند آلکسا و گوگل ترندز از رفتار برخط اجتماعی اطلاعاتی ارائه می‌دهند، شاخص‌های افکار و احساسات عمومی محتوای خود وب منبع بیش‌تری است. امروزه افراد و سازمان‌ها در تمامی اشکال و ابعاد و در هر رشته حضور وبی را ضروری می‌دانند. درحالی‌که زمانی هزینه‌ها و دانش فنی مورد نیاز برای انتشار محتوا می‌توانست عاملی باشد که بخش‌های عظیمی از جامعه را از انتشار محتوای آنها محروم کند، فناوری‌های وب ۲/۰ قابلیت نشر برخط محتوا را برای میلیاردها نفر فراهم کرده است. اکنون بزرگ‌ترین وبگاه‌های شبکه‌های اجتماعی صدها میلیون کاربر دارند که روی وبگاه‌های آنها محتوا منتشر می‌کنند و در مورد بزرگ‌ترین آن‌ها در فصل پنجم بحث می‌شود. مهم است که بدانیم طیف محتوایی که در وب منتشر شده بسیار وسیع‌تر از آن چیزی است که از وبگاه‌های شبکه‌های اجتماعی منتشر شده، اگرچه گرفتن داده‌ها موجب یکسری مشکلات می‌شود.

بیش تر بررسی‌های وب‌سنجی ارزیابانه که از ردگیری‌های برخط کاربران استفاده کرده‌اند در کل می‌توانند به عنوان تحلیل پیوندی دسته‌بندی شوند. از آنجایی که در ابتدا توصیه شد که داده های موتور جستجو می‌تواند در محاسبه ضریب تأثیر وبی در تجمیع صفحه‌های وبی استفاده شود (برای نمونه، یک کشور یا یک وبگاه)، ابرپیوندها پایه بسیاری از بررسی‌های ارزیابانه را شکل داده‌اند. اگرچه، همانگونه که در فصل سوم بررسی شد ماهیت متغیر ابزارهای موجود بدین معنا است که ابرپیوند تنها ردگیری برخطی نیست بلکه به عنوان مقیاس تأثیر استفاده می‌شوند.

ابرپیوندها: آلکسا، بهینه‌سازهای موتور جستجو و وب‌پیماها

هرچند امروزه شناسایی ابرپیوندها دشوارتر از زمانی است که اطلاعات پیوندی از طریق موتورهای جستجو به راحتی قابل دستیابی بود، با این وجود هنوز هم می‌توان آن‌ها را شناسایی کرد. وبگاه alexa.com اطلاعات رفت و آمد وبی و تعداد وبگاه‌های شناسایی شده پیوندی را به یک دامنه خاص می‌دهد. این وبگاه نه تنها تعداد

وبگاه‌های شناسایی شده پیوندی را به یک دامنه خاص اعلام می‌کند، بلکه ۱۰۰ وبگاه برتر متصل به یک دامنه خاص را نیز نشان می‌دهد. در حال حاضر وبگاه‌های پیوندزنی در[1] آلکسا به عنوان منبع داده‌ای جایگزینی در تحلیل ابرپیوند وبی پیشنهاد شده است (وگان، ۲۰۱۲)، و برای مقایسه تحلیل کلیدواژه با تحلیل پیوند استفاده شده است (وگان و رومرو-فاریاس، ۲۰۱۲).

آلکسا به عنوان منبع داده‌ای هنوز در خصوص پیوندهای وبی محدودیت‌هایی دارد. آنچنانکه در داده‌های رفت و آمد وبی، اطلاعات تنها در سطح دامنه وجود دارد —تعداد وبگاه‌هایی را که با یک وبگاه دانشگاه پیوند دارند را می‌توان دید (برای نمونه، www.wlve.ac.uk) اما تعداد وبگاه‌های پیوندی به یک زیردامنه یا راهنمای بخصوصی را نمی‌توان دید (برای نمونه، cybermetrics.wlv.ac.uk). آلکسا تعداد وبگاه‌هایی را که به داخل پیوند می‌شوند را نیز می‌شمارد (بر اساس نشانی وبی یک وب عملیاتی شده)، که بدین معنا است که امکان استفاده از گروه‌های جایگزین از قبیل صفحه‌های وبی یا راهنماهای پیوندی داخلی نیست. این وبگاه همچنین ۱۰۰ وبگاه برتر را فهرست می‌کند و به ازای هر وبگاه تنها یک پیوند نشان داده می‌شود. در نتیجه تنها یک تحلیل محتوایی بینهایت محدود می‌تواند نحوه قرار گیری پیوندها را در وبگاه تعیین کند.

با این وجود، می‌توان دید که چگونه داده های وبگاه‌های پیوندزنی در آلکسا می توانند بنیان وب‌سنجی‌های ارزیابانه و بررسی‌های تحلیل وبی برای کتابداران را تشکیل دهند. جدول ۴-۲ (در صفحه بعدی) تعداد وبگاه‌های پیوندی داخل گروه راسل[2] دانشگاه‌های بریتانیا را نشان می‌دهد. علی‌رغم وجود نسبتاً همگن مجموعه‌ای از دانشگاه‌ها، بنا به نظر روزنامه *گاردین*[3] (۲۰۱۲)، عدم اختلاف واقعی در مؤسسه‌های آموزشی بالاتر بریتانیا، تعداد وبگاه‌های پیوندی به دامنه هر دانشگاه با رتبه دانشگاه همبستگی منفی دارد (از لحاظ آماری معنی‌دار در سطح ۵٪ با استفاده از

1. sites linking in 2. Russell Group(www.russellgroup.ac.uk) 3. Guardian

رتبه اسپیرمن[1]). هر چند با وجود معنی‌داری آماری در سطح ۵٪ نمی‌توان رتبه دانشگاه‌های موجود را نادیده گرفت، با این وجود، منبع اطلاعاتی ارزشمندی را ارائه می‌کند که از آن می‌توان نه فقط به عنوان یک شاخص برای رتبه‌بندی بلکه برای حضور و رؤیت‌پذیری وبی استفاده کرد. حضور و رؤیت‌پذیری از منظر تحلیل‌های وبی مهم هستند، و اطلاعات تأثیر محتوای وبی را می‌دهند، بدون توجه به اینکه آیا چنین تأثیری را می‌توان به عنوان شاخصی برای رتبه‌بندی یک مؤسسه در نظر گرفت.

جدول ۴-۲. «وبگاه‌های پیوندزنی در» آلکسای دانشگاه‌های بریتانیا در مارس ۲۰۱۳

نام دانشگاه	نشانی وبی	رتبه آلکسا	
دانشگاه بیرمنگام	www.birmingham.ac.uk/	۳۰	۲۰۶۲
دانشگاه بریستول	www.bristol.ac.uk/	۱۸	۴۲۹۹
دانشگاه کمبریج	www.cam.ac.uk/	۱	۴۲/۵۷۴
دانشگاه کاردیف	www.cardiff.ac.uk/	۴۰	۹۱۰۲
دانشگاه دورهام	www.dur.ac.uk/	۷	۹۹۰۲
دانشگاه ادینبُرگ	www.ed.ac.uk/home	۱۵	۲۳/۹۳۳
دانشگاه اکسِتر	www.ex.ac.uk/	۱۰	۱۰/۵۹۰
دانشگاه گلاسگو	www.gla.ac.uk/	۱۴	۱۵/۷۰۸
امپریال کالج لندن	www3.imperial.ac.uk/	۱۳	۶۹۹۳
کینگز کالج لندن	www.kcl.ac.uk/	۳۱	۸۸۶۵
دانشگاه لیدز	www.leeds.ac.uk/	۳۷	۱۴/۸۴۳
دانشگاه لیورپول	www.liv.ac.uk/	۴۵	۷۷۸۹
مدرسه علوم اقتصادی و سیاسی لندن	www.lse.ac.uk/	۳	۱۳/۱۹۷
دانشگاه منچستر	www.manchester.ac.uk/	۴۱	۱۸/۰۰۲
دانشگاه نیوکاسل	www.ncl.ac.uk/	۳۳	۱۲/۸۵۳
دانشگاه ناتینگهام	www.nottingham.ac.uk/	۲۶	۱۱/۷۸۷
دانشگاه آکسفورد	www.ox.ac.uk/	۲	۴۳/۷۷۴
دانشگاه مری کوییِن لندن	www.qmul.ac.uk/	۳۶	۷۶۲۹
دانشگاه کوییِنز بلفاست	www.qub.ac.uk/	۵۳	۵۶۴۶
دانشگاه شفیلد	www.sheffield.ac.uk/	۴۲	۲۲۹۲
دانشگاه ساوسمپتون	www.soton.ac.uk/	۲۲	۱۳/۳۴۸
دانشگاه کالج لندن	www.ucl.ac.uk/	۶	۲۳/۴۸۴
دانشگاه وارویک	www2.warwick.ac.uk/	۵	۱۱/۹۲۷
دانشگاه یورک	/www.york.ac.uk	۱۷	۹۶۹۹

1. Spearman's rank

آلکسا تنها منبع اطلاعاتی پیوندهای درونی نیست. وبگاه webometrics.info که حضور و رؤیت‌پذیری رتبه‌بندی ۲۱۲۵۰ مؤسسات آموزشی بالاتر را عرضه می‌کند (مواردی که با ابزارهای قدیمی بی‌نهایت پرهزینه بوده اند) از داده های جمع‌آوری شده از آهرفس[1] و مجستیک اس‌ای‌او[2] استفاده می‌کند. این ابزارها ابتداً برای متخصصان بهینه‌سازی موتور جستجو طراحی شده‌اند، افرادی که تلاش دارند تا رتبه وبگاه های سایر افراد را ارتقاء دهند، هرچند پژوهشگران نیز می‌توانند از آن‌ها استفاده کنند. برای مثال، در تحلیل رابطه بین هزینه کتابخانه‌های دانشگاهی ایالات متحده و حضور وبی آن‌ها، اوردونا–مالیا و رگاززی[3] (۲۰۱۳) از مجستیک اس‌ای‌او و خدمات مشابه مرورگر وبگاه باز اطلاعات پیوند درونی جمع‌آوری کردند (www.opensiteexplorer.org). برخلاف وبگاه Alexa.com این خدمات علاوه بر دامنه‌ها اجازه بررسی پیوندزنی به زیر دامنه‌ها را نیز می‌دهند، که در بررسی اوردونا–مالیا و رگاززی (۲۰۱۳) یک جنبه مهم و اساسی است، چرا که دارای طیف وسیعی از زیر دامنه‌ها و راهنماهای مورد استفاده وبگاه ها و خدمات کتابخانه هستند. داده‌های کامل هر یک از این وبگاه‌ها تنها با عضویت قابل دسترسی هستند.

پیوندهای جایگزین: استنادهای نشانی وبی و اشاره‌های وبی

همانگونه که در فصل سوم ذکر شد، هر چند که هیچ یک از موتورهای جستجوی اصلی قابلیت جستجو پیوندی را به صراحت ارائه نمی‌کنند، و تنها در وبگاه بینگ رابط برنامه کاربردی مناسبی ارائه می‌کند، با این وجود هنوز برای ارائه اطلاعات می‌توان از طریق نماد اشاره‌های وبی نام یک سازمان یا استنادهای آدرس اینترنتی، از موتورهای جستجو استفاده کرد (در جایی که خود نشانی وبی قابل رویت است، بجای اینکه یک متن دیگری به نشانی وبی مشخصی ابرپیوند شده باشد).

اگر تعداد تقریبی استناد نشانی وبی یا اشاره‌های وبی برای مجموعه‌ای از پایگاه‌های وبی تمام چیزهایی باشد که لازم است، باز نمی‌توان جستجوها را دستی

1. ahrefs(http://ahrefs.com) 2. Majestic SEO(www.majesticseo.com) 3. Ordina-Malea and Regazzi

وارد کرد، مگر این که تعداد جستجوها بسیار زیاد باشد. در صورتی که تعداد جستجوها زیاد باشد، یا وقتی که پژوهشگری بخواهد نشانی وبی صفحه‌ای را که به یک وبگاه استناد شده، یا عبارت خاصی اشاره شده را تحلیل کند، خودکار کردن فرایند ترجیح دارد. این کار را هم می‌توان با نوشتن برنامه‌ای با استفاده از رابط برنامه کاربردی بینگ انجام داد. یک گزینه آسان‌تر آن، اینکه از تحلیلگر وب‌سنجی[1] گروه پژوهشی سایبرسنجی آماری[2] استفاده کرد. در فصل‌های پنجم و ششم با جزئیات بیش‌تری به تحلیلگر وب‌سنجی می‌پردازیم، که از آن برای تحلیل اظهار نظرهای یوتیوب و پیوند ارتباطی استفاده می‌شود.

رتبه‌بندی موتورهای جستجو

اگر بحث پیرامون تأثیر وبگاه ها را بدون اشاره به وجود غول درونی آن یعنی رتبه‌بندی موتور جستجو رها کرده و بگذریم، سهل‌انگاری کرده‌ایم: امروزه موتورهای جستجو در کمک به مردم برای یافتن اطلاعات برخط نقش حیاتی ایفا می‌کنند. اگر با چند کلمه کلیدی که به دقت وارد شده‌اند نتوان اطلاعات را پیدا کرد، به این معنا است که برای خیل عظیم کاربران نیز وجود ندارد. اساساً، این که موتوری جستجویی فقط بیاید و یک سند مرتبطی ارا بیابد کافی نیست، بلکه باید چیدمان رتبه‌بندی اسناد به نحوی باشد که آن اسنادی که به احتمال زیاد پاسخگوی نیازهای یک کاربر هستند نه تنها در صفحه نحست رتبه‌بندی شوند بلکه در ابتدای صفحه نخست قرار داده شوند. گوگل خیلی خوب مسیر خود را با استفاده از الگوریتم رتبه صفحه[3] به سمت موتور جستجو غالب آغاز کرد، که آن هم ساختار پیوند صفحه را نه تنها با بررسی تعداد پیوندهای دریافتی یک وبگاه بلکه با وزن‌دهی به هر یک از این پیوندها بر اساس تعداد پیوندی دریافت شده، مدّ نظر قرار داد (برین و پیج[4]، ۱۹۹۸). این فقط یکی از چند علامت متفاوتی است که گوگل از آن برای رتبه‌دهی محتوای خود

1. Webometric Analyst(http://lexiurl.wlv.ac.uk)
2. Statistical Cybermetrics Research Group
3. PageRank
4. Brin and Page

استفاده می‌کند، امروزه صنعت بهینه‌سازی موتور جستجو برای شناسایی و دستکاری این علائم تلاش می‌کند به نحوی که مردم بتوانند صفحه های خود را به بالای رتبه‌بندی موتورهای جستجو برسانند. با چنین بازار دستکاری شده‌ای، آیا رتبه‌بندی‌های موتور جستجو می‌تواند شاخص معنی‌داری از تأثیر باشد؟

دستکاری بالقوه بازار نه تنها بر رتبه‌بندی یک موتور جستجو تأثیر دارد، بلکه بر سایر سنجه‌های وبی مانند تعداد پیوندهای مورد اشاره یک وبگاه نیز تأثیر بگذارد. بین تعداد پیوندهای وبی که می‌توانند دیده شوند و دلایل ذکر شده در بررسی آنها اختلاف وجود دارد. در مقابل، الگوریتم رتبه‌بندی گوگل پنهان مانده است، هرچند مردم برای مهندسی معکوس آن مرتباً گام‌هایی برمی‌دارند. درحالی‌که بخش رتبه صفحه الگوریتم رتبه‌بندی با چندین ابزار و وبگاه شناخته شده و به سادگی قابل دسترسی است (مانند www.prchecker.info)، تنها یک عدد واحد در مقیاس ۰ تا ۱۰ ارائه می‌شود، که فضای اندکی را برای مقایسه بین وبگاه ها مشابه ایجاد می‌کند. به منزله جایگزینی، در ارزیابی تأثیر وبگاه ها کتابخانه دانشگاهی ایالات متحده، اوردنا- مالیا و رگاززی (۲۰۱۳) از ارزیابی تأثیر مشابه استفاده می‌کنند، دامنه موزرنک[1] از مرورگر اپن سایت[2] (www.opensiteexplorer.org) که در مقیاس ۱ تا ۱۰۰ است.

از آنجایی که موتورهای جستجو سعی دارند که در بهینه‌سازهای موتور جستجو و ارتقاء تجربه‌های کاربران جلو بیافتند، یکی از پیامدهایی که ارتقاء خواهد یافت شخصی‌سازی است، با توجه به دریافت رتبه موتور جستجوی مختلف توسط کاربران مختلف. ایده رتبه‌بندی موتور جستجو به عنوان شاخصی از تأثیر بیش از پیش بی‌معنی می‌شود.

سنجه‌های وبی باید بر اساس معیارهای باز باشند، تا جایی که ممکن تأیید شده و یا به چالش کشیده شوند، معیارها نباید برای حفظ منافع تجاری پنهان کرده شوند. یک رتبه بالاتر یا پایین‌تر در گوگل در واقع می‌تواند حاکی از کیفیت یک وبگاه

1. Domain MozRank 2. Open Site Explorer

باشد، اما به سادگی می‌تواند یک تغییر ناگهانی الگوریتم را نیز نشـان دهـد. ایـن نکتـه هم مهم است که بتوان سنجه‌ها را اعتباریابی کرد، بطـوری کـه بتـوان دلایـل اسـتقرار پیوندها و اشاره‌های صورت گرفته مورد کند وکاو قرار گیرند. تحقق ایـن موضـوع بـا تحلیل محتوایی و عاطفی است.

زیر-بخش‌های وب[1]

در جایی که اطلاعات مفصل‌تری برای تأثیر یک وبگاه یـا دامنـه لازم باشـد، وب‌پیمـا می‌توانند مناسب‌تر باشند. وب‌پیمای ساکسایبوت[2] برای پژوهش تحلیل پیوند طراحـی شده، و پایه و اساس بسیاری از بررسـی‌هـا را تشـکیل مـی دهـد. اگرچـه رابـطهـای برنامه‌نویسی موتور جستجو برای مدت کوتاهی آن‌ها را تصاحب کردنـد، امـا همیشـه برای مطالعه‌های وب‌سنجی خاص مناسب‌تر بوده‌انـد. بـرای مثـال، در بررسـی پیونـد درونی تعداد کمی از وبگاه‌ها شناخته شده، و پژوهشگر مـی‌خواسـته از نمایـه‌گـذاری کامل وبگاه‌ها مطمئن باشد. یک وب‌پیمای متن باز به‌نـام هریتـریکس[3] نیـز در آرشـیو اینترنت[4] در دسترس است. ایـن وب‌پیمـا در درجـه نخسـت بـرای اهـداف آرشـیوی طراحی شده است نه برای تحلیل پیوندی، اما از آنجایی که به صورت متن بـاز اسـت افراد دارای مهارت‌های لازمه برنامه‌نویسی می‌توانند آن را با نیازهایشان تطبیـق دهنـد. آرشیو اینترنت دسترسی به یکی از وب‌پیماهای خود را نیز فراهم کـرده اسـت، یعنـی ۸۰ ترابایت اطلاعات از ۲/۷ میلیارد نشانی‌های وبـی بـرای اشخاصـی کـه بـه تحلیـل بخش بزرگی از وب علاقه‌مندند(آرشیو اینترنت، ۲۰۱۲). کامن کـراول[5] نیـز دسترسـی به داده‌های وب‌پیما را امکان‌پذیر می‌کنـد، بطـوری کـه همـه افـراد مـی‌تواننـد بـه آن دسترسی داشته و آن را تحلیل کنند. هرچند، در بسیاری از مـوارد، ارزش کـار اضـافی که برای استفاده از داده‌های وب‌پیما باید انجام شود ممکن اسـت بسـیار کـم باشـد، و با انواع جایگزینی اطلاعات بتوان از زمان کتابداران استفاده بهینه کرد.

1. Sub-sections of the web
2. SocSciBot (http://socscibot.wlv.ac.uk)
3. Heritrix (http://crawler.archive.org)
4. The Internet Archive
5. CommonCrawl (http://commoncrawl.or.org)

اگر چه استهلاک قابلیت پیوند موتورهای جستجو اصلی، و محدودیت‌های رابط‌های برنامه‌نویسی آن‌ها، باعث شده که گردآوری اطلاعات پیوندزنی به یک وبگاه از سراسر وب به طرز روزافزونی با مشکل مواجه شود، با این همه می‌توان از تأثیر یک وبگاه در زیر بخش‌های وب مطلع شد. اندازه بزرگ وبگاه‌های شبکه های اجتماعی به این معنا است که آن‌ها می‌توانند منبع اطلاعاتی مهمی باشند. برای مثال، چند بار در توییتر یا گوگل پلاس به یک وبگاه اشاره شده است، یا چند بار وبگاه روی دلیشس[1] یا داگ[2] روی دیگ[3] علامت‌گذاری شده است. برخی از این نوع اطلاعات در بعضی خدمات پیوند بهینه‌سازی موتورهای جستجو (مانند http://ahrefs.com)، و بعضی دیگر در ابزارهای تخصصی‌تر (مانند http://topsy.com) وارد شده‌اند، و در بعضی موارد خود وبگاه‌ها نیز ممکن است جمع‌آوری کرده باشند (مانند http://delicious.com). در فصل بعد درباره تحلیل داده‌ها از وبگاه‌های شبکه‌های اجتماعی بطور مفصل‌تری بحث خواهیم کرد.

یک رویکرد نظام‌مند برای تحلیل محتوا

همانگونه که قبلاً هم اشاره شده است «داده‌های خام تنها نیمی از جریان را فاش می‌کنند» (مارک، ۲۰۱۱، ۶) و اگر قرار است از بینش‌ها حمایت شود باید گام‌هایی برای تعیین دلایل استقرار پیوندها برداشته شود. همانگونه که در فصل دوم ذکر شد، در مواردی که نیاز به تحلیل حجم عظیمی از داده‌های ساختار یافته باشد تحلیل عاطفی می‌تواند مناسب باشد (و به این موضوع در فصل بعد بیش‌تر می‌پردازیم). اما در مورد طیف گسترده محتوایی که می توانند با پیوندزنی یا اشاره به یک وبگاه پیدا شده باشند، احتمالاً یک تحلیل محتوایی مناسب‌تر است.

تحلیل محتوای یک سنجه وبی رویکرد نظام‌مندی است برای تحلیل دلایل ایجاد محتوا از طریق بازرسی انسانی از محتوا، و در صورت لزوم پیرامون محتوا. از تحلیل محتوا به منظور تعیین نوع نویسنده محتوا (مانند خصوصی یا عمومی، گروه

1. Delicious 2. Dug 3. Digg

پژوهشی، بخش دانشگاهی یا خدمات پشتیبانی)، ماهیت محتوای خاصی (مانند یک تصویر حاوی چه چیزی است)، دلیل نحوه ساخت یک پیوند (مانند برای بیان رابطه بین وبگاه پیوند شده یا فقط یک منبع ارزشمند)، یا دلیل ذکر عبارت خاصی (مانند ابراز نارضایتی از یک شرکت یا نمایش نقاط قوت) استفاده شود. تمرکز تحلیل وبی بر طیف عظیمی از محتوا است: پیوندهای درونی به وبگاه‌ها تجاری (وُگا، گوآ و کیپ[1]، ۲۰۰۶)؛ پیوندهای بیرونی در وبگاه ها دانشگاهی (استوارت، تلوال و هریس[2]، ۲۰۰۷)، بررسی رستوران‌ها (پانتلیدیس[3]، ۲۰۱۰)، اظهارات سیاست‌مدارها در تصاویر (اوزیل و پارک[4]، ۲۰۱۲) و اطلاعات واکسن HPV (مادن و همکارانش[5]، ۲۰۱۲). در تمامی این موارد یک رویکر اصولی برای بررسی ماهیت محتوا انجام شده است.

روش تحلیل محتوا عمدتاً شامل پنج مرحله است:

- شناسایی پرسش (یا پرسش‌هایی) که باید پاسخ داده شوند؛
- انتخاب نمونه‌ای از محتوا؛
- توسعه یک طرح مناسب برای دسته‌بندی؛
- دسته‌بندی محتوا؛
- تحلیل یافته‌ها.

شناسایی پرسش (یا پرسش‌هایی) که باید پاسخ داده شوند

همانگونه که در مثال‌های تحلیل محتوایی مورد بحث بالا می‌توان دید، طیف پرسش‌هایی که می‌توانند پایه و اساس تحلیل محتوا را شکل دهند فقط به تخیل پژوهشگر محدود می‌شود. یک بررسی تحلیل وبی ممکن است شامل بررسی دلایل استقرار پیوندها در وبگاه خود یا رقبا باشد، در حالی‌که یک مطالعه علم‌سنجی می تواند بررسی حوزه‌های پژوهشی مورد توجه یک رشته خاص باشد. بدون داشتن یک پرسش شفاف نمی‌توان دست به نمونه داده‌ای مناسب یا توسعه یک طرح دسته‌بندی مناسب زد.

1. Vaughan, Gao and Kipp
2. Stuart, Thelwall and Harries
3. Pantelidis
4. Ozel ad Park
5. Madden

انتخاب نمونه‌ای از محتوا

نحوه انتخاب نمونه‌ای از وبگاه‌ها به شدت به ابزاری وابسته است که برای جمع‌آوری داده‌ها استفاده می‌شود. اگر برای جمع‌آوری داده‌ها از یک وب‌پیما استفاده شده باشد، برای دسته‌بندی هزاران نتیجه بالقوه می‌تواند وجود داشته باشد. اگر برای جمع‌آوری داده‌ها از یک موتور جستجو استفاده شده باشد تنها تا سقف ۱۰۰۰ نتیجه وجود خواهد داشت. اگر تحلیل محتوا برای اظهارنظرهای ارائه شده درباره یک سازمان باشد، نتایج تنها می‌تواند انگشت‌شمار باشند. از آنجایی که تعداد زیادی از پدیده‌های وبی برای ظهور به یک قدرت قانونی نیازمندند (مانند تعداد فراوانی از پیوندها در تعداد انگشت‌شماری از وبگاه‌ها) باید برای لحاظ کردن این موضوع گام‌هایی برداشته شود.

برای نمونه، فرض کنید که یک پژوهشگر وب‌سنجی علاقه‌مند به بررسی دلایلی باشد که دانشگاه‌ها به اقلام موجود در انبار چاپ الکترونیکی arXiv.org پیوندزنی می‌کنند، و یا آیا پیوندها مترادف استنادهای قدیمی هستند، و یا به صفحه‌های خانگی که پژوهشگران مقاله‌های خود را در آن نشان می‌دهند. شاید وب پیمایی در فضای وبی دانشگاهی بریتانیا ۸۳۰۰۰ صفحه با پیوند به آرشیو شناسایی کند که برای پژوهشگر بسیار بیش از آن چیزی است که بتواند دستی تمام آن‌ها را دسته‌بندی کند. در نتیجه تصمیم می‌گیرد که یک نمونه ۵۰۰ تایی بردارد. هرچند، اینکه بخواهیم فقط ۵۰۰ نمونه تصادفی نشانی وبی از ۸۳۰۰۰ صفحه برداریم ناخواسته نتیجه را به نفع مؤسسه‌هایی برمی‌گردد که تعداد فراوانی پیوند به صورت غیر عادی دارند. برای مثال، دانشگاه ساوسمپتون[1] میزبان انعکاسی از arXiv.org در xxx.soton.ac.uk است. شمارش پیوندها به همان صورتی که در سطح دامنه مناسب است، شاید بهتر باشد که ابتدا از هر یک از ۱۰۰ عدد یا دامنه های دانشگاهی پیوندی به arXiv.org، ۵۰ نمونه تصادفی نشانی وبی نمونه‌گیری شده و سپس نمونه تصادفی از این مجموعه کوچک‌تر انجام شود.

1. Southampton University

اگرچه در صورتی که یک وب‌پیما داده‌ها را جمع‌آوری کرده باشد شاید مجموعه کاملی از نتایج در دسترس باشد، اما موتورهای جستجو اغلب برای کاربران به ۱۰۰۰ نتیجه اول محدود می‌شوند. تحلیل‌گر وب‌سنجی به آدرس http://lexiurl.wlv.ac.uk امکان بارگیری نتایج بینگ را می‌دهد، ولی اگر پژوهشگری بخواهد نتایج را از گوگل بگیرد باید آن‌ها را در صفحه‌های اچ.تی.ام.ال اسکرپ کند[1]. اگرچه نتایج می‌توانند به صورت یک فایل CSV با نوار ابزارهای وصل به[2] SEOQuake بارگیری شوند، اما برای فرستادن یک‌باره جستجوهای زیاد به گوگل باید جانب احتیاط رعایت شود، چرا که ممکن است گوگل آن آدرس آی.پی را برای چند ساعت مسدود کند. برای بدست آوردن گویاترین نمونه از نتایج، لازم است که از نتایج شخصی صرف نظر کرد، و اگر وبگاه Google.com هم‌چنان جستجوگر را به محلی متفاوتی تغییر جهت داد (برای نمونه، Google.co.uk) در اینصورت www.google.com/ncr پیوندی است که گوگل از آن تغییر مسیر نمی‌دهد.

گسترش یک طرح دسته‌بندی مناسب

لازم است که طرح (یا طرح های) دسته‌بندی نشانگر پرسش (یا پرسش هایی) باشند که تحلیل محتوا در مورد آنها است، و دسته‌ها تا حد ممکن از هم مجزا و مشخص باشند. برای مثال، به مثال بالا برکردیم که در آن پژوهشگری وب‌سنجی که به بررسی دلایل پیوندزنی دانشگاه‌ها به موضوعات موجود در انبار چاپ الکترونیکی arXiv.org علاقه‌مند است، پژوهشگر ممکن است تصمیم بگیرد که پرسش را به دو بخش تقسیم کند: از دانشگاه‌ها چه کسی به arXiv.org پیوندزنی می‌کند، و دلایل جایگذاری پیوند، و اینکه باید طرح دسته‌بندی مجزایی برای هریک از آنها ایجاد کند.

دو رویکرد اصلی وجود دارد که می‌توان از آن‌ها برای ایجاد طرح استفاده کرد: یک رویکرد تکراری یا خوشه‌بندی پس-همارا[3]. یک طرح تکراری در طی فرایند تحلیل محتوا، با ایجاد دسته‌های جدید مورد نیاز توسعه داده می‌شود. خوشه‌بندی

1. scraped 2. plug-in(www.seoquake.com) 3. post-coordinate

پس–همارا در جایی که از بیش از یک دسته‌بندی مورد استفاده است مناسب‌تر است و لازمه آن بازدید از گزینش تصادفی محتوا و سپس ابداع دسته‌های مناسبی است که از نظر محتوایی مشابه هستند. در خوشه‌بندی پس–همارا امکان ایجاد یک تفاهم‌نامه دسته‌بندی هست، بیانیه صریحی است برای دسته‌ها و محتواهای درونی آن، به منظور کمک به استحکام و توافق دسته‌بندی‌کننده داخلی.

دسته‌بندی محتوا

دسته‌بندی بندرت فرایندی سیاه و سفید است. حتی در بهترین طرح‌های دسته‌بندی طراحی شده ممکن است به نظر بیاید که برخی موضوعات را می‌توان در چندین دسته یا در جایی که محتوای آنها مبهم باشد (از جمله اینکه آیا یک اظهار نظر طعنه‌آمیز است یا جدی) قرار داد. در چنین شرایطی یک تفاهم‌نامه برای توضیح نحوه دسته‌بندی به بخش‌های علاقه‌مند کمک می‌کند.

بر اساس هدف تحلیل، چندین دسته‌بندی‌کننده می‌توانند مورد نیاز باشد. یک مطالعه مقدماتی برای اهداف داخلی ممکن است تنها یک دسته‌بندی‌کننده لازم داشته باشد. در مطالعات گسترده‌تری که ممکن است بر تصمیم‌های سیاسی تأثیر بگذارد شاید بیش از یک دسته‌بندی‌کننده مورد نیاز باشد. در جایی که از بیش از یک دسته‌بندی‌کننده استفاده شود، دسته‌بندی‌کننده دوم ممکن است بخشی از پیوندهایی را که دسته‌بندی‌کننده اول قبلاً دسته‌بندی کرده است را دسته‌بندی کند (برای نمونه، ۱۰٪)، و سپس با استفاده از یک آزمون آماری مانند آلفا کریپندورف[1] می‌توان ثبات دسته‌بندی‌کننده داخلی را مورد آزمون قرار داد (کریپندورف، ۲۰۱۲). اگر اختلاف‌های اساسی و بزرگی وجود داشته باشد شاید بازگشت به پله سوم، و تکرار فرایند با یک تفاهم نامه دسته‌بندی واضح‌تر، یا دسته‌بندی تفکیک شده‌تری لازم باشد.

یافته‌های تحلیل

گام آخر استنتاج نتایج یافته‌ها است. برای تحلیل مقیاس کوچک از اظهارنظرهای یک

1. Krippendorff's alpha

کتابخانه ممکن است کافی باشد که بگوییم «۶ اظهار نظر مثبت و ۳ اظهار نظر منفی بوده‌اند» اما اگر آزمایش وسیع‌تری انجام شده باشد ممکن است اعمال بررسی‌های آماری بیش تری نیز ضروری باشد. برای نمونه، پژوهشگران وب‌سنجی که علاقمند به دلایل پیوندزنی دانشگاه‌ها به موضوعات موجود در انبار چاپ الکترونیکی arXiv.org هستند ممکن است بخواهند مشخص کنند که آیا تفاوت مشاهده شده دلایل پیوندزنی در انواع مختلف صفحه‌ها از لحاظ آماری معنی‌دار است یا نه که در این مورد آزمون خی‌دو[1] مناسب خواهد بود.

نتیجه‌گیری

این فصل به طیف وسیعی از ابزارها و رویکردهای درونی و بیرونی که برای ارزیابی تأثیر محتوای وب مورد استفاده هستند پرداختیم. حتی برای چیزی به سادگی ارزیابی تأثیر یک وبگاه، مجموعه‌ای از سنجه‌ها از طیف منابع متنوعی به راحتی می‌تواند عرضه شوند. برخی از آنچه در بالا به آن اشاره شد شامل موارد زیر است:

- بازدیدهای صفحه (گوگل آنالیتیکز)
- نرخ پرش (گوگل آنالیتیکز)
- رتبه رفت و آمد (آلکسا)
- پیوندزنی به وبگاه‌ها (آلکسا)
- پیوندهای درونی (www.majesticseo.com)
- استنادهای نشانی وبی (بینگ)
- پیج رَنک (گوگل)
- دامنه موزرنک (www.opensiteexplorer.org)
- تعداد مکان‌یاب‌ها (http://delicious.com)
- تعداد پیوندهای توییت شده (http://topsy.com)

1. chi-square

یک فهرست جامع ظاهراً انتها ندارد، زیرا ممکن است داده‌ها به صورت نامتناهی برای ارائه دانه‌های بهتر ارزیابانه بهم متصل شده باشد (مانند بازدیدهای صفحه در افغانستان یا وبگاه‌های پیوند داخلی با اظهارنظرهای مثبت). مناسب‌ترین مجموعه سنجه‌ها به هدف بررسی بستگی دارد و به ناچار باید از عنصر آزمون و خطا استفاده کرد.

اگرچه توان سنجه‌های وبی برای اهداف تحلیل وبی آشکار است، باید مشخص باشد که وب می‌تواند در مورد طیف بسیار وسیع‌تر رفتار دنیای واقعی نیز اطلاعاتی ارائه کند. درحالی که وب بیش از پیش برای عرضه بینش‌های اقتصادی و بهداشتی مورد بررسی قرار می‌گیرد، ممکن است به نظر بیاید که ماهیت چنین بررسی‌هایی منحصراً به تخیل پژوهشگر و ابزارهای موجود محدود است.

۵

ارزیابی تأثیر رسانه‌های اجتماعی

مقدمه

بیش‌تر اطلاعاتی که به صورت برخط در دسترس هستند روی سرور شخصی یک مؤسسه یا یک شخص نیستند، بلکه از وبگاه‌ها و سرورهای بیرونی استفاده می‌کنند: تصاویر در فلیکر؛ ویدئوها در یوتیوب؛ نمایش‌ها در اسلایدشیر[1]؛ و اظهارنظرها در توییتر بارگذاری می‌شوند. این خدمات و وبگاه‌ها نه تنها فرایند انتشار را برای افراد آسان می‌کنند، بلکه می‌توانند محتوایی را که احتمال ویروسی شدن دارد را نیر تحویل دهند. سنجه های وبی دارای مزایا و معایبی هستند. از یک سو وبگاه‌ها و خدمات بیرونی می‌توانند سنجه‌های بیش‌تر و پیچیده‌تری فراهم کنند: آن‌ها حجم زیادی داده ساختاریافته را به روشی مشابه عرضه می‌کنند، و از لحاظ نظری تمام داده ها را با یک معیار خاص می‌توان بازیابی کرد. به عبارت دیگر، سنجه ها محدودیت شدیدی را به عملکرد خدمات مجاز اعمال می‌کنند، هرچند که ممکن است داده‌ها ساختار یافته باشد ولی وبگاه می‌تواند دسترسی به آن داده‌ها را فراهم نکند.

با چنین طیف گسترده ای از خدمات موجود، از نقطه نظر تجزیه و تحلیل وب ضروری است که درباره ارزشمندی هریک از خدمات سؤال کنند (ووکوویچ و همکارانش[2]، ۲۰۱۳)، و اینکه در چه نقطه‌ای متوقف شده و لذا برای یک خدمات

1. SlideShare 2. Vucovich

خاص تلاش نکنند. با در نظر داشتن ویژگی های سنجه های وبی مورد بحث در فصل پیش، اطلاعاتی که سنجه های رسانه‌های اجتماعی می‌توانند عرضه کنند در مقایسه بسیار بیش تر از اطلاعات حضور در وب خود کتابخانه است. تعداد انگشت شماری از خدمات و وبگاه های شخصی با صدها میلیون کاربر نیز می‌توانند اطلاعاتی از رفتار دنیای واقعی کاربر، که زمینه‌ساز طیف بسیاری از بررسی‌های وب‌سنجی باشند، را فراهم کنند.

این فصل با عنایت به بررسی انواع وبگاه‌های شبکه‌های اجتماعی موجود و انواع محتوای وبگاه های شبکه های اجتماعی که شاید کتابداران به ارزیابی آن‌ها علاقه‌مند باشند آغاز می‌شود. این بحث با نگاه اجمالی به برخی وبگاه‌های شبکه‌های اجتماعی عمومی که می‌تواند اساس یک بررسی سنجه وبی باشد، بررسی‌های انجام شده قبلی، و برخی از ابزارهای موجود دنبال می‌شود.

ابعاد وبگاه‌های شبکه‌های اجتماعی

وب تحت سلطه تعداد اندکی وبگاه است، که بسیاری از آن ها دارای برخی از کاربردهای شبکه های اجتماعی هستند. تعریف بوید و الیسون (۲۰۰۷) از وبگاه شبکه های اجتماعی شامل سه بخش است: اجازه به کاربران برای ایجاد یک پروفایل عمومی در یک سیستم محدود، برقراری ارتباط با سایر کاربران، هدایت اتصال هایی که آنها و دیگران برقرار می‌کنند. این تعریف شامل وبگاه هایی مانند فیسبوک و توییتر است و نیز وبگاه هایی که در درجه نخست ممکن است به عنوان وبگاه های شبکه‌های اجتماعی نباشند از قبیل ای بی[1] و ویکی پدیا.

در هنگام ثبت فهرست وبگاه های شبکه های اجتماعی در ویکی پدیا (a۲۰۱۳) ۲۰۰ وبگاه شبکه های اجتماعی عنوان فعال اصلی[2] را در بر می‌گرفت و هر اثری که قصد داشت سنجه‌ها را مطرح کند به سرعت تکراری و از رده خارج می‌شد چراکه وبگاه‌های شبکه‌های اجتماعی مختلفی ظهور کرده و از بین می‌روند و کاربردها تغییر

1. eBay 2. major active

می‌کند. برخی از این وبگاه‌ها و خدمات رابط‌های برنامه‌نویسی گسترده‌ای که قادر به محاسبه طیف گسترده‌ای از سنجه‌ها هستند را ارائه می‌کنند، بقیه آنها محتوای بسیار محدودی را نشان می‌دهند. این فصل با داده‌ای آغاز می‌شود که یک پژوهشگر سنجه‌های وبی، در صورت دسترسی به داده‌ها، تمایل به بررسی آن دارد.

در تعریف بوید و الیسون (۲۰۰۷) یک وبگاه شبکه‌های اجتماعی دارای سه بخش است: پروفایل عمومی، اتصال به کاربران دیگر، و قابلیت هدایت این اتصال‌ها و هر یک از این ها می‌توانند بنیان بررسی‌های سنجه وبی باشند.

پروفایل‌ها

هنگام بحث درباره پروفایل‌های روی وبگاه‌های شبکه‌های اجتماعی ایجاد تمایز بین بخش‌های اصلی پروفایلی که تا حدی ثبات دارد و پروفایل موقتی که غالباً امکان تغییر دارد ارزشمند است. برای مثال، پروفایل اصلی کاربر توییتری حاوی عکس‌ها، نام کاربری آنها، نام، مکان، وب وبگاه و شرح حال بیش از ۱۶۰ شخصیت است. گرچه احتمال دارد که هریک از این‌ها را کاربر تغییر دهد، اما معمولاً روزانه ثابت باقی می‌مانند. پروفایل موقتی کاربر توییتر در وهله نخست حاوی روزآمدسازی‌هایی است که کاربر می‌فرستد، و روزآمدسازی‌های سایر اشخاصی که در دسته دوست داشتنی‌های[1] کاربر قرار داده شده‌اند. ممکن است این محتوا دائما در دسترس باشد، اما از آنجایی که محتوای جدید ایجاد می‌شود محتوای قبلی به عقب رانده می‌شود. پروفایلی که کاربر می‌تواند آن را در یک وبگاه شبکه های اجتماعی مطابق وبگاه خاصی ایجاد کند، و نیز دوره حیات یک خدمت بسیار تغییر می‌کند چرا که عملکرد آن اضافه شده یا متوقف می‌شود. ماهیت محتوای پروفایل موقتی بین وبگاه‌های شبکه‌های اجتماعی اغلب بسیار متفاوت است: توییتر در وهله اول بر روزآمدسازی‌های کوتاه مدت مبتنی بر متن تمرکز دارد (اگرچه امکان انواع محتوای بیش‌تر را نیز می‌دهد)؛ اسلایدشیر برای ارائه ها است؛ و یوتیوب برای ویدئوها است.

1. favourites

بسیاری از وبگاه‌ها نیز این فرصت را فراهم می‌کنند تا بتوان روی فضای پروفایل موقتی سایر افراد به ایجاد یک عنصر محتوایی پرداخت. برای مثال، گذاشتن پیام بر دیوار فیسبوک یک کاربر دیگر، یا اظهار نظر روی ویدئو یوتیوب یک کاربر دیگر. پروفایل های اصلی نیز بسیار متفاوت هستند. برای نمونه، درحالی‌که زندگی‌نامه در توییتر متن ساختار نیافته ۱۶۰ نویسنده است، لینکداین می‌تواند زندگی‌نامه ساختاریافته‌تری با توجه به تجربه، آموزش و مهارت‌های آنها ارائه کند، که این کار لینکداین بیش‌تر در راستای نقش آن به عنوان یک وبگاه شبکه‌های اجتماعی برای حرفه‌مندان تجارت است.

بسته به در دسترس بودن اطلاعات، هر یک از ابعاد پروفایل می تواند پایه گذار یک بررسی سنجه وبی باشند. در موارد بسیاری یک پژوهشگر می‌تواند به ویژگی های جمعیت‌شناختی کاربران یک وبگاه یا زیر مجموعه خاصی از کاربران در پروفایل اصلی علاقه‌مند باشد: کتابداران قصد استفاده از یک وبگاه شبکه‌های اجتماعی داشته باشند تا تعیین کنند که آیا اعضای فعلی همان کسانی هستند که کتابداران می‌خواهند با آنها تعامل داشته باشند یا خیر؛ کتابداران کتابخانه‌های عمومی ممکن است داده های کاربران یک منطقه محلی را جمع آوری کنند (یا اینکه کسی که ارتباطی را با یک کتابخانه داشته است) چرا که می‌توانند مطمئن شوند که کتابخانه‌ها در موضوع‌های مختلف با علایق جامعه هم راستای هستند؛ مدیران سازمان‌های تجاری ممکن است بخواهند از ویژگی‌های جمعیت‌شناختی اشخاصی که به محصولات آن‌ها ابراز علاقه کرده‌اند اطلاعاتی داشته باشند؛ مدیران شرکت‌هایی که می‌خواهند رقابت مؤثرتری با رقبای موفق‌تر خود داشته باشند شاید برای شناسایی شکاف‌های حوزه تخصصی خود بخواهند اطلاعات مهارت‌های کارکنان رقبای خود را بدست آورند.

تمرکز بیش‌تر پژوهش سنجه وبی روی پروفایل موقتی است. این موضوع فرصتی را فراهم می‌کند تا بجای اینکه زمانی را صرف ایجاد پروفایل‌های خود بکنند

اطلاعاتی را از ادراک اشخاص از دنیای امروز بدست آورند: تعداد اظهارنظرهای مثبتی که یک محصول در ماه گذشته دریافت کرده می‌تواند اطلاعاتی بیش‌تری در خصوص نحوه دریافت آن بدهد تا لایک‌ها و تنفرهای[1] پروفایلی که شاید سال‌ها پیش ایجاد شده است. همانند رفتار جستجوی کاربران، آنطوری که از طریق گوگل ترندز به نمایش گذاشته می‌شود (فصل چهارم را ملاحظه فرمائید)، تجزیه و تحلیل حجم عظیمی از اطلاعات پروفایل موقتی فرصتی را برای عرضه شاخص های حوزه هایی چون بهداشت و تندرستی و اقتصاد ایجاد می‌کند. جای تعجب نیست که کانون توجه بیش‌تر پژوهش‌ها بجای ویدئوها و عکس‌ها بر محتوای مبتنی بر متن است، چرا که قابلیت تجزیه و تحلیل خودکار، یا حداقل نیمه خودکار را دارند. با این وجود، این امکان وجود دارد که انواع محتوایی که کانون تجزیه و تحلیل‌ها هستند غنی‌تر شوند، مانند موضوع ویدئوها و عکس‌هایی که برچسب زنی می‌شوند یا به یک شرکت یا محصول خاصی ملحق می‌شوند.

اتصال‌ها

غالباً اتصال بین پروفایل‌ها کانون توجه سنجه‌های وبی در اهداف تجزیه و تحلیل وبی هستند. اطلاعات اغلب بلافاصله در دسترس بوده و به نحو چشمگیری نمایش داده می‌شوند، خواه تعداد دنبال‌کننده‌ها در توییتر باشد، خواه تعداد دوستان در فیسبوک یا مشترکین یوتیوب. تشخیص این مسئله مهم است، هر چند که، ماهیت ارتباط می‌تواند بسیار متفاوتر از آن چیری باشد که بطور تلویحی از طریق واژه تخصیصی وبگاه بیان می‌شود: دوستان فیسبوکی با دوستان دنیای واقعی بسیار متفاوت هستند؛ دنبال کردن در توییتر به معنای این نیست که دنبال‌کننده‌ها با هم اتفاق نظر دارند؛ و اشتراک برای یک مشترک یوتیوب الزاماً آنقدر ارزش ندارد که برای آن حاضر به پرداخت پولی باشد.

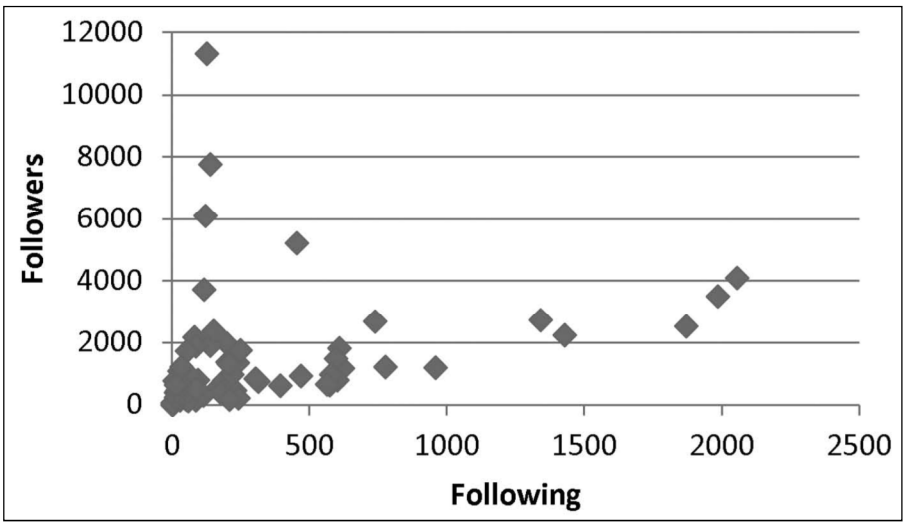
1. dislikes

اگرچه می‌تواند به نفع وبگاه‌های شبکه‌های اجتماعی باشد که کاربران برای ساخت تعداد اتصال های خود وقت صرف کنند، و می‌توان بحث کرد که اغلب به تعداد ساخت‌ها در یک رقابت امتیاز داده می‌شود، این مسئله مهم است که ارزش اتصال‌ها در چشم‌انداز است، بویژه آنکه تعداد اتصال‌های دریافت شده علاوه بر انعکاس مجموعه متنوعی از رفتارها می‌تواند کیفیت درک شده حساب کاربری یک کاربر را نشان بدهد.

اهمیت پرسشگری درباره سنجه‌ای خاص را با یک مثال به بهترین شکل می‌توان نشان داد: تعداد دنبال‌کننده‌های یک حساب توییتر نمادی از موفقیت حساب توییتر است. ساده‌ترین ارزیابی تأثیر حساب کاربری توییتر کتابخانه تعداد دنبال‌کننده‌های آن است. با مقایسه تعداد دنبال‌کننده‌های حساب کاربری توییتر دو کتابخانه به صورت نظری متوجه می‌شوید که کدام کتابخانه در توییتر موفق‌تر بوده است. برای مثال، حساب کاربری توییتر کتابخانه دانشگاه ساوسمپتون[1] ۴۸۴ دنبال‌کننده دارد درحالی‌که کتابخانه دانشگاه آنجلیا شرقی[2] ۸۹۷ دنبال‌کننده دارد، این امر حاکی از موفقیت بیش‌تر حساب کاربری دانشگاه آنجلیا شرقی است. متأسفانه، ممکن است تعداد دنبال‌کننده‌های یک حساب کاربری تحت تأثیر عوامل برخط و چه غیر برخط بسیاری باشند.

این احتمال وجود دارد که تعداد عواملی که بر تعداد دنبال‌کننده هایی یک حساب کاربری تأثیر می‌گذارند شامل تعداد حساب‌هایی که آن حساب کاربری آن‌ها را دنبال می‌کنند (شکل ۵-۱)، و تعداد روزآمدسازی‌های ارسال شده باشد (شکل ۵-۲). هر دو شکل ۵-۱ و ۵-۲ علیرغم بخش‌های پَرت[3] همبستگی‌های آماری معنی‌دار مثبتی را نشان می‌دهند (معنی‌دار در سطح ۵٪ با استفاده از آزمون همبستگی پیرسون).

1. University of Southhampton (@UniSotonLibrary)
2. University of East Angelia (@UEALibrary)
3. outliers



شکل ۵-۱. تعداد دنبال‌کننده‌ها و تعداد دنبال‌های کتابخانه‌های دانشگاهی بریتانیا (آوریل ۲۰۱۳)

برای نمونه، این واقعیتی که کتابخانه بودلیان[1]، با وجود دنبال کردن تنها ۱۲۶ جریان توییتر و روزآمدسازی حساب کاربری با میانگین ۱۲۹۷ بار، بیش از ۱۱۰۰۰ دنبال‌کننده دارد را بتوان به جای اینکه به هرچیز دیگری در حساب کاربری توییتر آن نسبت داد در این ارتباط دانست که این کتابخانه در جهان شناخته شده است.

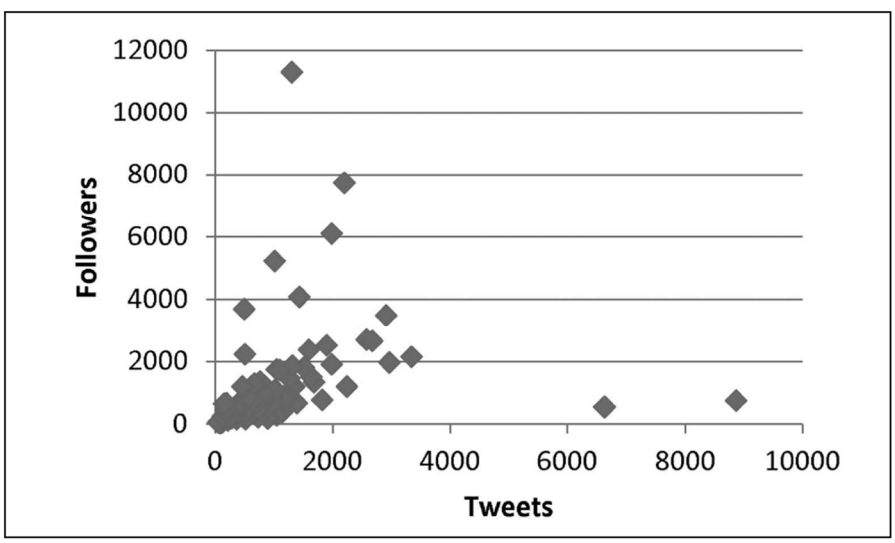
یک تجزیه و تحلیل رگرسیون[2] چندگانه بر اساس داده‌های این حساب‌های کاربری (به استثنای کتابخانه بودلیان) معادله زیر را به وجود می‌آورد:

دنبال‌کننده‌های مورد انتظار = ۱/۰۸ × دنبال‌کننده‌ها + ۱/۸۵ × روزآمدسازی‌های وضعیت

با چنین معادله ای می‌توان حساب های کاربری که در طول زمان‌های مختلف ارسال شده‌اند و تعداد افراد متفاوتی که دنبال می‌کنند را مقایسه کرد. در واقع، با استفاده از این فرمول پی می‌بریم که اگرچه تعداد دنبال‌کننده‌های کتابخانه آنجلیا شرقی بیش‌تر است، ممکن است این تصور پیش بیاید که کتابخانه دانشگاه ساوسمپتون حضور برخط موفق تری را داشته دارد. درحالی‌که کتابخانه آنجلیا شرقی تنها ۸۰ درصد تعداد دنبال‌کننده مورد انتظار را دارد، کتابخانه دانشگاه ساوسمپتون ۳۸ درصد دنبال‌کننده

1. Bodleian Library 2. regression

بیش از حد انتظار دارد! اما لازم نیست عواملی کـه بـرای چنـین تجزیـه و تحلیلـی در نظر گرفته می‌شوند فقط به اطلاعات بـرخط محـدود باشـند، بلکـه مـی‌تواننـد شـامل اطلاعاتی باشند که در دسترس عموم اسـت از قبیـل تعـداد دانشـجوهای دانشـگاه یـا بودجه کتابخانه. مسئله این نیست که کتابخانه نباید از تعداد دنبال‌کننده‌های تـوییتر بـه عنوان شاخص استفاده کند (گرچه تا حد زیادی ابتدایی به نظر می‌رسد) یا اینکـه بایـد به عوامل دنیای واقعی توجه داشته باشد (کـه بسـیار زیـاد بـه نظـر مـی‌رسـد)، بلکـه سنجه‌های پرسشگر و شناختن محدودیت‌های آن ها اهمیت دارد.



شکل ۵-۲. تعداد دنبال‌کننده‌ها و تعداد روزآمدی‌های کتابخانه‌های دانشگاهی بریتانیا (آوریل ۲۰۱۳)

اگرچه وقتی اندازه‌گیری تأثیر یک پروفایل مطرح می‌شود حتی یـک مقایسـه سـاده از تعداد اتصال‌ها نیز با مشکلات فراوانی مواجه می‌شود، اما این موضـوع بـه ایـن معنـی نیست که اتصـال‌هـای مسـتقیم نمـی‌تواننـد اطلاعـات ارزشـمندی را از ارتبـاط بـین کنشگرهای مختلف نشان بدهند، مخصوصاً در مـواردی کـه انـواع متفـاوتی از رابطـه منتقل می‌شود. برای مثال، از طریق استفاده از فهرست هـای تـوییتر (جـایی کـه در آن یک کاربر افراد دنبال‌کننده چیزی را بخاطر گزینه انتخابی آنها گروه‌بندی مـی‌کنـد) یـا

پشتیبانی‌های لینکداین (جایی که در آن یک کاربر مهارت خاص یک ارتباط را پشتیبانی می‌کند). اتصال‌های غیر مستقیم نیز دارای طیف گسترده‌ای هستند: کاربران ممکن است دوستان مشترک داشته باشند، علایق مشترکی داشته باشند، در مؤسسه مشترکی آموزش دیده یا عضو یک سازمان باشند.

ظرفیت اتصال‌ها بیش از سنجه‌های ارزیابانه است، اما بیش‌تر بنیان‌گذار بررسی‌های ارتباطی هستند که در فصل ششم بررسی خواهند شد.

قابلیت مرور: مشاهده محتوا

یکی از جنبه‌های مهم مشارکت در وبگاه‌های شبکه‌های اجتماعی قابلیت مرور اتصال‌های بین پروفایل‌ها و مشاهده محتوای یک فرد دیگر است: اساساً محتوا ایجاد شده تا دیده شود. در برخی موارد داده ها قویاً از مشاهده یا عدم مشاهده محتوا حکایت دارند، و در سایر موارد باید به جایگزین‌ها اعتماد کرد. اگرچه خدماتی مانند یوتیوب و اسلایدشیر داده های مربوط به تعداد مشاهده های یک محتوا را عرضه می‌کنند، اما با تعداد مشاهده یک اظهار نظر به هیچ طریقی نمی‌توان از روزآمدی توییتر و فیسبوک مطلع شد. شاید به این دلیل باشد که روزآمدی زمانی ارسال می شود که تمام دوستان یا دنبال‌کننده های شخص در رفت و بازگشت بوده و زمانی که دوباره وارد شدند محتوای یک نفر با محتوای شخص دیگری جایگزین شده باشد. از این رو باید به شاخص‌های تقریبی که تعداد مشاهدات محتوای خاص را نشان می‌دهد اعتماد کرد. درحالی‌که ممکن است این حالت ایده آل نباشد اما از جهتی که شاخص نوع تفکر مردم از محتوا هستند مفیدند.

تعداد اتصال‌های یک کاربر می‌تواند قابل دسترس‌ترین نمادی باشد که نشان دهد چه تعدادی از مردم چه محتواهایی را دیده‌اند. اگرچه شاید انتظار همبستگی بین تعداد اتصال ها و تعداد مشاهده ها وجود داشته باشد، بسته به نوع محتوا و دلیل اتصال، تعداد مشاهده های هر بخش محتوایی در برگیرنده بخش بسیار کوچکی از تعداد کل اتصال‌ها است. این موضوع را می‌توان در وبگاه شبکه‌های اجتماعی مانند

یوتیوب بسیار مشاهده کرد چرا که هم تعداد اتصال‌ها و هم مشاهده‌های یک ویدئو به وضوح قابل مشاهده هستند. یک کانال تلویزیونی به نام کمدی 4Od[1] به عنوان یکی از کانال های یوتیوب از شرکت تلویزیونی کانال چهار[2] بریتانیا، در زمان نگارش این متن ۱۲۵۴۷۳ مشترک داشت، اما بسیاری از ویدئوهای بارگذاری شده تا چندین ماه پس از ارسال فقط صدها بازدیدکننده داشته‌اند.

تعداد اتصال ها نیز نمی‌توانند ظرفیت گزینه پربازدیدترین[3] که در بسیاری از وبگاه‌های شبکه‌های اجتماعی وجود دارد را نشان دهد. ممکن است شخصی در یوتیوب تعداد انگشت‌شماری مشترک داشته باشد، اما به دلیل بارگذاری یک ویدئوی جالب توجه بارها دیده و به اشتراک گذاشته شود. داشتن ویدئویی که میلیون‌ها بار مشاهده شده برای اشخاصی که تعداد مشترک زیادی دارند غیر عادی نیست. تعداد دفعاتی که سایر کاربران با آن مشغول بوده‌اند می‌تواند شاخص بهتری برای تعداد دفعات مشاهده محتوا باشد. این حالت ممکن است در توییتر توییت مجدد باشد، در فیسبوک لایک باشد یا در بیش‌تر وبگاه‌های شبکه‌های های اجتماعی اظهار نظر باشد. هر یک می‌توانند نمادی تقریبی باشند از تعداد دفعاتی که بخشی از محتوا مشاهده شده است، البته تعداد قطعی مشاهده ارجحیت دارد، هرچند که نحوه محاسبه تعداد مشاهده‌ها غالباً روشن نیست.

درک نحوه مشاهده و بازخورد محتوا بر روی دیدگاه کتابدارانی که محتوای خود را منتشر می‌کنند تأثیردارد. کتابداران باید توجه داشته باشند که آیا از ویدئوهای آموزشی استفاده می‌شود یا خیر، آیا عکس های بارگذاری شده مورد توجه جامعه کاربران آنها قرار گرفته، یا اینکه آیا اظهارنظرها آموزنده تلقی می‌شوند یا خیر. آمار بازدیدهای محتوایی می‌تواند بنیاد مطالعات وب‌سنجی، کتاب‌سنجی یا علم‌سنجی را نیز تشکیل دهند: چه محتوایی مورد توجه ویژه مردم است؟ این موضوع حاوی چه اطلاعاتی از رفتار دنیای واقعی است؟

1. 4oD comedy 2. Channel 4 3. Viral

نوع شناسی وبگاه‌های شبکه‌های اجتماعی

محتوای پروفایل، اتصال‌ها و بازدید محتوا همگی پایه و اساس سنجه‌ها را شکل می‌دهند، با وجود گستردگی و تنوع انتخاب وبگاه‌های شبکه‌های اجتماعی قابل دسترس، توجه به انواع سنجه‌هایی که می‌توانند مورد نظر کتابداران باشند باید بنا به کاربرد یک وبگاه شبکه‌های اجتماعی و نیز نوع ارزیابی باشد.

تلوال و استوارت (۲۰۱۰) سه نوع وبگاه شبکه های اجتماعی را شناسایی کرده‌اند: وبگاه‌های جامعه‌پذیر[1]، شبکه‌ای[2] و مروری[3]. وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیر وبگاه‌هایی هستند که از ارتباط‌ها و اتصال‌های اجتماعی غیر رسمی حمایت می‌کنند، مانند وبگاه هایی مانند فیسبوک و توییتر که با افکار اکثر مردم درباره وبگاه شبکه‌های اجتماعی هم‌خوانی دارند. وبگاه‌های شبکه‌های اجتماعی شبکه‌ای آن‌هایی هستند که برای ارتباط‌ها و اتصال‌های درون فردی رسمی طراحی شده‌اند، مانند وبگاه شبکه‌های اجتماعی تجاری لینکداین یا وبگاه شبکه‌های اجتماعی دانشگاهی ریسرچ گیت[4]. وبگاه‌های شبکه‌های اجتماعی مروری آنهایی هستند که با اتصال‌های شبکه‌های اجتماعی از کشف محتوا حمایت می‌کنند. نمونه آنها یوتیوب برای ویدئوها، فلیکر برای عکس‌ها، و اسلایدشیر برای ارائه‌ها هستند. کاربردهای وبگاه‌های شبکه‌های اجتماعی اغلب با هر یک از این دسته‌بندی‌ها متناسب است. برای مثال، اگرچه فیسبوک در وهله نخست یک وبگاه شبکه‌های اجتماعی جامعه پذیراست، که برای ارتباط با دوستان مورد استفاده است، اما از آن برای مرور محتوا یا برای ارتباطات رسمی با همکاران نیز استفاده می‌شود. به همین ترتیب، با وجودی‌که بسیاری از مردم از فلیکر برای دسترسی به محتوا استفاده می‌کنند، اشخاصی که در صنعت تصویرسازی مشغول هستند از آن می‌توانند برای ارتباطات و اتصال‌های رسمی‌تر استفاده کنند در حالی‌که سایرین می‌توانند از آن برای ارتباطات شخصی‌تر استفاده کنند. طرز استفاده از وبگاه‌های شبکه‌های اجتماعی را نمی‌توان از بالا دیکته کرد، بلکه بیش تر از اجتماع یا اجتماعات کاربران ناشی می‌شود زیرا آنها از

1. Socializing 2. Networking 3. Navigating 4. ResearchGate (www.researchgate.net)

کارآمدی‌های متفاوتی استفاده می‌کنند. در جایی‌که یک وبگاه شبکه‌های اجتماعی بخواهد رفتاری را به کاربران خود تحمیل کند، اغلب آن کاربران به سرعت به سمت وبگاهی بعدی که در حال پیدایش است رو می‌آورند. شبیه این رخداد زمانی اتفاق افتاد که وبگاه شبکه‌های اجتماعی اولیه فرندستر[1] به حذف حساب‌های کاربری جعلی اقدام کرد (بوید و الیسون، ۲۰۰۷).

در هنگام بررسی وبگاه‌های شبکه‌های اجتماعی این تلقی نباید در بین نباشد که وقتی دو حساب کاربری از یک خدمت رسانه های اجتماعی استفاده می‌کنند هر دوی آنها لزوماً استفاده یکسانی از آن وبگاه دارند. برای مثال، یک تحلیل محتوا از حساب‌های کاربری کتابخانه سازمانی روی توییتر نشان داد که افراد برای پیگیری اخبار و اطلاعات از حساب‌ها استفاده می‌کنند (استوارت، ۲۰۱۰). قدرت وبگاه‌های شبکه های اجتماعی در ماهیت تعاملی آن‌ها است و شایستگی کتابداران نیز جلب مراجعه کنندگان است (میلستن[2]، ۲۰۰۹؛ استوارت، ۲۰۱۰). فیلدز[3] (۲۰۱۰) به درستی اشاره کرد که خدمات نباید محدود به استفاده های تعاملی بوده، و برای برجسته‌سازی پرسش هایی که در میز مرجع پرسیده شده اند از یک پرسش توییتری استفاده کرد. از آنجایی که حساب کاربری دیگر قابل دستیابی نبود و دنبال‌کننده‌های بسیار کمی نیز داشت (پاپگروچ[4]، ۲۰۱۱) می‌توان فهمید که حفظ یک حضور غیر اجتماعی در شبکه‌های اجتماعی اساساً دشوار است.

وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیر

جامعه‌پذیری هدف اصلی اغلب وبگاه های شبکه های اجتماعی عمومی در جهان است. در حال حاضر کاربر فعال ماهانه توییتر و فیسبوک به ترتیب بیش از ۲۰۰ میلیون (توییتر، ۲۰۱۳) و یک میلیارد (فیسبوک، ۲۰۱۲) تخمین زده می‌شود. اکنون پیدا کردن پیوندهای حساب های کاربری سازمانی در صفحه خانگی یک کتابخانه بسیار طبیعی است. بازدید از صفحه وب اصلی کتابخانه هر یک از ۱۱۹ کتابخانه بریتانیا موجود در

1. Freindster 2. Milstein 3. Fields 4. Popegrutch

رتبه‌بندی فوریه ۲۰۱۳ *گاردین* (۲۰۱۳) نشان داد که ۷۷ مؤسسه حساب تـوییتری و ۵۹ مؤسسه حساب فیسبوک خود را به نحو جشمگیری نشان دادند.

جدول ۵-۱. حساب‌های کاربری رسانه‌های اجتماعی از صفحه‌های کتابخانه اصلی کتابخانه‌های دانشگاهی در بریتانیا، ۲۰۱۳

رسانه‌های اجتماعی	تعداد کتابخانه‌های دانشگاهی بریتانیا (از بین ۱۹۹کتابخانه)
توییتر	۷۷
فیسبوک	۵۹
یوتیوب	۲۱
فلیکر	۴
گوگل+	۴
فورسکوئر	۳
ویموو[1]	۲
پینترست[2]	۱
گدجت گوگل[3]	۱

همانگونه که در جدول ۵-۱ می‌توان دید دو وبگاه شبکه اجتمـاعی مشـابه دیگـر نیـز شناسایی شده اند: گوگل پلاس و فورسکوئر[4]. کاربرد رسانه های اجتماعی را مـی‌تـوان به این صورت دید که از توزیع قدرت-قانون پیروی مـی‌کننـد، بـه همـان ترتیبـی کـه بخش عمده‌ای از رفت و آمد وبی را بـه تعـداد انگشـت‌شـماری از وبگـاه‌هـا نسـبت می‌دهند، لذا جای تعجب نیست که بخش اعظمی از حساب‌های رسانه‌های اجتمـاعی را ببینیم که فقط دو خدمت را مورد توجه ویژه قرار دهند.

ایـن بررسـی اطلاعـات توزیـع رسـانه‌هـای اجتمـاعی جـاری مـورد اسـتفاده در کتابخانه‌هـای دانشـگاهی بریتانیـا را نشـان داده، و در ادامـه ایـن فصـل یـک نمونـه از حساب‌ها را برای نمایش ظرفیت سنجه‌های مختلف ارائه می‌شود. علاوه بر حساب‌های کاربری شخصی کتابداران، به احتمال زیاد حساب‌هـای کـاربری شـبکه‌هـای اجتمـاعی سازمانی مربوط به مؤسسات است. حداقل از منظر خدمات عمومی، پرطرفدارترین نـوع وبگاه شبکه‌های اجتماعی وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیر هستند.

1. Vimeo 2. Printerest 3. Google Gadget 4. Foursquare

می‌توان انتظار این برتری را داشت. چرا که وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیردر درجه نخست تعامل قرار دارند، و ایجاد پروفایل‌ها در چنین وبگاه‌هایی، باعث کانال‌های جدیدی برای ارتباط با یک کتابخانه یا کاربران خدمات اطلاعات می‌شوند. از این رو تحلیل‌های وبی برای وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیراحتمالاً حول محور نشانه‌های تعامل است. انعکاس این موضوع دنبال کردن یا عضویت در محتوای یک حساب کاربری، و مشارکت با دارنده حساب کاربری است. این موضوع می‌تواند شامل عکس‌العمل نسبت به محتوا، و هم محتوای ناخواسته باشد. تمام این تعامل‌ها نیز مثبت نیستند، و برای تهیه مفهوم سنجه‌ها یک تحلیل محتوایی یا تحلیل معنایی می‌تواند ضروری باشد. انتقادگرایی فقط در صورتی که تحت ردگیری و نظارت بوده، می‌تواند بازخورد مفیدی از خدمات اطلاعات باشد. بیش‌تر سازمان‌ها حجم مشخصی انتقاد دریافت می‌کنند، اما بهر حال این موضوع اهمیت دارد که آن‌ها از دیدگاه انتقادی در یک جهت تنظیم شوند تا بتوان بسرعت به آنها رسیدگی کرد.

وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیرمباحثه‌های بسیار گسترده‌ای را ترغیب می کنند، و کتابداران نه فقط باید به ارتباط مستقیم (اظهارنظرهایی که متوجه آن‌ها و محتوای آن‌هاست) علاقه مند باشند، بلکه باید به محتوای وبگاه شبکه های اجتماعی نیز علاقه مند باشند. از دیدگاه تحلیل وبی این موضوع شاید نیاز به نظارت داشته باشد که چگونه مردم در روابط خود از خدمات اطلاعاتی و کتابخانه یاد می‌کنند. از دیدگاه وب‌سنجی هر موضوعی که می‌تواند از فعالیت‌های دنیای واقعی اطلاعاتی بدهد نیاز به نظارت دارد. وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیر برای تحلیل وب‌سنجی منابع ارزشمند دقیقی هستند چرا که حاوی اطلاعاتی در خصوص هر چیزی است که می‌تواند موضوع مکالمه دو نفر باشد. همانگونه که در دادگاه‌های بریتانیا افزایش روزافزون تعداد پیگردهای قانونی مردم را برای اظهارنظرهای برخط می‌توان دید، مردم ندرتاً و به اندازه‌ای که در یک محیط عمومی توصیه می‌شود خود را سانسور می‌کنند.

وبگاه‌های شبکه‌های اجتماعی شبکه‌سازی

هنگام توجه به سنجه‌های مناسب برای شبکه‌سازی وبگاه‌های شبکه های اجتماعی، احتمالاً بجای کمیت تعامل ها توجه وافری به کیفیت تعامل‌ها خواهد شد. این موضوعی که برای یک شغل چند نفر درخواست داده‌اند یا منابع مالی که انتقال تماس یک وبگاه شبکه های اجتماعی شبکه‌ای را تامین می‌کنند مهم نیست، بلکه کیفیت کاربردها و مناقصه ها است که اهمیت دارند. مهم نیست که چه تعداد اتصال برخط ایجاد شده بلکه موضوع حائز اهمیت تعداد اتصال هایی است که در دنیای واقعی منجر به نشست‌ها و و همکاری‌ها می‌شوند. غالباً کیفیت اتصال‌ها دشوارتر از کمیت آنها است، هرچند می‌توان برای آن تلاش بسیاری کرد. بجای تاکید بر تعداد اتصال‌ها یا حتی تعامل‌ها، می‌توان فقط به تعامل‌های چندگانه‌ای پرداخت، که سایر کاربران ارزش اتصال های دریافتی خود را از طریق اجرای مجدد آن در چندین بار به نمایش می‌گذارند.

تنوع اتصال‌های شبکه می‌تواند کیفیت آن را نشان بدهد. بر اساس نظریه قدرت روابط ضعیف از گرانوتر[1] (۱۹۷۳)، که در آن برخی از ارزشمندترین تماس های ما مربوط به دوستان نزدیک ما که با آن‌ها نقاط مشترک فراوانی داریم نیستند، بلکه ارزشمندترین تماس اتصال‌هایی هستند که دانش مشترک کم تری با آن‌ها داریم، چرا که از هر مسئله‌ای چشم‌اندازها و ایده‌های جدیدی بروز می‌کند.

هرچند دلیل شناسایی نشدن خدمات و وبگاه‌های شبکه های اجتماعی شبکه‌ای در بررسی کتابخانه‌های دانشگاهی بریتانیا می‌تواند این باشد، که بیش‌تر حساب‌های کاربری شخصی هستند تا حساب‌های کاربری سازمانی(حتی اگر در موقعیت حرفه‌ای مورد استفاده باشند). ماهیت فعال جامعه کتابداران را روی وبگاه‌های شبکه های اجتماعی شبکه‌سازی می‌توان دید. برای مثال، در صفحه‌های گروهی لینکداین برای سازمان‌هایی چون سیلیپ و انجمن کتابداران آمریکا[2].

1. Granovetter 2. American Library Association

توجه کتابداران به وبگاه‌های شبکه های اجتماعی شبکه‌سازی نه تنها بخاطر چشم‌انداز تحلیلی آن، بلکه از این جهت است که اطلاعات جوامع خاصی از کاربران را ارائه می‌کند. هرچند تنوع محتوای یک وبگاه شبکه اجتماعی شبکه‌سازی می تواند بسیار کم‌تر از یک وبگاه شبکه اجتماعی جامعه‌پذیرباشد، هم‌چنین از ظرفیت اختلال کم‌تر و بررسی‌های متمرکزتر روی تعامل‌های حرفه‌ای برخوردار است. برای مثال، درحالی‌که می‌توان یک شبکه از فیزیکدانان نظری را هم در وبگاه شبکه‌های اجتماعی جامعه‌پذیر (مثل توییتر) و هم در وبگاه شبکه های اجتماعی شبکه‌سازی (مثل ریسرچ گیت[1]) بررسی کرد، وبگاه شبکه‌های اجتماعی جامعه‌پذیر اتصال‌ها و محتواهای غیرحرفه‌ای را دربرگرفته، و حتی به احتمال زیادتر تعامل بین فیزیکدانان نیز از نوع غیرحرفه‌ای است. ممکن است محتوا و اتصال‌های بیش‌تری مورد توجه پژوهشگر باشد، ولی در غیر این صورت، فقط وقفه‌ای در پژوهش می‌افتد.

وبگاه‌های شبکه‌های اجتماعی مروری

وبگاه‌های شبکه‌های اجتماعی مروری[2] به آن وبگاه‌هایی گفته می‌شود که پشتیبان کشف محتوا از طریق اتصال‌های شبکه‌های اجتماعی است. جای تعجب نیست، که بر خلاف وبگاه‌های شبکه های اجتماعی جامعه‌پذیری و وبگاه‌های شبکه های اجتماعی شبکه‌سازی، وبگاه‌های شبکه های اجتماعی مروری بر تعداد یا قدرت اتصال‌ها تمرکز کم تری داشته (هرچند ممکن است که این ها یک عامل کمکی باشند) و بیش‌تر بر خلق محتوا و هدایت کاربران به آن‌ها متمرکزند.

از دیدگاه تحلیل وبی کتابداران به تأثیر محتوای خودشان علاقه مندند: نظرها درباره یوتیوب، فلیکر یا پینترست، و واکنش‌ها به وبگاه‌های نشانه‌گذاری[3] اجتماعی آنها و رخدادهای ارسال شده اجتماعی. اما، در وبگاه‌های شبکه های اجتماعی جامعه پذیرو شبکه‌سازی، آن‌ها امکانات اطلاعات گسترده‌تر را نیز فراهم می‌کنند:

1. ResearchGate 2. Navigation social network sites 3. bookmarking

نوشته‌ها و قیمت‌ها در وبگاه ای بِی درباره اقتصاد چه می‌گویند؟ گزینه‌های عکاسی در طی زمان چگونه تغییر می‌کنند؟

رسانه‌های اجتماعی بی قاعده – ویکی‌پدیا

ویکی‌پدیا، حداقل با توجه به تعریف بوید و الیسون (۲۰۰۷) می‌تواند یک وبگاه شبکه اجتماعی محسوب شود. کاربران در این وبگاه می توانند در یک سیستم محدود یک پروفایل عمومی بسازند، هم چنین ارتباط‌های بیانی با سایر کاربران و هدایت اتصال‌هایی با دیگران ایجاد می‌کنند. توجه به ابعاد شبکه های اجتماعی کم تر از محتوایی است که ایجاد می‌شود. گرچه این موضوع می‌تواند برای یک وبگاه شبکه اجتماعی مروری نیز به وجود آید، در ویکی‌پدیا محتوا مجزای کاربرانی که آن را ایجاد کرده‌اند در نظر گرفته می‌شود. درحالی‌که می‌توان مشارکت‌های خاص هر کاربر را شناسایی کرد، چراکه صفحه به صورت تکراری ایجاد شده و به صورت یک کل دیده می‌شود، اما در نظر گرفتن بخش‌هایی از صفحه به عنوان اجزای مجزا با نویسنده های متفاوت به همان صورتی که یک ویدئو یوتیوب یا نمایش یک اسلایدشیر در نظر می‌گیرد بی‌معنی است.

در حالی که بعید است کتابداران یک ویکی داخلی ایجاد کنند که داده‌های کافی برای بررسی‌های علم‌سنجی یا وب‌سنجی داشته باشد، اما ویکی‌پدیا به عنوان ششمین وبگاه دارای رتبه‌بندی جهانی (بر اساس الکسا) می‌تواند برای طیف گسترده‌ای از بررسی‌های علم‌سنجی یا وب‌سنجی ارزشمند باشد چرا که میلیون‌ها کاربر در ویکی‌پدیا مشارکت دارند.

پژوهش‌ها و ابزارهای خدمات و وبگاه‌های خاص

در این بخش به برخی از سنجه‌ها، ابزارها و پژوهش های مربوط به چند وبگاه شبکه اجتماعی می‌پردازیم.

توییتر

توییتر به عنوان زمینه‌ای برای وب‌نویسی کوچک[1] و وبگاهی از نوع شبکه اجتماعی جامعه‌پذیر یکی از پرطرفدارترین خدمات رسانه‌های اجتماعی است. در مقایسه با سایر وبگاه‌های شبکه‌های اجتماعی کاربرد اصلی آن نسبتاً ساده بوده، و تمرکز اصلی آن بر انتشار روزآمدسازی‌های وضعیت ۱۴۰ نویسنده است. اندازه توییت[2] در ابتدا به گونه‌ای ایجاد شد که پیام‌ها می‌توانست از طریق قابلیت‌های پیام متنی تلفن‌های همراه ارسال شود و از ابتدا باز بودن و دسترسی به داده‌ها ضمیمه توییتر شد تا بتوان برای تعامل با این خدمت از بسترهای مختلف برنامه کاربری ایجاد کرد. ترکیبی از تعداد زیادی از کاربران توییتر و ۴۰۰ میلیون روزآمدسازی‌های وضعیت روزانه‌ای که ظاهراً درباره هر موضوع قابل تصوری ارسال می‌شوند، و در راستای تعداد زیاد رابط‌های برنامه‌نویسی موجود، توییتر نقطه تمرکز بررسی‌های زیادی با موضوع‌های بسیار متنوع بوده است.

مانند گوگل ترندز، از توییتر نیز در بررسی‌های بهداشت عمومی و اقتصادی بسیاری استفاده می‌شود: زانگ، فوهرس و گلوور[3] (۲۰۱۲) دریافتند که توییتر نه تنها با نوسانات بازار مالی همبستگی دارد، بلکه حتی می‌تواند پیش‌بینی کند. آچرکار و همکارانش[4] (۲۰۱۱) دریافتند که توییتر یک شاخص زمان واقعی از بیماری‌هایی چون سرماخوردگی عرضه می‌کند. توییتر امکاناتی فراهم می‌کند تا از رفتارهای طیف گسترده‌ای از مردم اطلاعاتی بدست آید: گلدر و مسی[5] (۲۰۱۱) نشان دادند که چگونه حال و احوال مردم در توییتر در طول روز تغییر می‌کند. کانینگهام[6] (۲۰۱۲) از ابزارهای آن دو برای انجام پژوهش استفاده کرد تا نشان دهد که مصرف الکل و سیگار می‌تواند در توییتر منعکس شود، و پیشنهاد کرد که می‌توان از آن برای بررسی تأثیر سیاست‌های سلامت عمومی استفاده کرد.

1. microblogging
2. tweet
3. Zhang, Fuehres and Gloor
4. Achrekar et al.
5. Golder and Macy(http://timeu.se)
6. Cunningham

ماهیت شبکه‌های اجتماعی توییتر امکان بررسی‌های سطوح بیش‌تری را نیز می‌دهد. زمانی که هیملبویم، مک کریری و اسمیت[1] (۲۰۱۳) شبکه‌های توییتر را برای ده موضوع سیاسی حساس ترسیم کردند آنها خوشه هایی از افراد را با عقاید سیاسی مشابه یافتند. در فصل ششم درباره توییتر به عنوان منبعی برای بررسی‌های منطقی مفصل‌تر بحث خواهد شد.

استفاده از توییتر در داخل جامعه کتابداری کانون توجه بررسی‌های گوناگونی بوده است بطوری که پژوهشگران تلاش کرده‌اند تا مشخص کنند که کتابخانه‌ها چگونه از توییتر استفاده می‌کنند (مانند استورات، ۲۰۱۰؛ دل بوسکو، لیف و اسکارل[2]، ۲۰۱۲)، کسانی که حساب کاربری توییتر یک کتابخانه را دنبال می‌کنند (اسول[3]، ۲۰۱۲)، و هم چنین نحوه انتشار این اطلاعات به سایر کاربران توسط دنبال‌کننده‌ها (کیم، ابلس و یانگ[4]، ۲۰۱۲). اگر چه بیش از یک وبگاه شبکه اجتماعی جامعه‌پذیر، کتابداران از توییتر برای طیف گسترده‌ای از فعالیت‌ها استفاده کرده‌اند، از تأکید بر منابع کتابخانه و اخبار گرفته تا تعامل با سایر کاربران، و سنجه‌های مناسب برای بررسی تأثیر محتوای خود کتابداران قویاً منوط به گزینش خود آنها دارد.

ابزارهای توییتر

در نتیجه رواج توییتر و باز بودن نسبی داده‌های آن، تعدادی ابزار برای بررسی داده‌ها ساخته شده‌اند. تعدادی از این ابزارها برای این طراحی شده‌اند که از تأثیر یک حساب کاربری فردی اطلاعاتی ارائه کنند، و برخی دیگر برای دسترسی به داده‌ها از کل جریان توییتر طراحی شده‌اند. به همین دلیل بدون هیچ مهارت برنامه‌نویسی می‌توان طیف وسیعی از بررسی‌ها را انجام داد.

در ساده‌ترین سطح، ابزارهایی هستند که خدمات هشدار ارائه می‌کنند مانند توییت‌بیپ[5] که می‌تواند به اشاره‌های نام کاربران یا وبگاه‌ها یک کاربر هشدار دهد یا

1. Himelboim, McCreecy and Smith
2. Del Bosque, Leif and Skarl
3. Sewell
4. Kim, Abels and Yang
5. Tweet Beep (http://tweetdeep.com)

فلورس[1] که به کاربر امکان ردگیری افرادی را می‌دهد که در طی زمان او را دنبال می‌کنند (یا دنبال نمی‌کنند).

اغلب کاربران تنها به هشدار داده شده محتوای برخط علاقه‌مند نیستند، بلکه می‌خواهند تأثیر حساب کاربری خود را بدانند. رتبه‌دهنده توییت[2] برای هر نام کاربری توییتری که وارد می‌شود یک رتبه ساده ارائه کرده و یک نمره رتبه ای که تناسب افراد رتبه‌دار را با توجه به الگوریتم رتبه توییت امتیاز پایین‌تری می گیرند نشان می‌دهد. این الگوریتم بر اساس شش عامل است، هر چند که رتبه‌دهنده توییت نحوه اعمال وزن‌دهی هر یک از را منتشر نمی‌کند: تعداد دنبال‌کننده‌ها، قدرت دنبال‌کننده‌ها (آن‌هایی که رتبه بالایی دارند در عوض رتبه بالاتری می‌دهد)؛ تعداد روزآمدسازی‌ها؛ نحوه جدیدترین روزآمدسازی حساب کاربری؛ دنبال‌کننده و دنبال کردن اخیر؛ و مشغولیت (مانند توییت های مجدد و اشاره‌ها). چنین خدماتی گاهی اطلاعاتی را به صورت رایگان ارائه می‌کنند و برخی نیاز به عضویت برای یک حساب کاربری حرفه‌ای دارند. برای مثال، ریتوییت‌رنک[3] رتبه‌بندی خود را به صورت رایگان ارائه می‌کند اما سایر سنجه‌ها مانند دسترسی و افشاگری[4] تنها برای کسانی قابل دسترسی است که حساب کاربری حرفه‌ای دارند. رتبه‌بندی ریتوییت‌رنک بر اساس تعداد توییت‌های مجدد، تعداد دنبال‌کننده‌ها، دوستان و فهرست‌هایی است که یک کاربر در آنها قرار دارد. با این همه، وزنی که برای هر یک از آن عوامل اعمال می‌کند را ارائه نمی‌شود.

این گونه رتبه‌بندی‌ها تا حدود زیادی بی‌معنی هستند. اینکه حساب کاربری توییتر من دارای رتبه ریتوییت‌رنک ۴۷۹۰۱۱ است که آن را در ۹۱/۲۳ صدک رتبه‌بندی می‌کند یا دارای رتبه رتبه‌دهنده توییت ۱۶۸۰۹۳۷ و نمره ۸۸ از ۱۰۰ است اطلاعات ارزشمندی نمی‌دهد. بدون آگاهی از وزنی که مربوط به رتبه‌بندی‌های متفاوت است، تشخیص اینکه آیا رتبه، هدف حساب کاربری را به طرز مناسبی نشان می‌دهد یا خیر غیر ممکن است. اگر برای به اشتراک‌گذاری اخبار از یک حساب

1. FIIwrs (http://fllwrs.com)
2. Tweet Grader (http://tweet.garder.com)
3. Retweet Rank (www.retweetrank.com)
4. reach and exposure

کاربری توییتری استفاده شود شاید این احساس را بوجود آورد که تعداد توییت‌های مجدد مهم است، اما چنانچه در درجه اول برای آدرس‌دهی جستجوهای کاربر استفاده شود در آن صورت تعداد اشاره های سنجه می‌تواند مهم باشد. رتبه‌بندی‌ها بافت کاربران مشابهی را نیز نیاز دارند. در رتبه ریتوییت و رتبه‌دهنده توییت لازم است که هر یک از اسامی به نوبت وارد شوند

برای تحلیل بررسی‌های گسترده‌تر و با جزئیات بیش‌تر اغلب بارگیری محتوا از توییتر لازم است. این موضوع می‌تواند با استفاده مستقیم از مجموعه جامع رابط‌های برنامه‌نویسی توییتر، یا ابزارهای دسک‌تاپی[1] که بر اساس آنها ساخته شده‌اند تحقق یابد. رابط‌های برنامه‌نویسی توییتر می‌توانند به اطلاعات جاری در خصوص شبکه های دوستان یک کاربر، دنبال‌کننده‌ها، توییت‌ها و سایر اطلاعات مرتبط و به کل جریان عمومی توییتر دسترسی بدهد. جمع‌آوری هر توییتی که با یک آیین‌نامه پاکسازی خاص مطابقت دارد یا نمونه ای از توییت‌های عمومی امکان پذیر است. با مجوز ویژه حتی می‌توان به آتش‌نشانی توییتر[2] نیز دسترسی یافت، به گونه‌ای که می‌توان به هر توییتی که تاکنون منتشر شده دست یافت، که البته در بیش‌تر موقعیت‌ها این دسترسی مناسب نیست. برای بارگیری خودکار اطلاعات نیز ابزارهایی وجود دارد. با وبومتریک آنالیست[3] می‌توان جستجوی توییتری را در هر ساعت و بارگیری توییت‌های جدید را انجام داد، در حالی که قالب اکسل[4] نودکسل[5] برای بارگیری شبکه‌های داده‌ای توییتر طراحی شده است. در فصل ششم هنگام پرداختن به تحلیل شبکه‌ای مجدداً به این دو ابزار رجوع خواهیم کرد.

فیسبوک

فیسبوک مردمی‌ترین وبگاه شبکه های اجتماعی جهان بالغ بر یک میلیارد کاربر فعال ماهانه دارد. فیسبوک در مقایسه با توییتر متن- محور امکان به اشتراک‌گذاری طیف گسترده‌تری از محتوا را داده، و بستری را برای ابزارها و برنامه‌های کاربردی گروه-

1. desktop
2. Twitter firehouse
3. (http://lexiurl.wlv.ac.uk)
4. Excel
5. NodeXL (http://nodexl.codeplex.com)

ثالث فراهم می‌کند. فیسبوک علاوه بر سنجه‌های پروفایل یک کاربر (مانند تعداد دوستان، تعداد اظهارنظرهای موجود در دیوار[1] یک کاربر و لایک‌های محتواهای آنها)، امکان خلق صفحه‌هایی را برای ساخت انجمنی حول موضوع یا مؤسسه خاصی نیز میسّر می‌کند. به همین دلیل در فیسبوک بعد از اینکه حداقل ۳۰ نفر در آن صفحه اظهار لایک کردند شناخت های مضاعفی عرضه می‌کند. این شناخت‌ها سنجه‌های مضاعفی مانند تعداد افرادی که از صفحه بازدید می‌کنند، تعداد نفراتی که یک نوشته را می‌بینند و تعداد اظهارنظرهایی که یک صفحه دریافت کرده است را فراهم می‌کنند. این داده شناختی به شکل تجمیع شده بوده و افراد را نمی‌توان شناسایی کرد و اگر شخص بازدیدکننده عضو فیسبوک نباشد (یا ارتباط برقرار نکرده باشد) در تحلیل‌ها لحاظ نمی‌شود. به روش‌های زیادی صفحه‌ها جایگزین گروه های پیشین شده‌اند هرچند گروه‌ها از این مزیت برخوردارند که می‌توانند با هم جمع شده و باعث شوند تا کاربران در آن‌ها عضو شوند به جای اینکه فقط لایک کنند.

درحالی‌که از ابتدا توییتر دارای ساختاری باز بود، فیسبوک بر اساس پیش‌فرض ابزاری شخصی بوده، و در آن به بسیاری از اطلاعاتی که می‌تواند بنیان یک بررسی سنجه وبی ارزشمند باشد دسترسی نیست. با این وجود، برای صفحه‌های عمومی برخی از داده‌های مفید به قرار زیر است:

- تعداد اشخاصی که درباره صفحه صحبت می‌کنند؛
- تعداد اشخاصی که صفحه را لایک می‌کنند؛
- پرطرفدارترین هفته؛
- پرطرفدارترین گروه سنی؛
- تعداد عکس‌هایی که به صفحه پیوست شدند؛
- پر بازدیدترین هفته؛
- بیش‌ترین تعداد اشخاصی که در یک زمان از صفحه بازدید می‌کنند.

1. wall

به‌هرحال، با این آمارهای محدود می‌توان بین چند صفحه فیسبوک مقایسه انجام داد. جدول ۵-۲ (صفحه بعد) ده مورد از برترین صفحه فیسبوک کتابخانه‌های دانشگاهی بریتانیا که بر اساس تعداد کل لایک های دریافت شده صفحه رتبه‌بندی شده‌اند را نشان می‌دهد. اگرچه به جای اینکه به پرسش مؤسسه‌ای با موفق‌ترین صفحه فیسبوک پاسخ دقیق بدهد، در واقع پرسش‌های بیش‌تری به وجود می‌آورد: اگر کتابخانه دانشگاه گلاسگو بیش‌ترین تعداد لایک‌ها را دارد، چرا تعداد افرادی که درباره ال.اس.ای صحبت می‌کنند دو برابر افرادی است که درباره آن دانشگاه حرف می‌زنند؟ چرا عکس های الحاقیه دانشگاه ساوسمپتون بسیار کم است؟ و علت بازدیدکننده‌های بسیار زیاد غیر عادی دانشگاه پورتسماوس[1] در یک هفته چیست؟

جدول ۵-۲. آمار صفحه‌های فیسبوک برای ۱۰ کتابخانه دانشگاهی بریتانیا که صفحه آن‌ها بیش‌ترین تعداد لایک را در ژوئن ۲۰۱۳ داشت

کتابخانه	تعداد کل لایک‌ها	اشخاصی که درباره این صحبت می‌کنند	عکس‌هایی که اینجا پیوست شده‌اند	بیش‌ترین بازدیدکننده در یک هفته	ورود بزرگ‌ترین گروه
دانشگاه گلاسگو	۴۶۶۲	۴۵	۱۳۸۵	۷۷	۶
دانشکده اقتصاد لندن	۴۲۷۷	۱۰۷	۲۷۹۵	۱۳۰	۹
دانشگاه وارویک	۴۲۴۲	۳۸	۱۳۳۹	۶۴	۱۰
دانشگاه پورتسماوس	۳۲۸۲	۲۱	۲۰۲۲	۳۶۲	۱۳
دانشگاه لیکستر[2]	۳۲۰۴	۵۰	۲۳۶۰	۹۴	۱۴
دانشگاه برونل[3]	۲۷۸۷	۲۶	۱۰۴۴	۸۷	۱۷
دانشگاه دورهام	۲۷۶۵	۲۵	۱۳۲۱	۵۱	۳۴
دانشگاه ساوسمپتون	۲۱۳۴	۱۰	۸۹	۱۲	۴
دانشگاه شفیلد	۲۰۱۲	۲۳	ناموجود		
دانشگاه ساسکس[4]	۱۹۸۰	۴۱	۱۸۷۶	۷۶	۱۵

تمرکز هزاران مطالعه دانشگاهی بر فیسبوک است، هرچند به دلیل ماهیت متفاوت محتوای آنها و نیز تفاوت در استخراج اطلاعات از وبگاه شبکه‌های اجتماعی، کلاً بر

1. University of Portsmouth
2. University of Leicester
3. Brunel University
4. University of Sussex

رفتار مجموعه‌های خاصی از کاربران یا رفتار در صفحه‌هایی خاص متمرکزند، به جای تمرکز بر ظرفیت فیسبوک برای شاخص‌های جهانی به شیوه روندهای سرماخوردگی گوگل متمرکز هستند. برای نمونه، ژانگ، هی و سنگ[1] (۲۰۱۳) ۱۳۵۲ پیام ارسال شده یک گروه دیابتی فیسبوک را تحلیل کردند، در حالی‌که پونس و همکارانش[2] (۲۰۱۳) پروفایل‌های عمومی فارغ التحصیل‌های علوم پزشکی را بررسی کردند تا نشان بدهند که آیا محتوای منتشر شده آنها مطابق دستورالعمل شورای اعتباربخشی آموزش پزشکی و تخصصی[3] غیرحرفه‌ای هست یا خیر. فیسبوک منبع محتوایی غنی را برای کتابداران، نه فقط برای بررسی تأثیر صفحه‌های خود کتابداران و سازمان‌های مشابه، بلکه برای بررسی صفحه‌هایی با محتوای مرتبط با کاربران آن‌ها، ارائه می‌دهد.

ابزارهای فیسبوک

مانند توییتر، خدماتی را برای رتبه‌بندی پرطرفدارترین صفحه‌های فیسبوک، بر اساس بیش‌ترین تعداد لایک یا صحبت افراد درباره یک صفحه (چون www.pagedatapro.com) ارائه می‌کنند، هرچند که مثل تمامی رتبه‌بندی‌های این‌چنینی باید جانب احتیاط را داشت. مقایسه‌ها فقط وقتی ارزشمندند که بین سازمان‌های مشابه انجام شده و بر اساس اهداف یکسان نیز از صفحات استفاده شود. قالب نودکسل افزونه‌هایی دارد[4] که امکان بارگیری شبکه‌ها را برای صفحه‌ها، گروه‌ها و شبکه‌های شخصی فیسبوک را فراهم می‌کنند. در فصل آتی به این نرم‌افزار باز خواهیم گشت.

همانگونه که در بیش‌تر موارد مطرح است، شاید یک پژوهشگر نتواند از ابزارهای عمومی رایگان برای کاربرد ضروری مورد نیاز بهره بگیرد، و لذا راه‌اندازی یک وسیله سفارشی می‌تواند ضروری باشد. فیسبوک از طیف وسیعی از رابط‌های برنامه‌نویسی[5] برای تعامل با صفحه‌های شخصی و نمودار باز برخوردار است. درحالی‌که برخی از این

1. He and Sang 2. Ponce et al 3. Accreditation Council for Graduate Medical Education
4. (http://socialnetimporter.codeplex.com) 5. (https://developers.facebook.com/docs/reference/apis)

رابط‌ها نیاز به احراز هویت دارند، با این وجود آنها را می‌توان با مرورگر مخصوص[1] بررسی کرد (https://developers.facebook.com/tools/explorer).

یوتیوب

احتمالاً کتابدارانی که از یک وبگاه شبکه اجتماعی مروری استفاده می‌کنند علاقه خاصی دارند که بدانند چند بار از محتوا بازدید شده است، و این موضوع وابستگی شدیدی به قابلیت وبگاه دارد. یوتیوب یک وبگاه اشتراک ویدئویی است، و پرطرفدارترین وبگاه شبکه اجتماعی مروری شناخته شده برای بررسی کتابخانه‌های دانشگاهی بریتانیا بوده که دارای طیف گسترده‌ای از تحلیل‌گرهای موجود می‌باشد[2]. همین امر منجر به ارائه اطلاعاتی می‌شود که کتابداران علاقه‌مند به تأثیر محتوا می‌خواهند در باره بازدیدها، ویژگی‌های جمعیت‌شناسی، اظهارنظرها، لایک‌ها، تنفرها، اشتراک‌ها و علایق بدانند.

یوتیوب علاوه بر جذابیت از منظر تحلیلی وب، نقطه تمرکز چندین مطالعه دانشگاهی نیز بوده است. این مطالعه‌ها در فیسبوک در وهله اول حول موضوع‌ها و گروه‌های کاربری مشخصی بوده‌اند. مراقبت‌های بهداشتی موضوع پرطرفداری است برای بررسی، چرا که پژوهشگران صحت ویدئویی موضوع‌های متفاوت سلامتی را ارزیابی می‌کنند. برای مثال، بیماری‌های التهابی روده (موکوار و همکارانش[3]، ۲۰۱۳)، برداشتن لوزه کودکان (استریچووسکی و همکارانش[4]، ۲۰۱۳) و بیماری صرع (ونگ، استیونسون و سلوا[5]، ۲۰۱۳). سایر حوزه‌هایی که موضوع مطالعه‌های اخیر بوده‌اند شامل ماهیت اظهارنظر سیاسی در کانال یوتیوب کاخ سفید[6] (هالپرن و گیبس[7]، ۲۰۱۳) و یک کانال یوتیوب کتابخانه‌ای به عنوان ارائه بخشی از طرح رسانه‌های اجتماعی گسترده‌تر آن (ووکوویچ و همکارانش[8]، ۲۰۱۳).

1. Graph API explorer
2. (www.youtube.com/analytics)
3. Mukewar et al.
4. Strychowsky et al.
5. WongT Stevenson and Selwa
6. Whitehouse
7. Halpern and Gibbs
8. Vucovich et al.

انواع قدیمی‌تر بررسی‌های سبک اطلاع‌سنجی[1] نیز موجود است. چودهاری و مکاروف[2] (۲۰۱۳) الگوهای رشد تعدادی از بازدیدهای دسته‌های مختلف ویدئوهای یوتیوب را بررسی کردند، و دریافتند که درحالی‌که بازدیدهای اولیه برای برخی دسته‌ها به عنوان شاخصی از اعتبار آتی محسوب می‌شود برای بقیه دسته‌ها این مورد مطرح نیست. سوگیموتو و همکارانش[3] (۲۰۱۳) به بررسی رابطه بین صحبت‌های مجری‌های تی.ای.دی[4] (مخفف فناوری، برنامه نمایشی، طراحی) و تأثیر این صحبت‌ها پرداخته و بازدیدها، اظهارنظرها و لایک‌ها را ارزیابی کردند. لایی و ونگ[5] (۲۰۱۳) به بررسی تأثیر محتوای یوتیوب در وبگاه‌های بیرونی بر تعداد بازدیدهای دریافت شده یک ویدئو پرداختند.

از آنجایی که وبگاه‌های شبکه های اجتماعی مروری برای کشف محتوا طراحی شده‌اند، این بحث می تواند پیش بیاید که جامعه کتابداران به آن‌ها توجه ویژه‌ای دارند، هر چند که بررسی‌های مربوط به قابلیت اعتمادپذیری ویدئوهایی که برای کلیدواژه‌های خاص بوده، یا ماهیت اظهارنظرهای کانال های بخصوص، می‌توانند بدون هیچ ابزار تخصصی محقق شوند (البته دقیق و پر زحمت). ابزارهای جمع‌آوری داده‌های خودکار می‌توانند بررسی‌های جامع‌تری را میسّر سازند.

ابزارهای یوتیوب

یوتیوب همانند بسیاری از وبگاه‌های شبکه های اجتماعی مجموعه جامعی از رابط‌های برنامه‌نویسی را برای توضیح محتوای خود ارائه می‌کند (کارکردی که ممکن است از طریق مرورگر گوگل بررسی شود[6]) و مورد استفاده تعدادی از وسیله ها است. وبومتریک آنالیست[7] امکان بارگیری طیف وسیعی از محتوا را، از اطلاعات مجموعه ای از ویدئوهای کاربران گرفته تا اظهارنظرهای همراه با مجموعه‌ای از ویدئوها، و هر شبکه‌های مرتبطی را میسّر می‌کند. در بخش زیر به قابلیت وبومتریک

1. informetric
2. Chowdhury and Makaroff
3. Sugimoto et al.
4. Technology, Entertainment, Design
5. Lai and wang
6. (https://developers.google.com/apis-explorer)
7. (http://lexiurl.wlv.ac.uk)

آنالیست برای جمع‌آوری اظهارنظرهای یوتیوب خواهیم پرداخت، که از آن برای پایه ریزی اساس یک تحلیل عاطفی استفاده می‌شود. قالب نودکسل امکان بررسی شبکه‌های یوتیوبی را می‌دهد که بر مبنای یک شبکه کاربر یوتیوب و یک شبکه ویدئویی هستند (مانند دسته اشتراکی یا اظهارنظرهای کاربران مشترک شده).

همانگونه که در فصل چهارم ذکر شد، گوگل ترندز اطلاعات جستجوی یوتیوب را نیز ارائه می‌کنند. چنین دیدگاهی می‌تواند منجر به بررسی‌های بیش‌تری شده و تحلیل بیش‌تری ارائه کند برای موج مطالعاتی موجودی که به بررسی محتوای ویدئوها برای جستجوهای جستجوی خاص می‌پردازند. سابقاً، امکان شناسایی ویدئوهایی که پژوهشگر فکر می‌کند یک کاربر بجای جستجو برای نیاز واقعی خود چه چیزی را وارد می‌کند وجود داشت.

ویکی‌پدیا

ویکی‌پدیا با سایر پایگاه‌های وب رسانه‌های اجتماعی که در بالا مورد بحث قرار گرفته‌اند تفاوت چشمگیری دارد، چراکه با مشارکت اشخاصی که کلاً به خروجی مشترک علاقه‌مندی کم‌تری دارند، اعتبار و اندازه وبگاه را به مورد خاصی تبدیل کرده، و از سال ۲۰۰۵ ظرفیت آن برای پژوهش وب‌سنجی مورد اذعان قرار گرفته است (وُب[1]، ۲۰۰۵). در تابستان ۲۰۱۳ ویکی‌پدیا ششمین وبگاه در رتبه‌بندی جهانی بوده (الکسا، b۲۰۱۳) و بالغ بر ۴ میلیون صفحه به زبان انگلیسی و تمامی موضوعات قابل تصور را تحت پوشش داده است.

در ویکی‌پدیا مانند یوتیوب، محتواها به جای کاربر از اهمیت وافری برخوردارند، و در مقایسه با منابع ویرایش شده سنتی تعداد زیادی مقاله است که باید اعتبار مقاله‌ها بررسی شود. برای مثال، مقایسه نتایج مقاله‌های علمی (گیلس، ۲۰۰۵) و مقاله‌هایی درباره مواد مخدر (کلاسون و همکارانش، ۲۰۰۸) قابل قبول بوده اما برای مقاله‌های تاریخی نتیجه ضعیف‌تری مشاهده شده است (رکتور، ۲۰۰۸). اگرچه

1. Vob

کیمونس[1] (۲۰۱۱) مطرح می‌کند که تعداد ویرایش یا تعداد تجدید نظرهای مطالعاتی که فقط به بخش منتخب کوچکی از پژوهش‌ها پرداخته‌اند بسیار بیش‌تر از تعداد ویراستاران ویرایش یک مقاله معمولی است. با اینهمه، تنوع محتوا را نمی‌توان انکار کرد و به چندین زبان موجود است، و برای طبقه‌بندی اسناد به زبان لهستانی استفاده شده است (کیسیلکی و همکارانش[2]، ۲۰۱۲)، وقتی که پس از قطع قابلیت دامنه پیوند[3] در یاهو طیفی از وبگاه‌های شبکه‌های اجتماعی مورد بررسی قرار گرفتند، ویکی‌پدیا یکی از دو وبگاه شبکه های اجتماعی (همراه با مکان‌یاب اجتماعی دلیشس) بود که همبستگی بالایی را با تعداد کل پیوندهای داخلی با دانشگاه‌های اسپانیایی نشان داد (اوردنا–مالیا و اونتالبا–روپرز[4]، ۲۰۱۳).

گستردگی موضوعات ویکی‌پدیا به این معنا است که طیف متنوعی از عناوین ممکن است بررسی شده باشند: یک بررسی علم‌سنجی می‌تواند به بررسی طریقه‌ای بپردازد که تعداد ویرایش‌ها یا ویرایش‌گرهای ویکی‌پدیا رشد یک حوزه جدید را منعکس می‌کنند؛ انواع ویرایش‌های صفحه یک شرکت می‌تواند چشم‌اندازی مثبت یا منفی را برای شرکت منعکس‌کنند؛ یا بازدید از صفحه‌های مربوط به بهداشت می‌تواند منعکس‌کننده نگرانی‌های سلامتی عمومی را در یک بیماری خاص باشد. سوگیری‌های ویرایش‌گرهایی را باید مدّ نظر داشت، هر چند که همگی اغلب مرد و ساکن آمریکا شمالی یا اروپای غربی هستند (مؤسسه ویکی‌مدیا، ۲۰۱۱)، و جدول زیر می‌تواند بخشی از دلیل اختلاف عظیمی که در ویرایش و بازدید از دو صفحه برای وقایع یکسان در ۱۵ آوریل ۲۰۱۳ بود را نشان دهد: بمب‌گذاری ماراتون بوستون[5] و مجموعه‌ای از بمب‌گذاری‌های همان روز در عراق[6]. در جدول ۵-۳ (صفحه بعد) مقایسه‌ای از صفحه اطلاعات پنج روز پس از حملات این دو صفحه را نشان می‌دهد.

1. Kimmons 2. Ciesielski et al. 3. linkdomain
4. Orduna-Malea and Ontalba-Ruiperez
5. Boston Marathon bombings(http://en.wikipedia.org/wiki/Boston_Marathon_bombing)
6. (http://en.wikipedia.org/wikis/15_April_2013Iraq_attacks)

این مثال هم‌چنین نشان می‌دهد که سنجه‌های ویکی ممکن است بتوانند برای مفاهیم ایجاد شده‌ای چون ارزش خبری اطلاعات قابل ارزیابی فراهم کنند. بین دو مجموعه بمب‌گذاری تفاوت‌های بیش‌تری غیر از تعداد کشته‌ها و مجروح‌ها و موقعیت جغرافیایی آن‌ها وجود دارد به این دلیل که حملات عراق بخشی از شورش در حال پیشروی بود، درحالی‌که بمب‌گذاری‌های بوستون یک گزارش خبری ممتد بود که به عنوان شکار بمب‌گذارها در سراسر اخبار پخش شد. بهر حال براحتی می‌توان دید که یک مطالعه عمیق چگونه می‌تواند نظرها را بررسی طیف متنوعی از متغیرها در موضوعات متفاوت خبری در طی زمان جلب کند.

جدول ۵-۳. ویکی‌سنجی برای دو بمب‌گذاری در ۱۵ آوریل ۲۰۱۳

فوت	حملات عراق	بمب گذاری ماراتون بوستون
≥۷۵۰	۷۵۰≥	۳
زخمی	۳۵۰>	۱۸۳
تعداد ویرایش‌ها	۱۰۶	۳۸۰۸
تعداد نویسندگان متفاوت	۳۶	۶۶۳
تعداد بازدیدهای صفحه	۱۲۹۰۹	۴۲۲۵۲۰

ابزارهای ویکی‌پدیا

اساساً سوابق تمام ویرایش‌های تمامی صفحه‌های ویکی‌پدیا به صورت عمومی قابل رؤیت است و بازدید صفحات نیز به صورت عمومی در دسترس هستند[1]. این موضوع موجب شد که برای کسب اطلاعات در ویکی‌پدیا چند ابزار ایجاد شود:

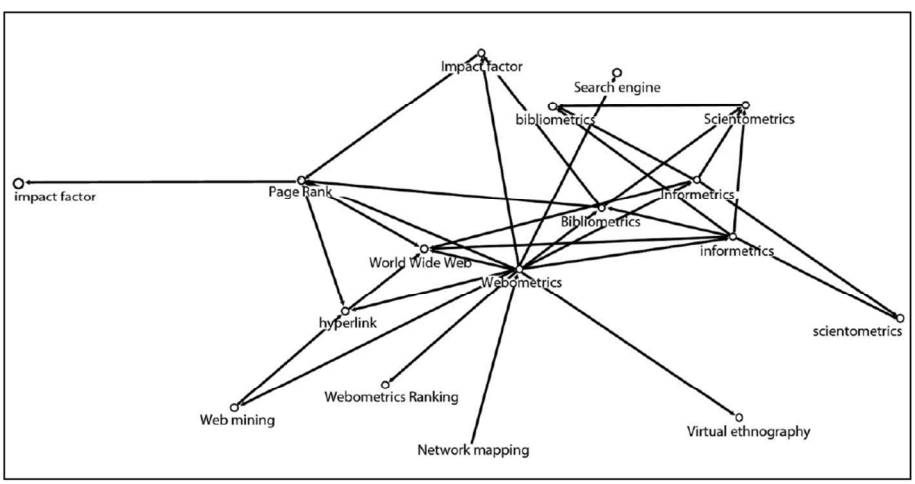
- *ویکی ترندز*[2] – مردمی‌ترین صفحه‌های ویکی‌پدیا، و نیز صفحاتی را نشان می‌دهد که در طی روز هفته یا ماه جاری برای تمام نسخه‌های زبانی ویکی‌پدیا با سرعت بسیار زیادتری در بین مردم می‌چرخد.
- *آمار رفت و آمدی مقاله ویکی‌پدیا*[3] – امکان دسترسی به آمار رفت و آمدی برای هر صفحه خاص را فراهم می‌کند.

1. (http://dumps.wikimedia.org/other/pagecounters-raw)
2. Wikitrends (http://toolserver.org/~johang/wikitrends)
3. Wikipedia Article Traffic Statistics(Http://state.grok.se)

- *آمار سوابق صفحه ویکی‌پدیا*[1] – اطلاعات ویرایش‌هایی که یک صفحه ویکی‌پدیا دریافت کرده، را خلاصه می‌کند و زمان انجام این ویرایش‌ها و کسی که بیش‌ترین ویرایش‌ها را انجام داده نشان می‌دهد.
- *تحلیل‌گر تعامل ویرایش‌گر*[2] – امکان مقایسه کار چند ویرایش‌گر در صفحه‌های مختلف را فراهم می‌کند.

هم چنین یک قالب نودکسل نیز وجود دارد که موجب بررسی شبکه‌های مدیاویکی می‌شوند[3]. این موضوع موجب ایجاد شبکه‌های ساده بر اساس ابرپیوندهای بین مقاله‌ها و همکاری نویسندگان می‌شود.

شکل ۵-۳ شبکه های مقاله ویکی‌پدیا را بر اساس پیوندهای بیرونی از صفحه وب‌سنجی نشان می‌دهد.



شکل ۵-۳. شبکه‌های مقاله ویکی‌پدیا بر اساس پیوندهای بیرونی از صفحه وب‌سنجی

گرچه ویکی‌پدیا داده‌های زیادی را قابل دسترس می‌سازد، که این اطلاعات لزوماً برای شخصی که قصد تحلیل آنها را دارد در قالب ارزشمدی نباشد. در عوض شاید

1. Wikipedia Page History Statistics(http://vs.aka-online.de/cgi-bin/wppagehiststat.pl)
2. Editor Interaction Analyzer (http://toolserver.org/~snottywong/editorinteract.html)
5. MediaWiki (http://wikiimporter.codeplex.com)

لازم باشد که از صفحه‌ای بازدید شده و داده‌ها به صورت را بصورت دستی جمع‌آوری کند، یا از همان صفحه با تعدادی رخدادهای مختلف بازدید مجدد کند.

وبگاه‌های شبکه‌های اجتماعی دیگر

مباحثه‌های رسانه‌های اجتماعی ابتدا تنوع زیاد وبگاه‌ها را خاطر نشان ساخته، و سپس به بحث درباره توییتر یا فیسبوک می‌انجامد. همانگونه که در بررسی کتابخانه‌های دانشگاهی بریتانیا مشاهده شد، زمانی که صحبت از کاربرد رسانه‌های اجتماعی می‌شود بحث‌های دیگری پیش می‌آید، که در بحث تفصیلی هر یک از این وبگاه‌ها کم‌تر تأکید می‌شود. با این وجود مبنای بسیاری از پژوهش‌های دانشگاهی وبگاه‌های شبکه‌های اجتماعی دیگر بوده، که تمرکز برخی از آنها به خصوص بر جامعه کتابداری است.

تورنتون[1] (۲۰۱۲) استفاده از پینترست را در کتابخانه‌های دانشگاهی بررسی کرده است. کووان و چان[2] (۲۰۰۹) ظرفیت پیوندزنی فالکسونومی از وبگاه نشانه گذاری اجتماعی دلیشس[3] را با سرعنوان‌های موضوعی کتابخانه کنگره[4] بررسی کردند. انگس، تلوال و استوارت (۲۰۰۸) سودمندی برچسب‌ها را در فلیکر برای تعیین ظرفیت آن به عنوان منبع تصویری دانشگاهی بررسی کردند. حتی اگر محتوای وبگاه شبکه های اجتماعی مورد توجه پژوهشگر نباشد، از شبکه می‌توان برای درک دیدگاه‌های استفاده کرد. برای مثال، اویلود و بامیگبولا[5] (۲۰۱۲) از لینکداین برای بررسی نظرات مردم برای نقش کتابخانه استفاده کردند. هر یک از این‌ها می‌توانند نقطه تمرکزی باشد برای بررسی یک کتابدار، چه در اهداف تحلیل وبی چه در اهداف وب‌سنجی.

کوتاه‌کننده‌های نشانی وبی- پیوندهای تحلیل وبی در هر وبگاه

سنجه‌ها در خدمات گروه- ثالث ممکن است با اطلاع‌رسانی کمی یا بدون هیچ اطلاع‌رسانی تغییر کنند، و کاهش ناگهانی اطلاعاتی شود که کتابداران می‌توانند از تأثیری که ایجاد محتوای آن‌ها داشته مطلع شوند. استفاده از خدمات کوتاه‌سازی

1. Thornton
2. Kwan and Chan
3. Delicious
4. Library of Congress Subject Headings
5. Oyelude and Bamigbola

نشانی وبی راه حل این موضوع در موقعیت‌های خاص است. بر اساس طراحی یک وبگاه و سیستم مدیریت محتوای خاص، یک نشانی وبی خاص ممکن است بسیار طولانی بوده، و هزاران نویسنده داشته باشد. به اشتراک‌گذاری این آدرس‌های اینترنتی بسیار طولانی می‌تواند برای افراد دشوار باشد، چه به علت اینکه رسانه‌های ارتباطی محتوا را مجدد شکل‌دهی کند (برای نمونه، ممکن است یک بستر پیام الکترونیکی یک نشانی وبی را در چندین خط پخش کند) و یا به این علت که فضای موجود بستر را برای ارتباط‌ها محدود کند (برای نمونه، توییتر پیام‌های کاربران را به ۱۴۰ نویسنده محدود می‌کند)، و خدمات کوتاه‌سازی نشانی‌های وبی را برای ساده‌سازی فرایند به اشتراک‌گذاری نشانی‌های وبی توسعه داده‌اند. یک خدمت کوتاه‌سازی نشانی وبی یک نشانی وبی کوتاهی را ارائه می‌کند، بطوری که هنگام بازدید کاربران را مجدداً به نشانی وبی اصلی هدایت کند.

فرایند هدایت مجدد نه تنها امکان به اشتراک‌گذاری آدرس‌های اینترنتی فوق‌العاده بلند را میسّر می‌کند، بلکه فرصتی را پیش می‌آورد تا تأمین‌کنندگان محتوا بتوانند اطلاعات کاربران را در حین سفر آن‌ها از یک وبگاه به وبگاه دیگر گردآوری کنند. برای مثال، شاید کتابداری بخواهد پیوند به مقاله جالبی را که گروه ثالث ارسال کرده به اشتراک بگذارد. اگر به لاگ‌های سرور یا تحلیل‌های وبی وبگاه هدف دسترسی وجود نداشته نباشد آن‌ها به سنجه‌هایی محدود می‌شوند که فیسبوک می‌تواند آن را از تعداد بازدیدکننده‌های مردم روی یک پیوند در دسترس قرار دهد. هرچند، در صورتی که از یک خدمت کوتاه‌کننده نشانی وبی استفاده شود، اطلاعات کاربران می‌تواند در حین کلیک کردن آن‌ها جمع‌آوری شود.

Bit.ly به عنوان پرطرفدارترین خدمت کوتاه‌کننده نشانی وبی، که امکان دسترسی بازدید هر از چند گاهی یک نشانی وبی کوتاه شده، نحوه به اشتراک‌گذاری یک نشانی وبی کوتاه شده توسط سایر مردم، تعداد دفعات کلیک بر یک آدرس اینترنتی، جای ارجاع یک نشانی وبی و موقعیت مکانی کاربران را نشان می‌دهد(شکل ۵-۴ را ملاحظه کنید).

هرچند، خدمات کوتاه‌سازی نشانی وبی ابزارهای کاملی نیستند، و کتابداران باید محدودیت‌های استفاده از خدمات کوتاه کردن نشانی وبی را مدّ نظر داشته باشند چرا که باید این کار را برای استفاده از هر محتوای برخطی که ایجاد می‌کنند اعمال نمایند. برای مثال، مشکلات احتمالی تنزل پیوند را در شرایطی که خدمات یک کوتاه‌ساز نشانی وبی قطع می‌شود یا سیاست آن تغییر کند باید در نظر داشت.

شکل ۵-۴. توزیع جغرافیایی کلیک‌های یک گزارش خبری BBC بر اساس مشاهده‌های Bit.ly

تأثیر عمومی رسانه‌های اجتماعی

علاوه بر وبگاه‌هایی که یک رتبه‌بندی یا یک درجه را برای یک حساب کاربری در یک شبکه اجتماعی مانند توییتر ارائه می‌کنند، وبگاه‌های دیگری نیز هستند که تحلیلی از تأثیر یک پروفایل شخصی یا سازمانی در چندین وبگاه ارائه می‌کنند. برای نمونه، کالوت[1] بر اساس داده‌های طیف گسترده‌ای از وبگاه‌های شبکه‌های اجتماعی امتیاز واحدی از ۱ تا ۱۰۰ عرضه می‌کند. این‌ها نه تنها شامل وبگاه‌های شبکه‌های اجتماعی جامعه‌پذیری چون فیسبوک و توییتر هستند، بلکه وبگاه‌های شبکه‌های اجتماعی شبکه‌سازی (مانند لینکداین)، وبگاه‌های شبکه‌های اجتماعی مروری (مانند یوتیوب، فلیکر و لست‌اف‌ام[2]) و وبگاه‌های وب‌نویسی نیز در بر می‌گیرند (مانند

1. Klout (http://klout.com) 2. LastFM

تامبلر[1]، بلاگر[2] و وردپرس[3]). در راستای ساده‌انگارانه رتبه‌بندی یک وبگاه واحد، یک رتبه‌بندی رسانه اجتماعی جهانی الزامی ندارد که رسانه‌های اجتماعی در یک موسسه خاص استفاده شود و مقایسه انواع متفاوتی از سازمان‌ها را ترغیب می‌کند. بعلاوه، یک رتبه‌بندی جامع رسانه‌ای می‌تواند رفتار خاصی را در بستری مشخص برای ارتقا رتبه خود ترغیب کند، و استفاده از خدمات نامناسب را نهی کند. برای مثال، برای اغلب کتابداران بعید است که شبکه لست‌اف‌ام به عنوان بخش منسجمی از ارائه خدمات کاربران باشد، هرچند اگر برای داوری در مورد تأثیر رسانه های اجتماعی پیشنهادی آن‌ها از یک امتیاز کالوت استفاده شود، آن‌ها شاید وسوسه شوند که برای آن وقت صرف کنند.

وی‌فالو[4] یک وبگاه رتبه‌بندی رسانه‌های اجتماعی عمومی دیگر است که حائز بررسی است. این وبگاه مانند کالوت رتبه‌بندی را در مقیاس ۱ تا ۱۰۰ انجام می‌دهد. این رتبه مانند پیج رنک در پیج رنک توییتر، فیسبوک، اینتساگرام[5] و لینکداین رفتار می‌کند. برخلاف کالوت، در این وبگاه کاربران مجاز به ابراز علایق خود هستند تا کتابداران بتوانند خود را با افراد دیگری که ادعای کتابداری دارند یا به کتابداری علاقه‌مندند مقایسه کنند. با این وجود، چنین رتبه‌بندی فقط باید به عنوان تفریح در نظر گرفته شود و هرگز قابل اعتماد نیست.

تحلیل عاطفی

یکی از بزرگ‌ترین مزایای وبگاه‌های شبکه‌های اجتماعی این است که امکان دسترسی به حجم عظیمی از داده‌ها را که دارای ساختاربندی مشابهی هستند فراهم می‌کنند. این بدان معنی است که تحلیل عاطفی ممکن است در حجم زیاد داده‌ها اعمال شده، تشخیص صحت آن با ترازهای انسانی است که متن‌ها مثبت هستند یا منفی. تحلیل محتوا در بسیاری از موارد هم‌چنان مناسب‌ترین روش است، یا به علت اینکه پرسشی که باید پاسخ

1. Tumblr
2. Blogger
3. Wordpress
4. Wefollow (http://wefollow.com)
5. Instagram

داده شود درباره احساسات نیست، یا اینکه نوع رسانه ها برای تحلیل عاطفی مناسب نیستند (مانند ویدئو یا تصویر). همچنین ممکن است در بسیاری از موارد کتابدار مایل به تعیین احساسات باشد (مانند اظهارنظرهای روی نوشته‌ای که خودشان ارسال کرده‌اند)، و شاید نمونه آنقدر کوچک باشد که بتوانند به صورت دستی تحلیل شود. البته، مواردی هستند که تحلیل عاطفی با مقیاس بزرگ‌تر برای آن‌ها لازم است.

از آنجایی که توییتر بیش از پیش به عنوان یک منبع بالقوه‌ای برای طیف گسترده‌ای از موضوعات، از وضعیت اقتصادی تا بهداشت عمومی، شناخته می‌شود، جای تعجب نیست که دریابیم خدماتی وجود دارد که تحلیل عاطفی توتیت‌های جدید را برای عبارت‌های جستجویی ویژه ارائه می‌کنند. برای مثال، سنتیمنت۱۴۰[1] امکان دسترسی سریع به احساسات همراه با توییت‌های جدید را برای یک عبارت یا اصطلاح خاص فراهم می‌کند. یک خدمت تحلیل عاطفی توییتر دیگر (http://smm.streamcrab.com) نه تنها تحلیل توییت‌های اخیر را فراهم می‌کند، بلکه توییت‌ها را بلافاصله به محض انتشار نشان می‌دهد، که برای موضوعات داغ روز و پرطرفدار مناسب‌تر است. اگرچه توییتر شاید به عنوان نبض وب در نظر گرفته شود، خدمات رسانه های اجتماعی دیگری نیز هستند که مورد توجه تحلیل عاطفی هستند. اشاره اجتماعی[2] اطلاعات احساسات مستتر در محتواهای مختلف را نشان می‌دهد (مانند، نشانه‌گذاری‌های اجتماعی، گزارش‌های خبری، اظهارنظرها و تصاویر) و حتی امکان بارگیری داده‌ها را نیز می‌دهد.

درحالی‌که شاید بتوان حجم مشخصی از داده‌ها را با استفاده از ابزارهای مبتنی بر وب تحلیل کرد، تحلیل‌های بیش‌تر غالباً به قدرت پردازش نرم‌افزار رایانه رو میزی نیاز دارند. طراحان زیادی تحلیل معنایی را انجام می‌دهند، اما درحالی‌که کدهای رایانه ای و منابع واژگانی زیادی قابل دستیابی هستند، ابزارهای رایگان قابل دسترسی برای کاربران غیر فنی کم است. یکی از ابزارهایی که برای استفاده غیر تجاری به صورت رایگان در

1. Sentiment140 (www.sentiment140.com)
2. Social Mention (http://socialmention.com)

دسترس است سنتی‌استرنت[1] از گروه سایبرسنجی آماری[2] در دانشگاه وولورهمپتون است. استفاده از فایل‌های عبارت‌های وزن‌داری[3]، که برای انعکاس خصوصیات یک مجموعه داده‌ای خاص می‌تواند به صورت خودکار و یا دستی درجه‌بندی کند، سنتی‌استرنت قادر است عواطف مجموعه‌ای از متن‌ها را در قالب یک فایل متنی ساده با یک متن در هر خطی به صورت خودکار دسته‌بندی کند.

رویکردهای متفاوتی برای تحلیل عاطفی وجود دارد؛ برای نمونه، سنتیمنت۱۴۰ از یک رویکرد یادگیری ماشینی بر مبنای تحلیل توییت‌هایی که با احساس همراه هستند استفاده می‌کند (گوو، هایانی و هانگ[4]، ۲۰۰۹)، درحالی‌که بقیه رویکردها هر جا که لغت‌ها مختص معنی خاصی هستند از واژه‌نامه استفاده می‌کنند (مانند سنتی‌وردنت[5]). همانطور که مکرراً خاطر نشان شده، برخی متون برای تجریه و تحلیل عاطفی مناسب‌تر از متون دیگر هستند، و برخی متون ممکن است برای رویکردهای خاصی مناسب‌تر باشند تا تحلیل عاطفی به نسبت بقیه رویکردها. از این رو، بررسی صحت ابزارهای تحلیل عاطفی با مجموعه داده ای خاص از این جهت مهم است که آیا آن ابزار شالوده یک بررسی وسیع است یا اینکه بر مبنای تحلیل عاطفی که با یک ابزار خاص انجام شده، تصمیم‌های مهمی گرفته می‌شود.

با این وجود، با نرم‌افزار تحلیل عاطفی مناسب می‌توان طیف گسترده‌ای از بررسی‌ها را انجام داد. برای مثال، در مطالعه موردی زیر از وبومتریک آنالیست و سنتی‌استرنت برای بررسی سریعی در خصوص واکنش مثبت مردم به ویدئوهای برخط کتابخانه‌های دانشگاهی استفاده شد. تحلیل عاطفی برای هر موقعیتی مناسب نیست، اما بدون شک ابزار ارزشمندی است برای تحلیل خودکار کمیّت‌های عظیم داده‌ای که بیش از پیش در دسترس پژوهشگران می‌باشد، البته برای برخورداری از نتایجی قاطع لازم است که نرم‌افزار آزموده شده و برای مجموعه خاصی از داده‌ها درجه‌بندی شود.

1. SentiStrength (http://sentistrength.wlv.ac.uk)
2. Statistical Cybermetrics Group
3. weighted terms
4. GoT Bhayani and Huang
5. SentiWordNet (http://sentiwordnet.isti.cnr.it)

یک مطالعه موردی یوتیوب

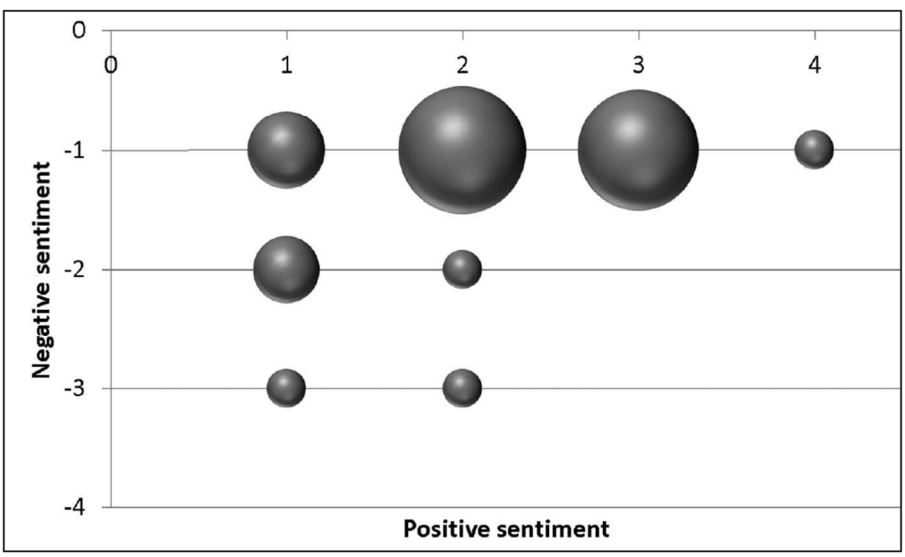
در بررسی وبگاه‌ها کتابخانه‌های دانشگاهی بریتانیا، ۲۱ مورد به عنوان موارد پیوندی به یوتیوب شناسایی شده‌اند. از بین این‌ها، هفت مورد آنها ویدئوهای کتابخانه میزبان به عنوان بخشی از یک حساب کاربری سازمانی گسترده‌تر بودند، و داده ها از ۱۴ حساب کاربری دیگر جمع‌آوری شدند.

درحین ایجاد یک مستند متنی ساده برای فهرست نام کاربران ۱۴ حساب کاربری، وبومتریک آنالیست به سرعت می‌تواند کل داده های وابسته به آن حساب‌های کاربری را جمع‌آوری کند. پیش از هدایت پنجره پاپ‌آپ برای هدایت برنامه به سمت فایل مرتبط با اسامی کاربری، این کار صرفاً نیاز به گزینه «جستجو با نام کاربر» و کلیک روی دکمه «جستجو برای ویدئوهای همخوان با هر جستجو در فایل» دارد. اما قبل از آن پنجره پاپ‌آپ برای هدایت برنامه به سمت فایل مرتبط با اسامی کاربری هدایت می‌شد. سپس یک فایل اضافی حاوی آخرین اطلاعات ۱۰۰۰ ویدئوی مرتبط با هر یک از این حساب‌های کاربری و آمار مرتبط با تمام این حساب‌ها ایجاد می‌شود (مانند، لایک ها، تنفرها، مدت زمان ویدئوها و تعداد اظهارنظرها). در بین ۱۴ حساب کاربری ۴۳۰ ویدئو وجود داشت. یوتیوب امکان سه سطح کنترل کاربر را بر اظهارنظرها می‌دهد: اجازه داده شده[1] (پیش‌فرض)، تعدیل یافته[2] و رد شده[3]. در ۹ ویدئو اظهارنظر مجاز نبود و ۲۴ مورد تعدیل یافتند (نیازمند تأیید بیش از انتشار است). از بین ۴۲۱ مورد برخوردار از اظهارنظر، تنها ۲۱ مورد واقعاً اظهارنظر داشتند و اکثریت آن‌ها تنها یک اظهارنظر داشتند: یک ویدئو شش اظهارنظر داشت، دو ویدئو پنج اظهارنظر داشتند، دو ویدئو چهار اظهارنظر داشتند، یک ویدئو سه اظهارنظر داشت، دو ویدئو دو اظهارنظر داشتند و ۱۳ ویدئو یک اظهارنظر داشتند.

اظهارنظرهای منتشره با ایجاد یک فایل متنی از ویدئویی نتایج جستجوهای قبلی بارگیری شده و سپس «اظهارنظرهای یوتیوب برای فهرست ویدئویی» در وبومتریک

1. allowed 2. modereted 3. denied

آنالیست انتخاب می‌شوند. این امر منتج به ۳۶ اظهارنظر شد، که دو مورد انگلیسی نبودند و یک مورد پیام ناخواسته[1] بوده، که سپس در سنتی‌استرنت با انتخاب «تحلیل قدرت احساسی» و سپس «تحلیل تمام متون در فایل» قابل تحلیل می‌شوند. نسخه معمولی برای هر متن دو امتیاز عاطفی به دنبال دارد، که هم احساس مثبت (۱ تا ۵) هم احساس منفی (۱- تا ۵-) را دربر دارد. نسخه جاوا امکان تحلیل تک واحدی (۴- تا ۴+)، دوگانی[2] (مثبت یا منفی) و سه‌گانی[3] (مثبت، منفی یا خنثی) را نیز فراهم می‌کند. شکل ۵-۵ عواطف ویدئوهای کتابخانه را با اندازه حبابی که نشانگر تعداد ویدئوهای متصل شده به موقعیت است نشان می‌دهد. همانگونه که می‌توان دید بیش‌تر اظهارنظرها منفی نبوده‌اند (۱-) اما آن‌ها را می‌توان نسبتاً مثبت در نظر گرفت (۲ یا ۳).



شکل ۵-۵. نمودار حبابی نشانگر ویدئوهای کتابخانه در یوتیوب

1. spam 2. binary 3. trinary

نتیجه‌گیری

فناوری رسانه‌های اجتماعی به روش‌های متفاوتی مورد استفاده قرار می‌گیرند، و خیلی مهم است که یک فناوری فقط به این خاطر که در یک مورد خاص کار نکرده خارج از رده دانست. اگرچه بحث‌های زیادی برای طیف خدمات و وبگاه‌های رسانه‌های اجتماعی موجود وجود دارد، خدمات و وبگاه‌های انگشت‌شماری برای اکثریت رسانه‌های اجتماعی وجود دارند. در حالی‌که این امر را می‌توان یک منفی تلقی کرد، امتیازهای انحصاری موجب کاهش انتخاب-کاربر می‌شوند و به بزرگ‌ترین شرکت‌های رسانه‌های اجتماعی امکان فعالیت با درجه‌ای از بخشودگی را می‌دهند، هم چنین مهم است که مزایای ارائه شده به پژوهش سنجه وبی را به رسمیت بشناسیم چرا که وبگاه‌های شبکه‌های اجتماعی حجم زیادی از داده‌های ساختار یافته را عرضه می‌کنند. خدماتی مانند توییتر که صدها میلیون کاربر دارد و در طی روز چندین بار وضعیت خود را به روز می‌کنند، و امروزه امکان شناخت حالات و عقاید افراد در زمان واقعی را فراهم می‌کند، عبارت «نبض» جهان متصل را برای عنوان کتاب هوبارد[1] (۲۰۱۱) تأیید می‌کند.

وبگاه‌های شبکه‌های اجتماعی از دیدگاه وب‌سنجی یک وسیله غیر قابل انکار هستند: حجم عظیمی از داده‌های ساختار یافته برای تحلیل در دسترس بوده و مردم بیش از پیش اطلاعات بیش‌تر ابعاد زندگی خود را وارد وب می‌کنند. این موضوع به این معنی نیست که همان خدمات جزو بخش های ارزشمند فعالیت‌های برخط کتابداران هستند و تحلیل وبی برای تعیین مشارکت ابزارها در کار کتابداران اهمیت دارند. ظرفیت و واقعیت همیشه یکی نیستند، بویژه هنگامی که پای ظهور فناوری‌ها در میان باشد؛ درحالی‌که به صورت نظری وبگاه‌های شبکه های اجتماعی باید ابزارهای ارزشمندی برای دانشجوها باشند، یک بررسی نشان داد که زمان صرف شده در وبگاه‌های شبکه‌های اجتماعی برای عملکرد دانشگاهی شاخصی منفی است (پایول، باکر و کوچران[2]، ۲۰۱۲).

1. Hubbard 2. Paul, Baker and Cochran

۶

بررسی ارتباط بین کنشگرها

مقدمه

سنجه‌های وبی ارتباطی، ارتباط بین کنشگرهای برخط را بررسی می‌کنند چه آن‌ها افراد باشند، چه سازمان‌ها، چه حساب‌های کاربری و چه صفحه‌های وبی. ممکن است آنها مستقل از سنجه‌های وب ارزیابانه در نظر گرفته شوند زیرا آن‌ها الزاماً برای این طراحی نشده‌اند که اطلاعات تأثیر محتوا را ارائه کنند، بلکه برای ارائه اطلاعات ساختار شبکه به عنوان یک کل، یا موقعیت کنشگرها در درون یک شبکه‌طراحی شده‌اند. تفاوت بین سنجه‌های ارتباطی و ارزیابانه می‌تواند فقط یک ادراک باشد؛ محوری بودن یک شبکه می‌تواند به عنوان شاخص تأثیر، یا می‌تواند تنها شاخصی برای محوری بودن شبکه باشد. این فصل به برخی از روش‌های تحلیل شبکه های اجتماعی که مورد استفاده انجمن سنجه‌های وبی هستند و نیز نحوه استفاده از این فناوری‌های تجسمی و روش‌شناسی‌هایی که برای کسب اطلاعات می‌توانند مورد استفاده کتابداران باشند می‌پردازد.

روش‌های تحلیل شبکه‌های اجتماعی

روش‌های تحلیل شبکه‌های اجتماعی بر اساس این ایده در علوم اجتماعی توسعه یافت که کنشگرها نباید مجزا در نظر گرفته شوند، بلکه رفتار آن‌ها در زمینه اتصال

کنشگرها به بهترین شکل درک می‌شود. اینک اهمیت به اشتراک‌گذاری اطلاعات در شبکه‌های اجتماعی و تغییر رفتار کاربر به صورت گسترده شناخته شده است (کریستاکیس و فولر[1]، ۲۰۰۹)، و کاربرد روش‌شناسی‌های تحلیل شبکه های اجتماعی علاوه بر اینکه برای بررسی‌های کتاب‌سنجی سنتی در جامعه علم اطلاعات (اوتا و روسیو[2]، ۲۰۰۲) توصیه می‌شود، شالوده تحلیل بررسی‌های وب‌سنجی ابرپیوندهای بین صفحه‌های وبی را نیز تشکیل داده‌اند (مانند پارک، بارنت و نام[3]، ۲۰۰۲؛ هولمبرگ[4]، ۲۰۰۹).

اگرچه روش‌شناسی‌های تحلیل شبکه های اجتماعی می‌تواند به وب و اطلاعات درون آن اعمال شود اما مجموعه کاربردهای آن بسیار گسترده‌تر بوده، آنچنان که کارکرد واژه‌شناسی عمومی استفاده شده یکی از گره‌ها و مرزها[5] است به جای صفحه‌های وبی و ابرپیوندها. یک شبکه (یا نمودار[6]) شامل گره‌ها (گاهی به عنوان رئوس[7] نیز به آن‌ها اشاره می‌شود) و مرزها است،که اتصال بین گره‌ها محسوب می‌شوند. این مرزها ممکن است با توجه به ماهیت اتصال هدایت نشده یا هدایت شده باشند (که در این مورد ممکن است به آن‌ها قوس[8] گفته شود). برای مثال، صفحه‌های وبی از طریق پیوندهای جهت‌دار[9] متصل شده‌اند، و داشتن پیوندهای یک وب‌نوشت به گزارش خبری بی‌بی‌سی[10] به معنای اتصال خودکار گزارش بی‌بی‌سی به وب‌نوشت نیست. در مقایسه، مرزهای دوستی (حداقل در دنیای فیزیکی) اغلب بی جهت[11] در نظر گرفته می‌شوند: چنانچه فرد الف با فرد ب دوست است پس فرد ب نیز با فرد الف دوست است.

نمودارها علاوه بر اینکه می‌توانند جهت‌دار یا بی‌جهت باشند، هم‌چنین می‌توانند دوگانی یا وزن‌دار[12] باشند. برای مثال، در توییتر دنبال کردن یک رابطه دوگانی است یعنی یک حساب کاربری یا حساب کاربری دیگری را دنبال می‌کند یا خیر. هر تعداد

1. Christakis and Fowler
2. Otte and Rousseau
3. Park, Barnett and Nam
4. Holmberg
5. nodes and edges
6. graph
7. verticals
8. arc
9. directional
10. BBC
11. undirected
12. weighted

پیام می‌تواند بین دو حساب کاربری توییتر ارسال شده باشد و یک تحلیل نمودار می‌تواند به ارتباطات بر اساس تعداد پیام‌ها ارسال شده بین دو کاربر وزن بدهد.

اندازه‌گیری‌های تحلیل شبکه‌ای می‌تواند به عنوان یکی از این سه نوع توصیف شود: اندازه‌گیری‌های ویژگی گره‌ها و مرزها، آن‌هایی که همسایگی گره‌ها را توصیف می‌کنند، و آن‌هایی که ساختار کل شبکه های را تحلیل می‌کنند (بورنر، سانیال ووسپیگنانی[1]، ۲۰۰۷). این فصل برای معرفی کامل موضوع تحلیل شبکه‌های اجتماعی طراحی نشده، بلکه به منظور نمایش ظرفیت روش‌های تحلیل شبکه‌ای برای کتابداران تدوین شده است. این بخش بر سه حوزه تمرکز دارد: مرکزیت[2] گره‌ها، تحلیل‌های خوشه‌ای[3] و خواص آماری نمودار. هریک از اینها می‌توانند منشأ شناخت‌های ارزشمندی برای کتابداران باشند و از طریق بسته‌های نرم‌افزاری تحلیل شبکه های اجتماعی بسیار مختلف موجود می‌توانند تجسم و تحلیل شوند. برخی هزینه بر هستند (مانند یوسی‌آی‌نت[4])، برخی به صورت رایگان در دسترس هستند (مانند پاجک[5]) و برخی دیگر متن–باز هستند (مانند گِفی[6]). در طی این فصل از دو قطعه نرم‌افزار استفاده شده است: نودکسل، به علت سادگی بارگیری محتوا از وب و به تصویر کشیدن شبکه‌ها، و گِفی، به علت طیف وسیع وصل شدنی‌هایی که موجب می‌شود برای موقعیت‌های بسیار فراوانی مناسب باشد.

مرکزیت گره

به‌جای بررسی مرکزیت یک گره در درون ساختار کل یک نمودار، بیش‌تر بررسی‌های وب‌سنجی اولیه تأثیر سند وبی مبتنی بر پیوندهای درونی را ارزیابی کرده‌اند که ناشی از مجاورت بی واسطه است. هرچند، طیف متنوع‌تر روش‌های تحلیل در درون تحلیل‌های شبکه اجتماعی گسترش یافته، و توسط دانشمندان علم اطلاعات و کامپیوتر بسط یافته است؛ از آن زمان تحلیل‌ها به شبکه‌های اسناد وبی و حساب‌های کاربری رسانه‌های اجتماعی اعمال شده‌اند.

1. Boner, Sanyal and Vespignani
2. centrality
3. cluster analysis
4. UCINet
5. Pajek
6. Gephi

طیف گسترده‌ای از اندازه‌گیری‌های مرکزیت ابداع شده است، دیرینه‌ترین و بطور سنّتی مهم‌ترین آنها که دارای مرکزیت درجه بوده‌اند، مرکزیت نزدیکی و مرکزیت بینیّت[1] است. اخیراً توجه روزافزونی به مرکزیت بردار ویژه[2]، به‌خصوص پیج رَنک شده است، که ممکن است انواع آن مدّ نظر قرار گیرد. هر یک از اندازه‌گیری‌های مختلف مرکزیت تلاش دارند تا ابعاد مختلف ارتباط بین گره و شبکه را اندازه بگیرند:

- مرکزیت *درجه* یک گره بر اساس تعداد اتصال‌های مستقیمی است که با سایر گره‌ها دارد و معادل شمارش خام پیوندهای درونی و پیوندهای بیرونی در مطالعات اولیه وب‌سنجی است.
- *مرکزیت نزدیکی* یک گره را به عنوان مرکز یک شبکه شناسایی می‌کند که بتواند با تمام گره‌های دیگر به سرعت تعامل داشته باشد.
- *مرکزیت بینیّت* یک گره را به عنوان مرکزیت می‌شناسد که در کوتاه‌ترین مسیر بین جفت گره‌های دیگر قرار داشته باشد.
- *مرکزیت بردار ویژه* و *پیج رَنک* گره‌هایی هستند که مرکزیت بالایی به گره‌های پر اتصال دیگر می‌دهد.

هیچ اندازه‌گیری مرکزیتی نیست که بر دیگری ارجحیت داشته باشد، هرچند که برخی اندازه‌گیری‌های مرکزیت برای موقعیت‌های خاصی مناسب‌تر هستند. پیج رَنک گوگل در مقایسه با مرکزیت درجه به علت تغییرپذیری بالا در کیفیت صفحه‌های وبی و ماهیت دسترسی عمومی شبکه می تواند برای رتبه‌بندی صفحه‌های وبی از الگوریتم مفیدتری برخوردار باشد. هرچند، در مواقعی که در مورد اهمیت گره‌ها عدم توافق کم‌تری وجود دارد، و بدست آوردن اطلاعات بیش‌تر در خصوص کل شبکه های با دشواری زیادی همراه است، مرکزیت درجه می‌تواند مفیدتر از اندازه‌گیری مرکزیت باشد. به همین صورت، درحالی شاید تصور شود که مرکزیت

1. betweenness 2. eigenvector

درجه، پیج رَنک و مرکزیت بردار ویژه می‌توانند اطلاعات تغییرات یک گره را بدهند، چنین اطلاعاتی همواره مهم‌ترین جنبه نیست. مرکزیت نزدیکی شاید در موقعیت‌هایی که مردم علاقه مندند اطلاعات را در سریع‌ترین شکل ممکن دریافت کرده یا به اشتراک بگذارند بسیار مهم به نظر بیاید، اما مرکزیت بینیّت می تواند شاخص بهتری برای اهمیت یک گره در یک شبکه باشد.

شناسایی خوشه

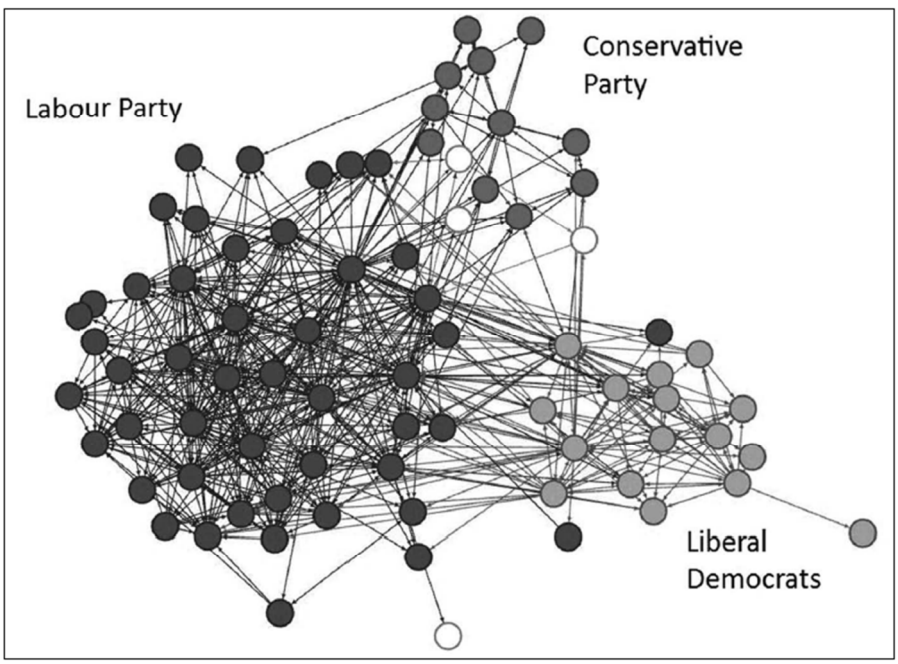
دنیای برخط در کل همگن‌تر از خود جامعه نیست: وبگاه‌های دارای منافع مشترک می توانند با یکدیگر پیوند برقرار کنند؛ کاربران شبکه های اجتماعی دارای منافع مشترک می توانند یکدیگر را دنبال کنند؛ و اشخاصی که در دنیای واقعی یکدیگر را می‌شناسند می توانند در دنیای مجازی با یکدیگر دوستی برقرار کنند. این اولویت‌ها باعث ایجاد خوشه ها یا جوامع شده و شناسایی این خوشه‌ها استفاده عملی زیادی دارد، هم به عنوان موضوع تحلیل و هم برای اشاعه اطلاعات.

خوشه‌بندی در تحلیل شبکه های اجتماعی یک شبکه را به شبکه‌های کوچک‌تر، یا چندین خوشه، را بر اساس ساختار پیوند شبکه تقسیم می‌کند، خوشه‌ها اغلب با مرزهای بیش تر در درون یک خوشه مشخص می‌شوند تا بیرون آن. نرم‌افزار تحلیل شبکه‌های ممکن است شامل الگوریتم‌های خوشه‌بندی باشند؛ برای نمونه، نرم‌افزار متن‌باز گِفی شامل هر دو الگوریتم خوشه‌ای مارکوف[1] و روش مادوله لووین[2] است.

اعمال یک الگوریتم خوشه‌بندی به یک شبکه اطلاعات بیش تری راجع به شناسایی خوشه‌های ناشناخته یا حتی خوشه‌بندی‌هایی که تاکنون شناخته شده اند می‌دهد. برای مثال، شکل ۶-۱ شبکه های ۷۹ نماینده مجلس بریتانیا که در سال ۲۰۰۹ در توییتر بودند را نشان می‌دهد. این شکل با توجه به الگوریتم فورس اطلس ۲[3] موجود در گِفی پخش شده است، که در آن گره‌ها یکدیگر را دفع می‌کنند و مرزها تلاش می‌کنند تا گره‌ها جذب یکدیگر شوند. هر یک از سه حزب سیاسی

1. Markov Cluster Algorithm 2. Louvain Modularity 3.. Force Atlas 2

اصلی بریتانیا با هاشور متفاوتی نشان داده می‌شود (گره‌های سفید وابسته به سایر احزاب هستند)، و می‌توان دید که گره‌های احزاب سیاسی می‌توانند با یکدیگر هم خوشه شوند.



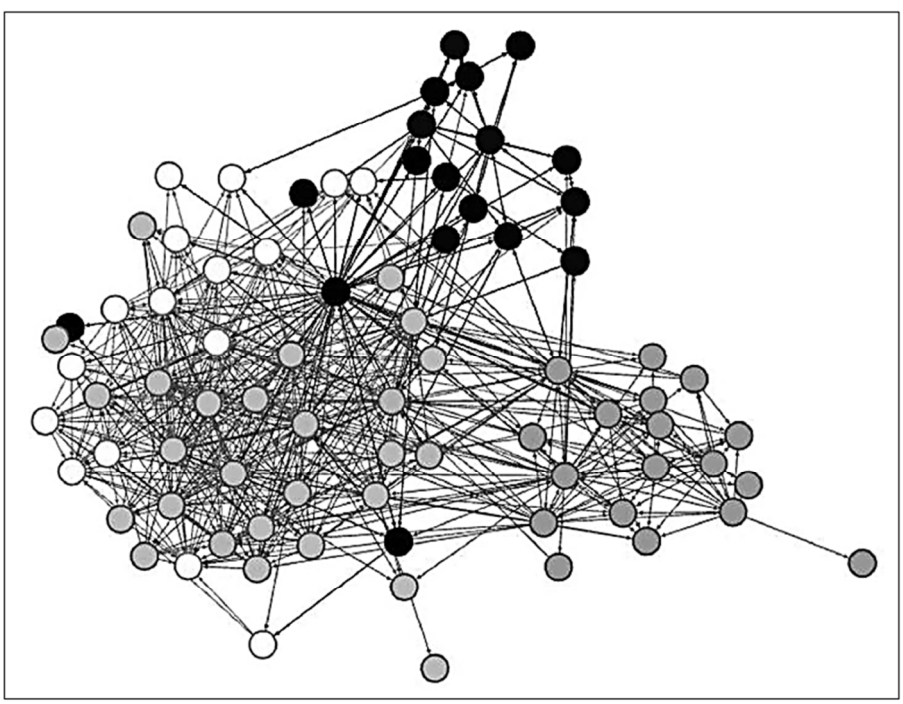
شکل ۶-۱. نمایندگان مجلس بریتانیا بر اساس وابستگی حزبی هاشور خورده

با اعمال روش مادوله لووین در شبکه، خوشه‌هایی شناسایی می‌شوند که در درون خوشه‌ها مرزهای بیش‌تری دارند تا بین آنها، نمایندگان مجلس در کل بر اساس حزب سیاسی نیز خوشه‌بندی می‌شوند (شکل ۶-۲ را ملاحظه کنید). با این وجود، درحالی‌که لیبرال دموکرات‌ها[1] و محافظه‌کارها[2] به عنوان گروه‌های مجزا شناخته می‌شوند، نمایندگان مجلس حزب کارگر در حقیقت دو گروه را شکل می‌دهند. برای آن‌هایی که به اتحاد حزب کارگر علاقه مندند این خوشه‌بندی ممکن است موجب راه‌اندازی بررسی‌های بیش‌تری شود: آیا خوشه‌بندی تفاوت‌های ایدئولوژیک را

1. Liberal Democrats 2. Conservatives

منعکس می‌کند؟ آیا نمایی از یک تفاوت جغرافیایی است؟ یا اینکه فقط یک ناهنجاری است که اکثریت نمایندگان مجلس کارگر که در آن زمان از توییتر استفاده می‌کنند آن را پدید آورده‌اند؟

مطالعات مشابه می‌توانند برای تفاوت بین ساختارهای سازمانی رسمی و ارتباطات بین کارکنان یا تفاوت بین رشته‌های علمی بیان شده در مجلات و ارتباطات نشان داده شده در شبکه‌های دانشمندان اطلاعات جدیدی بدهند.



شکل ۶-۲. نمایندگان مجلس بریتانیا بر اساس مدل مدوله لووین

خواص آماری نمودار

علاوه بر موقعیت گره‌ها در درون یک نمودار، خود نمودارها نیز می‌توانند تحلیل شوند: تفاوت ساختاری بین شبکه‌ها ممکن است برای مجموعه‌های متفاوت

کنشگرهایی مدّ نظر قرار بگیرند که از فناوری‌های مشابهی استفاده می‌کنند، یا کنشگرهای مشابهی که از فناوری‌های مختلفی استفاده می‌کنند. سنجه‌های زیادی هستند که اطلاعات ساختار کلی یک نمودار را ارائه می‌کنند، اما درحالی‌که شاید آن‌ها ویژگی‌های یکسانی را اندازه بگیرند تفاوت‌های ظریف احتمالاً به معنای این است که یک سنجه مناسب‌تر از سنجه دیگری است.

مرکزیت میانگین

همانطور که در بالا بحث شد، برای مرکزیت اندازه‌گیری‌های متفاوتی (مانند، مرکزیت درجه، مرکزیت بینیّت و مرکزیت نزدیکی) وجود دارد، و ممکن است از مرکزیت میانگین برای عرضه شاخصی در خصوص اینکه آیا گره‌های یک شبکه بیش از گره‌های شبکه‌ای دیگر به هم متصل هستند یا خیر استفاده شود، یا اینکه به طور متوسط گره‌های یک شبکه های نسبت به گره‌های سایر شبکه‌ها سریع‌تر می‌توانند با سایر گره‌های درون شبکه‌های تعامل برقرار کنند یا خیر.

تراکم نمودار

با تقسیم تعداد واقعی مرزهای درون یک نمودار بر تعداد بیشینه مرزهای احتمالی یک نمودار تراکم یک نمودار محاسبه می‌شود. برای مثال، در یک نمودار جهت دار با ده گره، تعداد بیشینه مرزها ۹۰ است زیرا هر گره می‌تواند به هر گره دیگری متصل شده، و در نتیجه اگر نمودار ۴۵ مرز داشته باشد آن‌گاه تراکم نمودار ۰/۵ خواهد بود. هرچند تراکم نمودار به مرکزیت درجه مرتبط است، اما با توجه به اندازه نمودار یکی می‌تواند مناسب‌تر از دیگری باشد. برای مثال، پایین بودن تراکم نمودار در یک نمودار بسیار بزرگ (مانند شبکه جهانی وب[1]) اجتناب‌ناپذیر است و بررسی مرکزیت درجه میانگین معنادارتر است، درحالی‌که شبکه‌های بسیار کوچک با دادن اطلاعات تراکم شبکه‌ها می‌توانند بینش بیش‌تری ایجاد کنند.

1. world wide web

ضریب خوشه‌بندی

خوشه‌بندی گره‌ها در نمودارهای دنیای واقعی جهان کوچکی را خلق می‌کنند که در آن شبکه‌های بزرگ گره‌ها به طور میانگین تنها چند گام بین هر دو گره فاصله دارند. ضریب خوشه‌بندی شبکه بطور کلی به دو صورت ظهور می‌کند (هاردیمن و کاتزیر[1]، ۲۰۱۳): ضریب خوشه‌بندی میانگین و ضریب خوشه‌بندی جهانی. برای هر گره یک ضریب خوشه‌بندی محلی می‌تواند محاسبه شود، که نسبت مجاورت گره ای که می‌توانسته متصل شوند و هم اکنون متصل شده‌اند را نشان می‌دهند، و ضریب خوشه‌بندی میانگین میانگینی است برای تمام گره‌های درون یک شبکه. ضریب خوشه‌بندی جهانی تعداد بسته‌های سه‌تایی باز را در یک نمودار را (جایی که سه گره با دو مرز متصل شده‌اند) با تعداد بسته‌های سه‌تایی بسته را در یک نمودار مقایسه می‌کند (جایی که سه گره با سه مرز متصل شده‌اند).

ماژول

ماژول قدرت خوشه‌ها را در یک شبکه های اندازه می‌گیرد. یک شبکه با ماژول بالا دارای لبه‌های زیادی در خوشه‌هایش است نه در بین آن‌ها. الگوریتم‌های خوشه‌بندی خاص، همانند روش ماژول لووین که در نمونه توییتر بالا استفاده شد، از ماژول برای بهینه‌سازی فرایند خوشه‌بندی استفاده می‌کنند.

منابع تحلیل شبکه‌های ارتباطی

وب پر از شبکه‌هایی است که اساس تحلیل‌های شبکه های ارتباطی را شکل می‌دهند، که شامل رابطه‌هایی است که عمداً و صراحتاً بیان شده و رابطه‌هایی که از داده‌ها بروز می‌کنند. در کتاب‌سنجی‌های قدیمی تعداد محدودی رابطه مستقیم بیان شده، که شبکه‌های استناد و هم- نویسندگی[2] از نمونه آنهاست، بررسی‌های بسیار زیادی بر اساس روابط غیرمستقیم مانند اتصال کتاب‌شناسی[3] (دو مقاله پیوند شده چنانچه آن‌ها به یک یا چند مقاله استناد کرده باشند) و تحلیل هم‌استنادی (دو مقاله پیوند شده

1. Hardiman and Katzir 2. co-authorship 3. bibliographic coupling

چنانچه یک یا تعداد مقاله بیش‌تری به آنها استناد کرده باشند) انجام شده است. بررسی‌های وب‌سنجی اولیه روش های تحلیل استناد مشابهی را به صفحه‌های وبی و ابرپیوندهای بین آن‌ها اعمال می‌کردند: تحلیل هم‌پیوندی مشابه تحلیل هم‌استنادی است (لارسون[1]، ۱۹۹۶) و تحلیل هم‌پیوندی مشابه اتصال کتابشناختی است (تلوال و ویلکینسون[2]، ۲۰۰۴). داده های برخط موجود بسیار غنی‌تر از داده‌هایی است که به صورت قدیمی در پایگاه داده‌ای کتابشناختی موجود بود، و وبگاه‌های شبکه‌های اجتماعی درحالی ظهور کرده‌اند که علاوه بر تنوع روزافزون روابط غیر مستقیم امکان بررسی طیف بسیار گسترده‌تری از رابطه‌های مستقیم را میسّر می‌ساختند.

پژوهش‌های مربوط به سنجه وبی ارتباطی برای بررسی رابطه بین سازمان‌ها از پیوندهای بین صفحه‌های وبی انواع سازمان های متفاوتی استفاده کرده‌اند (استوارت و تلوال، ۲۰۰۶؛ مینگویللو[3] و تلوال، ۲۰۱۲). برای بررسی تفاوت‌های زبانی و فرهنگی بین بخش های انگلیسی زبان کانادا و فرانسوی زبان کانادا از تحلیل هم‌پیوندی وبگاه‌ها دانشگاه کنیدین[4] استفاده شد (وگان، ۲۰۰۶)، درحالی‌که بار-ایلان و آزولی[5] (۲۰۱۳) ساختار و پیوند بین وبگاه های غیر انتفاعی را در فلسطین اشغالی بررسی کردند. پژوهش‌های ارتباطی وبگاه‌های شبکه های اجتماعی اغلب بر توییتر متمرکز بوده‌اند چرا که در توییتر شبکه ها در بازترین حالت بوده، و سیاست پرطرفدارترین حوزه برای بررسی محسوب می‌شود. برای نمونه، هاسو و پارک (b۲۰۱۲) شبکه‌های سیاست‌مداران کره‌ای را در توییتر، و همچنین در وب‌نوشت‌ها و صفحه‌های خانگی آن‌ها را با هم مقایسه کردند، و برونس و هایفیلد[6] (۲۰۱۳) شبکه‌های سیاسی را در توییتر در طی انتخابات ایالت کوئینزلند[7] بررسی کردند. مهم‌تر اینکه، سنجه‌های وبی ارتباطی نه تنها امکان بررسی رابطه‌های موجود را میسّر می‌کند، بلکه برای شناسایی رابطه‌های جدید نیز می‌تواند مورد استفاده قرار گیرند. برای مثال، کازییوکی و همکارانش[8] (۲۰۱۳) وبگاه‌های شبکه های اجتماعی را برای شناسایی مشتری‌های کسب و کار جدید مطرح می‌کنند.

1. Larson
2. Wilkinson
3. Minguillo
4. Canadian university
5. Azoulay
6. Bruns and Highfield
7. Queensland
8. Kazienko et al.

اکنون برخی از ابزارها و روش‌های مطرح با داده های دو منبع بررسی می‌شوند: وب و وبگاه شبکه‌های اجتماعی توییتر.

تحلیل شبکه‌های وب-دولتی محلی در میدلندز[1]

وبگاه‌های شبکه های اجتماعی در مقایسه با صفحه‌های وبی قدیمی توجه بسیار بیش تری را به خود جلب کرده‌اند چرا که در سال‌های اخیر صدها میلیون کاربر برای عضویت در وبگاه‌های شبکه های اجتماعی سرازیر شده‌اند. هرچند، اساساً صفحه‌های وبی به نسبت وبگاه‌های شبکه‌های اجتماعی نوع متفاوتی از ارتباط و مجموعه بسیار متنوع‌تری از محتوا را نشان می‌دهند.

همانگونه که در فصل سوم بحث شد گردآوری اطلاعات فعالیت‌های پیوندزنی بین وبگاه ها سخت‌تر شده است چرا که موتوهای جستجو قابلیت‌های خاص پیوند را تنزل داده و دسترسی به رابط‌های برنامه‌نویسی را کاهش داده‌اند. با این وجود، هم‌چنان بررسی استنادهای نشانی وبی را از طریق رابط‌های بینگ می‌توان انجام داد، و در این مثال استفاده شده از آنجا که داده‌ها برای نمایش پیوندزنی درونی بین سازمان‌های دولتی محلی در میدزلند بریتانیا جمع‌آوری شدند. شاید که کتابداری بخواهد این بررسی را انجام داده تا بتواند از بازیکن‌های حوزه خاصی، یا حتی توابع شرکت مشابهی اطلاعات بدست آورد.

لازمه پژوهش‌های ارتباطی استناد محور نشانی وبی ارسال به تعداد زیادی جستجو به موتور جستجو است. در یک مطالعه تأثیر ورود دستی عبارات جستجو به یک موتور جستجو شدنی است. برای مثال، ارزیابی تأثیر از ۱۰۰ وبگاه مختلف می‌تواند بر اساس ۱۰۰ عبارت جستجو باشد. هرچند، اگر استناد نشانی وبی یا اشاره‌های وبی برای برخی از انواع تحلیل شبکه‌های استفاده شوند، در همه چیز غیر از کوچک‌ترین مطالعات احتمالاً جمع‌آوری خودکار داده‌ها از یک موتور جستجو ضروری است. بررسی استنادهای نشانی وبی بین ۱۰۰ نشانی وبی مختلف مستلزم وارد کردن ۹۹۰۰ جستجو

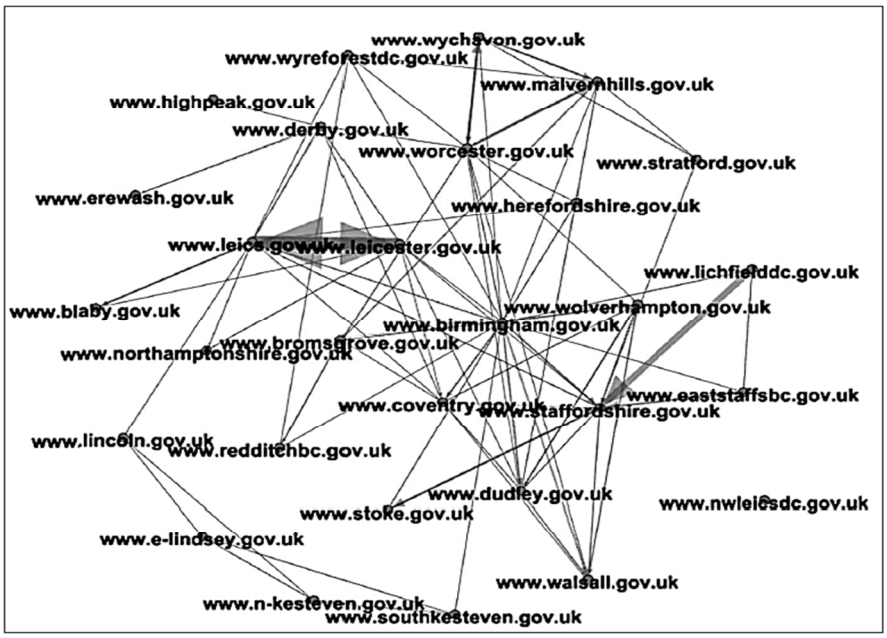
1. Midlands

است، چرا که برای هر ۱۰۰ نشانی وبی باید یک جستجو ایجاد شود تا تعیین کند آیا در درون هر یک از ۹۹ وبگاه دیگر مورد استناد بوده است یا خیر.

در حال حاضر رابط برنامه کاربردی بینگ تنها امکان ۵۰۰۰ جستجو رایگان را در ماه می‌دهد، و امکان دسترسی بیش‌تر برای آن‌هایی که اشتراک دارند هست. هرچند کتابداران ممکن است بخواهند برای استفاده از این داده‌ها برنامه خود را بنویسند، در حال حاضر ساده‌ترین راه جمع‌آوری داده‌ها از طریق وبومتریک آنالیست[1] است. ابزاری که برای تحلیل وب‌سنجی از گروه پژوهشی سایبرسنجی آماری در دانشگاه ولورهامپتون که برای جمع‌آوری داده ها از طیف وسیعی از منابع شامل بینگ طراحی شده است. از آنجایی که به صورت خاص برای جامعه وب‌سنجی طراحی شده است. علاوه بر اینکه می‌تواند جستجوها را به طور خودکار برای بینگ بفرستد، به طور خودکار نیز می‌تواند جستجوهای لازم را برای بسیاری از پرطرفدارترین مطالعات ایجاد کند.

شکل ۶-۳ یک نمودار شبکه‌های استناد نشانی وبی را برای پیوند داخلی بین مؤسسه‌های دولتی محلی در دو منطقه بریتانیا نشان می‌دهد: میدلندز شرقی و میدزلند غربی. برای خلق چنین نمودار شبکه‌ای در ابتدا به فهرست آدرس های اینترنتی که باید بررسی شوند نیاز است. انتخاب چنین صفحه‌هایی در بررسی‌های وب‌سنجی به حاشیه رفته است، هرچند اگر قرار است مطالعه معنی‌داری انجام شود باید با رویکرد روش‌شناختی به آن نزدیک شد. در این مورد، پیوند درونی مؤسسه‌های دولتی محلی شامل مجموعه شناخته شده‌ای از سازمان‌ها است، که هر یک نشانی وبی مجزایی دارند. برای مطالعات مشابه دیگر، مانند بررسی پیوند درونی کسب و کارهای محلی در یک منطقه بعید است که فهرست معتبری وجود داشته باشد، و برای نحوه ایجاد چنین فهرستی باید تصمیم‌گیری کرد. در این مورد برای تهیه فهرست معتبر مؤسسه‌ها و منابع نشانی وبی از اوپنلی لوکال[2]، یک وبگاه برای ایجاد دسترسی بیش‌تر به داده‌های دولتی محلی، استفاده شد.

1. (http://lexiURL.wlv.ac.uk) 2. Openly Local(Http://openlylocal.com)



شکل ۶-۳. نمودار شبکه‌های استناد نشانی وبی از رابطه‌های بین مؤسسه‌های دولت محلی در میدزلند بریتانیا

از آنجایی که اطلاعات دولت محلی به صورت XML قابل دستیابی است، نشانی وبی‌ها را به صورت خودکار می‌توان با دستور importXML()= استخراج کرد. هنگامی که فهرست نشانی وبی‌ها در سندی متنی ذخیره و کلید دولوپر[1] برای جستجوی بینگ درخواست شد، و در عرض چند دقیقه جستجوها می‌توانند به وجود آمده و داده‌ها با وبومتریک آنالیست به صورت خودکار جمع‌آوری شود. اگرچه وبومتریک آنالیست امکان نمایش اولیه نمودارهای شبکه‌های را می‌دهد، فرمتی که نتایج را در آن ذخیره می‌کند با نرم افزارهای بصری‌سازی نموداری دیگر مانند گِفی و پاجک، که کاربردهای بیش تری دارند نیز همخوانی دارد. در این مورد نمودار با گِفی نمایش داده شده است.

هنگامی که داده‌ها در گِفی موجود باشد طیف وسیعی از آمارها با توجه به مرکزیت، خوشه‌بندی و مادوله می‌توانند به سرعت ایجاد شوند.لازم است کتابداران

1. developer key

بدانند که یک سنجه را فقط برای اینکه قابل محاسبه است محاسبه نکنند. دلیل این امر می‌تواند این باشد که چیزی که ضروری است دیداری‌سازی شود تا به بررسی‌های آتی رابطه بین گروه‌های نمایشی کمکی شده باشد.

تحلیل شبکه توییتر – کتابخانه‌های دانشگاهی بریتانیا در توییتر

مثال پیشین مستلزم استفاده از چندین ابزار است: از ایجاد فهرستی از وبگاه‌ها برای بررسی، از طریق جمع‌آوری داده گرفته تا دیداری‌سازی نتایج. وبگاه‌های شبکه‌های اجتماعی نه تنها موجب افزایش میزان داده های ساختاریافته موجود در وب شده‌اند، بلکه تعداد اشخاصی را هم که به نمودار اجتماعی خود علاقه مندند را نیز افزایش داده‌اند. این موضوع منجر به ابزاری مانند نودکسل شده است، که وسیله واحدی را برای بررسی پشت سرهم [1] برخی از پرطرفدارترین وبگاه‌های شبکه‌های اجتماعی عرضه می‌کند.

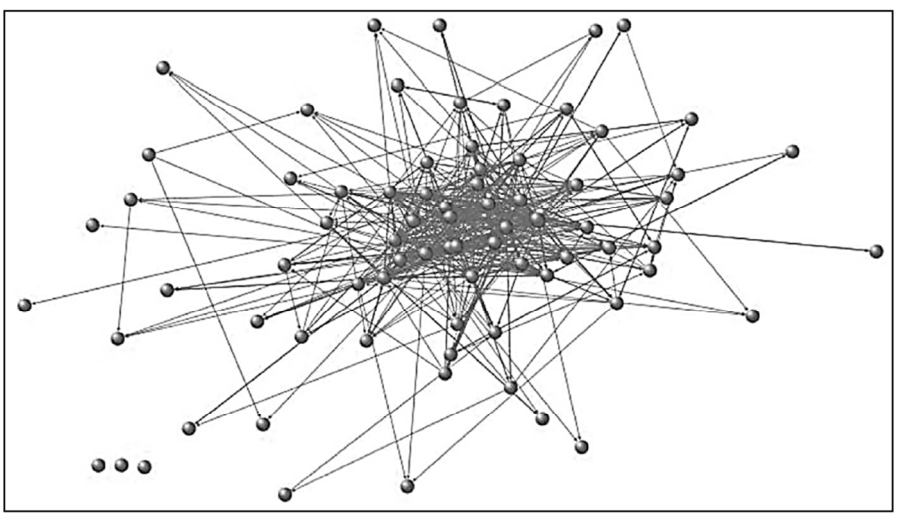
این مثال شبکه های متشکل از ۷۷ حساب کاربری توییتر در بررسی کتابخانه‌های دانشگاهی بریتانیا را مدّ نظر قرار می‌دهد. در این مورد شبکه ها بر اساس یک بررسی از وبگاه ها کتابخانه است، اما نودکسل قابلیت وارد کردن شبکه‌های کاربران توییتر بر اساس شبکه های لایک‌های کاربر، اشاره مردم به یک عبارت خاص و همچنین آن‌هایی که در فهرست توییتر هستند را نیز فراهم می‌کند. از آنجایی که کاربران توییتر با طیف گسترده‌ای از تجربه و تخصص در فهرست‌های توییتری همه موضوعات گردآوری شده‌اند، مجموعه وسیعی از نمونه‌ها آماده تحلیل هستند. تعدادی از آن نمونه ها برای کسانی که می‌خواهند با نودکسل کسب تجربه کنند و ببینند که حساب‌های کاربری چگونه پیوند و خوشه‌بندی می‌شوند عبارتند از:

- ۱۲۷۰ کتابدار بریتانیا (https://tewitter.com/philbradley/uk-libraries)
- ۴۵۸ عضو مجلس بریتانیا (https://twitter.com/tweetminster/ukmps)
- ۴۱۵ ورزشکار المپیک تیم بریتانیا

(https://twitter.com/ACMLDN2012/team-gb-olympics-athletes).

1. end-to-end

شکل ۶-۴ پیوند درونی ۷۷ حساب کاربری کتابخانه دانشگاهی را با استفاده از نودکسل نشان می‌دهد. پنج حساب کاربری دارای بیش‌ترین تعداد دنبال‌کننده در کل توییتر: کتابخانه بودلین[1]، کتابخانه دانشگاه ولز جنوبی[2]، کتابخانه دانشگاه کمبریج، کتابخانه دانشگاه منچستر و کتابخانه دانشگاه ساسکس هستند. هرچند، وقتی تنها به شبکه های کتابخانه‌های دانشگاهی بریتانیا نگاه می‌کنیم مجموعه متفاوتی از وبگاه ها برتر پدیدار می‌شوند (جدول ۶-۱ را ببینید).



شکل ۶-۴. نمودار شبکه روابط بین حساب‌های کاربری توییتر کتابخانه‌های دانشگاهی بریتانیا با استفاده از نودکسل

جدول ۶-۱. کتابخانه‌های دانشگاهی بریتانیا با بالاترین درجه و مرکزیت بینیّت

رتبه	مرکزیت درجه		مرکزیت بینیّت	
۱	متروپولیتن لیدز[3]	۲۹	دانشگاه لینکولن	۱۳۳۷
۲	دانشگاه منچستر	۲۷	دانشگاه شهر	۵۴۵
۳	دانشگاه کمبریج	۲۳	متروپولیتن لیدز	۴۵۳
۴	دانشگاه شهر[4]	۲۳	دانشگاه ساسکس	۴۰۱
۵	دانشگاه ساسکس دانشگاه دورهام	۲۲ ۲۲	دانشگاه لیورپول	۲۴۰

1. Bodleian Library 2. University of South Wales 3. Leeds Metropolitan 4. City University

کتابخانه بودلین و کتابخانه دانشگاه ولز جنوبی برای مرکزیت درجه و یا برای مرکزیت بینیّت در پنج رده اول درون شبکه های ۷۷ کتابخانه نیستند، و فقط کتابخانه دانشگاه ساسکس برای مرکزیت درجه و مرکزیت بینیّت در پنج رده اول قرار گرفته‌اند. گرچه کتابخانه دانشگاه لینکولن[1] فقط توسط ده کتابخانه دیگر در شبکه های دنبال می‌شود، با این وجود به دلیل برخورداری از تعداد دنبال‌های کتابخانه‌های زیادی دارای مرکزیت بینیّت فوق‌العاده‌ای است، از بین ۷۷ کتابخانه ممکن ۴۵ کتابخانه را دنبال می‌کند.

محاسبه مرکزیت درجه و مرکزیت بینیّت گره‌ها موجب می‌شود که از روابط کتابخانه‌های دانشگاهی بریتانیا در توییتر اطلاعات متفاوتی بدست آید. ممکن است این بحث پیش بیاید که مرکزیت درجه در بین سازمان‌های مشابه کیفیت جریان توییتر را بهتر نشان می‌دهد تا کل دنبال‌کننده‌ها، چرا که وابستگی کم‌تر به اعتبار و تعداد دانشجویان دارد، درحالی‌که مرکزیت بینیّت حساب‌های کاربری را که در جامعه مشارکت دارند بهتر نشان می‌دهد. مرکزیت درجه متوسط برای نمودار ۶-۴ برابر ۸/۳۱ است، آنچنانکه ممکن است برای چنین مجموعه سازمان‌هایی که از لحاظ جغرافیایی پراکنده هستند نسبتاً بالا باشد. مجموعه هم‌اندازه نماینده‌های مجلس دارای درجه متوسط ۹/۹۵ است، که با توجه به اینکه بیش‌تر نماینده‌های مجلس عضو یک حزب بودند و آنها را می‌توان مرتباً در همان ساختمان دید، خیلی بیش‌تر نیست.

نتیجه‌گیری

این فصل ابزارها و روش‌شناسی‌های مختلفی را برای بررسی سنجه‌های وب ارتباطی معرفی کرد، که اگرچه با روش‌های متفاوت تنها به سطح سنجه‌های وبی ارتباطی توجه نشان داد، کانون توجه این فصل بجای برخی از ارتباط‌های غیر مستقیم‌تری که کانون توجه سنّتی کتاب‌سنجی به شکل جفت کردن کتابشناختی و هم استنادی، و تحلیل هم استنادی، و تحلیل هم پیوندی بوده، به رویکرد تحلیل شبکه های اجتماعی

1. University of Lincoln Library

و پیوندهای مستقیم بین کنشگرها است. با این وجود، باید به کتابداران بـرای مـواقعی که از سنجه‌های وبی ارتباطی در فعالیت‌های حرفه‌ای خود می‌تواننـد اسـتفاده کننـد و برخی ابزارها و منابع داده‌های در دسترس ایده‌هایی ارائه کرد.

بسیاری از سنجه‌های مورد بحث این فصل می‌توانند بـرای اهـداف ارزیابانـه نیـز مورد استفاده قرار گیرند: شاید کسی بخواهد از مرکزیـت بینیّـت بـه عنـوان شاخصـی برای مشارکت درون یک شبکه های اجتماعی، و یا از تـراکم شـبکه‌هـای مختلـف بـه عنوان شاخصی برای ارزش شبکه‌هـا اسـتفاده کننـد. اگرچـه همانگونـه کـه در مـورد حساب‌های کاربری توییتر کتابخانه‌های دانشگاهی بریتانیا دیده شد، تاکید سـنجه‌هـای مختلف قطعاً بر ابعاد مختلف شبکه اسـت و الزامـاً یکـی بهتـر از دیگـری نیسـت. در عصر برتری روزافزون سنجه‌های ارزیابانه، در خصوص ماهیت توصـیفی سـنجه‌هـای ارتباطی مسائل زیادی است که باید آموخت.

۷

بررسی انتشارات قدیمی در محیطی جدید

مقدمه

همانگونه که در فصل دوم توضیح داده شد، در این کتاب از عبارت «کتاب‌سنجی وبی» برای اشاره به فصول مشترک کتاب‌سنجی و سنجه‌های وبی استفاده شده، و بررسی کمّی انتشارات قدیمی (مانند مجلات و کتاب‌ها) از طریق وب را نیز در بر می‌گیرد.

پژوهش‌های کتاب‌سنجی قدیمی به بررسی کتابخانه جامعی که دارای تمامی منابع کتابشناختی است نپرداخته‌اند، بلکه از طریق یک منبع کتابشناختی محدود، و بیش‌تر با استفاده از نمایه استنادی وب آو ساینس انجام شده‌اند. چنین منبع کتابشناختی تنها به بخش کوچکی محدود است که در درون یک کتابخانه جهانی قرار می‌گیرد، و اطلاعات ذخیره شده این موارد نیز محدود است.

وب دسترسی به منابع کتابشناختی قدیمی را امکان‌پذیرتر کرده و ظرفیت بررسی‌های کتاب‌سنجی را به سه طریق گسترش داده است. این روش‌ها:

- امکان دسترسی به بخش بزرگ تری از مواردی که می‌توانند در کتابخانه جهانی باشند را فراهم می‌کند.
- به‌جای دسترسی تنها به فراداده[1] می‌تواند دسترسی به متن کامل را فراهم کند.

1. metadata

- تحلیل محتوای کتابشناختی را در زمینه نحوه استفاده از آن امکان‌پذیر می‌سازد.

چنین گسترشی از کتاب‌سنجی سنتی به صورت بالقوه می‌تواند منجر به ارائه خدمات بهتر توسط کتابداران شود، با ارائه نقشه‌هایی با جزئیات بیش‌تر علوم یا توسعه اندازه‌گیری دگرسنجه‌ها تأثیر انتشارات‌های قدیمی برخط را تسهیل کنند.

این فصل هر یک از این سه حوزه کتاب‌سنجی وبی را بررسی کرده و با نگاه دقیق‌تری به تحقیق‌های انجام شده و منابع موجود می‌پردازد.

موضوعات کتابشناختی بیش‌تر

در کتاب حاضر استفاده از عبارت «کتاب‌سنجی» به روح بنیادی یونانی آن، یعنی «کتاب[1]» برمی‌گردد، و به تحلیل کمّی کتاب‌ها، مجلات و انواع مشابه انتشاراتی مربوط می‌شود که به صورت سنتی پایه و اساس انتشارات کاغذی را شکل می‌دهند. با حفظ معنی محدودتر آن، ممکن است از تعریف اینگونه به نظر برسد که برای بررسی هیچ چیز جدیدی وجود ندارد، اما حجم عظیمی از نشریات هرگز در پایگاه اطلاعات کتابشناختی کتابشناسی در نظر گرفته نشدند تا پایه و اساس بسیاری از بررسی‌های پیشین باشند.

ایجاد پایگاه‌های اطلاعات کتاشناختی مانند نمایه استنادی علوم دارای منابع بسیار کاملی بوده و از این رو در راستای محدود کردن تعداد آثار مورد توجه خدمات چکیده‌نویسی و نمایه‌سازی است. قانون برادفورد[2] بیان می‌دارد که مقالات مربوط به یک حوزه خاص به صورت گسترده‌ای در بین تعداد زیادی از مجلات توزیع خواهند شد، هرچه از مجلات هسته دورتر شویم با افت فزاینده‌ای مواجه می‌شویم(دبلیس[3]، ۲۰۰۹)، درحالی‌که قانون تجمع بیش‌تر گارفیلد مطرح می‌کند که همپوشانی بین رشته‌ها به این معناست که متون اصلی کل علم در کم‌تر از ۱۰۰۰ مجله پراکنده

1. biblio 2. Bradford's law 3. De Bellis

می‌شوند (گارفیلد، ۱۹۸۳). از این رو می‌توان بحث کرد که متون علمی اصلی می‌توانند تعداد محدودی از مجلات یک پایگاه اطلاعات کتابشناختی را در بر بگیرند.

با این وجود، چنین رویکردی با محدودیت‌های جدّی مواجه است. مقالات مجلات به هیچ وجه تنها نوع انتشاراتی نیستند، و حتی در حوزه‌های زیادی از نوع انتشارات هسته نیز محسوب نمی‌شوند. برای مثال، در علوم انسانی و هنر مهم‌ترین نوع انتشارات جزوه و مقالات هستند، درحالی‌که در علوم کامپیوتر مجموعه مقالات کنفرانس‌ها نقش چشمگیری دارند. حتی اگر انتشار مجلات ابزارها اصلی تلقی شوند، انتشار مقالات الزاماً در بین مجلات علمی «هسته» نیست؛ در عوض ممکن است حوزه‌های جدید و نو ظهور در نشریات حاشیه‌ای و کمتر شناخته شده پدیدار شوند.

این موضوع یک دیدگاه بسیار علمی نشر نیز به شمار می‌رود. برای سازمان‌هایی از دیگر بخش‌های جامعه، از قبیل سازمان‌های صنعتی یا بخش عمومی، متون خاکستری[1] می‌توانند نقش بزرگ‌تری ایفا کنند. متون خاکستری شامل اسنادی است که کم‌تر به صورت رسمی منتشر شده‌اند، شامل طیف وسیعی از گزارش‌های بخش‌های دولتی، مخازن فکری[2] و سایر سازمان‌ها است.

از آنجایی که وب تبدیل به بستر انتشاراتی بسیار مهمی برای تمامی انواع سازمان‌ها شده است، رهگیری برخی از آثاری که زمانی مشکل بود در حال حاضر از طریق نسل جدید موتورهای جستجو کتاب‌شناسی[3] تخصصی به سادگی پیدا می‌شوند.

گوگل اسکالر

احتمالاً گوگل اسکالر معروف‌ترین و پر استفاده‌ترین ابزار جدید کتابشناختی، و یکی از بهترین نمونه‌های نظام زیستی کتابشناختی جدید است، که جعبه جستجوی حداقلی را ارائه می‌کرد، که از ابتدا موتور جستجوی خود را برای ارائه راهی ساده در جستجوی طیف وسیعی از متون علمی از سراسر وب متمایز نمود.گوگل اسکالر

1. grey literature 2. think tanks 3. bibliographic search engines

دسترسی به طیف وسیعی از مقالات مجلات، کتاب‌ها، چکیده‌ها و پایان‌نامه‌ها و متون خاکستری را فراهم می‌کند. چندین نسخه متفاوت از همان مقاله را در کنار یکدیگر می‌توان یافت. مقالات مجلات هم ترازخوانی شده را ممکن است در کنار خروجی‌های مخازن فکری یا پایان‌نامه دانشجویان کارشناسی ارشد دیده شوند.

گوگل اسکالر از زمان ظهورش به محیط پیچیده‌ای تبدیل شده است. نه تنها امکان جستجوی اسناد علمی و اطلاعات استناد این اسناد را می‌دهد، بلکه کاربرد معمول یک وبگاه شبکه‌های اجتماعی را نیز فراهم می‌کند. در حال حاضر از کاربران دعوت می‌شود تا پروفایل بسازند، علایق پژوهشی خود و مؤسسه‌های وابسته را بیان کرده، و آثار خود را فهرست کنند. فهرست انتشارات امکان محاسبه خودکار تعداد استنادها را در کنار دو سنجه دیگر نویسنده، یعنی شاخص اچ و شاخص آی ده[1] میسّر می‌کند. گوگل اسکالر از کاربران دعوت می‌کند تا رابطه خود را با سایر پژوهشگران به شکل هم–نویسندگان[2] عمومی بیان کنند، همچنین اجازه می‌دهد که مقاله‌های جدید سایر نویسندگان یا استنادها را نیز محرمانه دنبال کنند، شاخص جدیدی از تاثیر را برای پژوهشگران فراهم می‌کند: تعداد دنبال‌کننده‌هایی که در گوگل اسکالر دریافت می‌کنند. البته این سنجه هم اکنون بسیار محدود به نظر می‌رسد زیرا بیش‌تر پژوهشگران پروفایل یا دنبال‌کننده‌ای ندارند.

گوگل اسکالر برای یافتن محتوا دو راه دیگر را نیز به کاربران معرفی می‌کند. «روزآمدی‌های من[3]» که بر اساس استنادهای کاربران پیشنهاداتی می‌دهد، و «سنجه‌ها» امکان مرور آثار برتر را با توجه به زبان و دسته، با بیش‌ترین انتشارات استنادی هر مجله که در فهرست رتبه‌بندی قرار دارند را برای کاربران میسّر می‌کنند.

گوگل اسکالر کتاب‌سنجی‌هایی کارهای پژوهشی که قبلاً مانند آن ارائه نشده را به متخصصان اطلاعات و پژوهشگرانی می‌دهدکه به دنبال آنها هستند. نویسندگان، مجلات و مقالات همه در حوزه های سنجه‌های استنادی خود دیده می‌شوند.

1. h-index and i10-index 2. co-authors 3. My updates

الگوریتم‌های پیشنهاد و رتبه‌بندی از دید کاربران مخفی هستند، اما با این وجود بر آثاری که کاربران آن‌ها را می‌بینند تأثیر می‌گذارند یعنی استمرار اثر متیو[1] که موجب ثروتمندتر شدن ثروتمند و فقیرتر شدن فقیر می‌شود.

بدون شک، گوگل اسکالر به عنوان یک ابزار جستجو دسترسی به مدارکی را فراهم می‌کند که شناسایی آن‌ها به شکل‌های دیگر مشکل بود، و غالباً در مقایسه با پایگاه‌های اطلاعاتی استناد قدیمی برای جستجوهای متون عمومی (مانند بکمن و وون وردن[2]، ۲۰۱۲) و جستجوهای منابع استناد شده (مانند برگمن[3]، ۲۰۱۲) مثبت تلقی می‌شود. با این وجود، رویکرد خودکار برای شمول داده ها نه تنها موجب افزایش کیفیت داده‌ها می‌شود، بلکه برای اهداف کتاب‌سنجی داده‌های دارای ارزش مشکوک[4] را نیز معرفی می‌کند. آگویللو (۲۰۱۲) اشاره می‌کند که نمایش بیش از حد متون علمی عمومی و موضوعات آموزشی در گوگل اسکالر بدین معنی است که نتایج کتاب‌سنجی با نتایج پایگاه اطلاعات کتاب‌سنجی قدیمی قابل قیاس نیستند، و هرچند توسعه‌های اخیر گوگل اسکالر آن را برای تحلیل استنادی ارزشمندتر می‌کنند، اما استفاده از آن هم‌چنان استفاده از آن برای تصمیم‌های ارتقا و استخدام توصیه نمی‌شوند (جاسکو[5]، ۲۰۱۲). با این وجود، از گوگل اسکالر برای بررسی‌های کتاب‌سنجی استفاده می‌شود (وگستف و کولیر[6]، ۲۰۱۲).

اینکه نتایج کتاب‌سنجی گوگل اسکالر با خدمات کتاب‌سنجی قدیمی قابل قیاس نیستند باعث نفی ارزش آن نمی‌شود. بلکه باید به عنوان سنجه‌های تکمیلی تلقی شوند، که بینش جدیدی را در مورد تاثیر تحقیقات ارائه می‌دهند، چه این تأثیر در بین انواع انتشارات شناسایی شده متفاوتی آگویللو باشد (۲۰۱۲)، یا تأثیر فرمت‌هایی باشد که در پایگاه داده‌های کتاب‌سنجی سنتی کم تر ارائه شده است. کوشا، تلوال و رضایی[7] (۲۰۱۱) دریافتند که استنادهای کتاب برخط بیش‌تر از استنادهایی است که

1. Mathhew effect
2. Backmann and von Wehrden
3. Bergman
4. dubious value
5. Jacso
6. Wagstaff and Culyer
7. Kousha, Thelwall and Rezaie

در پایگاه‌های اطلاعات قدیمی یافت می‌شوند، و پیشنهاد می‌کنند که از آن‌ها می‌توان برای حمایت از ارزیابی پژوهش استفاده کرد.

گوگل اسکالر منبع بالقوه‌ای برای کتاب‌سنجی‌های وب ارتباطی و ارزیابانه است. اورتگا[1] و آگویللو (۲۰۱۳) نتیجه گرفتد که گوگل اسکالر برای مطالعات همکاری، و ایجاد یک نقشه علم بر مبنای کلیدواژه‌ها مناسب است (اورتگا و آگویللو، ۲۰۱۲). استخراج داده‌ها از گوگل اسکالر آسان نیست، چرا که در حال حاضر یک رابط برنامه کاربردی وجود ندارد. با این وجود، ارزش بالقوه داده‌ها به معنای این است که مردم برای استخراج داده‌ها ابزارهایی خلق کرده‌اند.

برنامه نرم‌افزاری پابلیش یا پریش[2] برای بارگیری داده های استنادی از گوگل اسکالر توسعه یافته است، و تسهیلات دسترسی به نتایج پژوهش‌های گوگل اسکالر، و نیز امکان دسترسی به سنجه‌های استنادی را فراهم می‌کند. مهم‌تر اینکه، این نتایج را تنها محدود به جستجوهای سطح اول می‌کند. برای مثال، درحالی‌که امکان بازیابی تعداد استنادهایی که یک مجموعه اسناد دریافت کرده‌اند وجود دارد، امکان بازیابی خودکار مدارک استنادی وجود ندارد، چرا که استفاده معقول از گوگل اسکالر در همین حد است، و کاربران موقتاً خود را از این خدمت محروم می‌کنند. متأسفانه این موضوع موجب دشواری تحلیل‌های عمیق‌تر می‌شود، برای نمونه در بررسی خوداستنادی، یا بررسی شبکه‌های استناد و نویسنده.

برای آن‌هایی که توانایی‌های فنی بیش تری دارند، یک طرح گوگل اسکالر به زبان پایتون[3] نوشته شده است که به عنوان یک ابزار خط فرمان یا به عنوان بخشی از یک برنامه گسترده‌تر می‌تواند استفاده شود (htmlwww.icir.org/christian/scholar.).

جستجوی دانشگاهی مایکروسافت

گوگل تنها موتور جستجو دارای پایگاه اطلاعات کتابشناختی نیست. جستجوی

1. Ortega 2. Publish or Perish (www.harzing.com/pop.htm) 3. Python

دانشگاهی مایکروسافت[1] در حال حاضر یک مجموعه رابط برنامه کاربردی دارد، و مهم‌تر این که در حال حاضر امکان دسترسی خودکار به داده‌ها را بدون اینکه پژوهشگران نگران باشند که یافته های آنها موقتاً قطع می‌شوند فراهم می‌کند. اگرچه کاربرانی که می‌خواهند از این رابط‌ها استفاده کنند، باید درخواستی مبنی بر یک کلید از رابط‌ها به اضافه جزئیات نحوه استفاده از این خدمات را ارائه بدهند.

جستجوی دانشگاهی مایکروسافت علاوه بر امکان مرور محتوا امکان جستجو را هم می‌دهد. به کاربران امکان مرور نویسنده‌ها، مجله‌ها، انتشارات‌ها، کنفرانس‌ها، سازمان‌ها و کلیدواژه‌های برتر را در یک حوزه خاص تحقیقاتی می‌دهد. برخلاف گوگل اسکالر جستجوی دانشگاهی مایکروسافت نقطه تمرکز تعداد بررسی‌های زیادی نبوده است، چرا که شواهد زیادی برای قابلیت تداوم این ابزار در بررسی کتاب‌سنجی نیست. رابط برنامه کاربردی جستجوی دانشگاهی مایکروسافت قابلیتی را ارائه می‌کند که در حال حاضر در ابزاری مانند پابلیش ار پریش نادیده گرفته می‌شود، چرا که نه تنها امکان دسترسی به تعداد استنادها را می‌دهد، بلکه امکان بازیابی آثار استنادکننده را نیز می‌دهد (مایکروسافت، ۲۰۱۲).

ابزارهای جانبی

وب برای پژوهشگران امکان استفاده از جمع‌کننده‌های[2] جدیدی را فراهم می‌کند، برای جمع‌آوری داده‌ها از خود ناشران، و انتخاب استانداردهایی که امکان برداشت فراداده‌ها از واسپارگاه‌های سازمانی را دارند. چنین مطالعاتی می‌تواند حاوی داده‌هایی باشند که در پایگاه اطلاعات کتابشناختی استاندار وارد نشده اند، مانند فراداده های مربوط به اشکال و جداول، و همچنین داده‌های مجلاتی که در پایگاه داده‌های کتابشناختی وجود ندارند. بسیار مهم است که قبول کنیم اهداف ناشران مجله و ارائه دهندگان خدمات کتاب‌شناختی با هم متفاوت است، بنابراین درحالی‌که خدمات کتابشناختی نیاز به کسب درآمد از طریق دسترسی به فراداده ها دارند، ممکن است

1. Microsoft Academic Search (http://academic.research.microsoft.com) 2. aggregator

ناشران از ساختار بزرگ‌تر نظام زیستی توسعه دهنده پشتیبانی نمایند تا بتوانند رفت و آمد را به سمت محتوای مجله واقعی ارتقا دهند.

ناشران دیدگاه ناشر گرایانه و مجله گرایانه نظام زیستی کتابشناختی را فراهم می‌کنند، درحالی‌که واسپارگاه‌ها یک دیدگاه سازمان محور و موضوع محور ارائه می‌کنند، هرچند همیشه تفاوت بین این دو مشخص نیست. arXiv.org یکی از واسپارگاه‌های تثبیت‌شده‌تری برخوردار است و برای محتوای خود رابط برنامه کاربردی مشروح‌تری[1] داشته، و هم‌چنین منبعی که مالکیت حقوقی آن تأیید شده است. در فهرست گوگل اسکالر انتشارات برتر هفت حوزه واسپارگاهی arXiv.org از بین ۵۰ انتشارات برتر آن وارد شده است. در حال حاضر بر اساس سازمان یا بر اساس حوزه موضوعی هزاران واسپارگاه برای میزبانی پیش‌چاپ‌ها[2] وجود دارد. از ماه ژوئن ۲۰۱۳ پایگاه داده‌ها واسپارگاه اوپن دوار[3] ۲۳۳۶ واسپارگاه‌ها را فهرست‌بندی کرد، که بیش‌تر آن‌ها فراداده‌های ساختاریافته بر اساس پروتکل پیشگام بایگانی‌های باز برداشت فراداده[4] هستند. بسیاری از ناشران نیز رابط‌هایی را برای دسترسی به داده‌های خود فراهم می‌کنند، مانند اسپرینگر[5] و الزویر[6]. جستجوی فراداده در کراس رف[7] امکان دسترسی به فراداده‌ها را از چندین ناشر فراهم می‌کند.

اگرچه ناشرها و واسپارگاه‌ها منابع غنی فراداده‌ای را برای بررسی‌های کتابشناختی فراهم می‌کنند، اما این حوزه در جامعه کتاب‌سنجی کلاً نادیده گرفته شده است. بخشی از آن می‌توان به این علت باشد که ابزارهای مناسبی برای بررسی ساده این منابع جدید وجود ندارد. احتمالاً توجهات بیش‌تری در راه است چرا که تعامل با ناشران و واسپارگاه‌ها نه تنها دسترسی به فراداده‌ها را امکان‌پذیر می‌کند، بلکه هم‌چنین

2. http://arxiv.org/help/api/index
2. preprints
3. OpenDOAR (www.opendoar.org)
4. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
5. Springer (http://dev.springer.com)
6. Elsevier(www.developers.elsevier.com)
7. CrossRef Metadata Search(http://search.crossref.org/help/api)

می‌تواند دسترسی روزافزونی را به متن کامل فراهم نماید. برای نمونه، رابط پلوس[1] دسترسی به فراداده‌ها و کل متن مقالات را میسّر کرده، و امکان جستجوی محدود را به بخش‌های خاصی مانند مقدمه یا نتایج میسّر می‌کند. علاوه بر ایجاد دسترسی به فراداده های بیش از ۵ میلیون سند برخط، اسپرینگر یک رابط نیز دارد که دسترسی آزاد به متن کامل بیش از ۸۰۰۰۰ مقاله را را فراهم می‌کند. دسترسی متن کامل به عنوان بخش بسیار مهمی از کشف منابع محسوب می‌شود چرا که ابزارهای استخراج داده ها به شناسایی مفاهیم و ارتباطات بین مقالات کمک می‌کنند، و در خصوص تحلیل استنادی می‌توانند اطلاعات جدیدی بدهند (برتین و آتاناسوا[2]، ۲۰۱۲).

با وجود افزایش تعداد موارد کتابشناختی در موتورهای جستجو کتابشناختی جدید، اما اذعان این نکته که آن‌ها هنوز همه چیز را نمایه نمی‌کنند اهمیت دارد. هنوز طیف گسترده‌ای از اسناد به صورت برخط در دسترس نیستند و هر موتور جستجو تنها گزیده‌ای از وب را نمایه‌سازی می‌کند. این گزینش به شدت به مواردی که می‌خواهد وارد کند و نحوه شناسایی این موارد بستگی دارد. ما هنوز فاصله زیادی داریم تا از هر موضوع کتابشناختی موجود نمایه‌ای داشته باشیم، چه رسد به هر موردی که تاکنون وجود داشته است.

تحلیل متن کامل

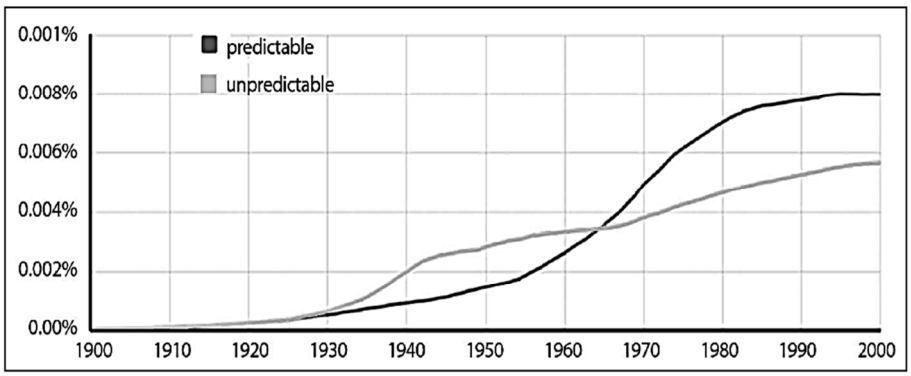
عبارت «سنجه‌های وبی» در سرتاسر این کتاب برای اشاره به اندازه‌گیری کمّی راه‌اندازی و استفاده از محتوای وب استفاده شده است. با این وجود، تمایز بین سنجه‌های وبی و کتاب‌سنجی‌سنتی اغلب روشن نیست، بویژه با توجه به موضوع کتاب‌سنجی وبی که در آن مرز بین آن دو به شدت نامشخص است؛ درحالی‌که استفاده از گوگل اسکالر می‌تواند به عنوان ابزار سنجه وبی در نظر گرفته شود، وب آو ساینس، که خدمت مشابهی را روی وب فراهم می‌کند، بیش‌تر پژوهشگران از آن

1. PLoS(http://api.plos.org) 2. Bertin and Atanassova

ابتداً به عنوان یک ابزار وب‌سنجی قدیمی یاد می‌کنند. بین این دو تفاوت‌های آشکاری از لحاظ کیفیت داده، دسترسی به خدمات و سوابق پیش از وب وجود دارد. این کتاب یک رویکرد تجربی داشته، و تمرکز آن بر حوزه‌هایی است که امکان ظهور آنها به صورت غیر برخط نبوده است، یا دست کم به طور غیربرخط سخت‌تر بوده است. تحلیل متن کامل یک نمونه از بررسی‌هایی را ارائه می‌کند که پیش از وب بسیار هزینه بر بوده‌اند، حتی اگر محتوای خودش بومی وب نباشد. برای مثال، سیلور[1] (۲۰۱۲) از طریق یک تحلیلی که از مقاله‌های کتابخانه دیجیتال جی‌استور[2] و در کتاب‌های داستان با استفاده از برنامه کاربردی انگرم در کتاب‌های گوگل[3] انجام داده نشان می‌دهد که چگونه بسامد استفاده از عبارت قابل پیش‌بینی[4] و غیر قابل پیش‌بینی[5] در طی زمان تغییر کرده است. این دو در آغاز قرن بیستم با بسامدی بسیار مشابه مورد استفاده قرار می‌گرفتند، اما پس از رکود بزرگ و جنگ جهانی دوم عبارت غیر قابل پیش‌بینی بیش از عبارت قابل پیش‌بینی مورد استفاده قرار گرفت، ولی در اواخر قرن بیستم عبارت قابل پیش‌بینی به طرز چشمگیری بیش از عبارت غیر قابل پیش‌بینی استفاده شد (شکل ۷-۱ را ملاحظه کنید). از آنجایی که هر دو عبارت دارای مجموعه داده‌های محدودی هستند، از لحاظ نظری انجام چنین مطالعاتی پیش از ظهور وب امکان‌پذیر بوده، هر چند بیش از حد گران. انجام مطالعات بر روی وب زمانی ارزشمند است که میلیون‌ها کاربر احتمالی برای یک خدمت وجود داشته باشد.

متن کامل اسناد این فرصت را فراهم کرده تا مزیت برخی واژه‌های خاص را در طی زمان نشان دهد: امکان پردازش زبان طبیعی و روش‌های استخراج داده ها میسّر است. از دیدگاه بررسی‌های کتاب‌سنجی سنتی این موضوع می‌تواند موجب ارائه اطلاعات بیش تری در خصوص تحلیل استنادی باشد (برتین و آتاناسوا، ۲۰۱۲) هرچند واقعاً امیدواریم که «دانش عمومی کشف نشده» پیدا شود.

1. Silver 2. JSTOR 3. Google Books Ngram Viewer (http://books.google.com/ngrams/info)
4. predictable 5. unpredictable



شکل ۷-۱. استفاده از عبارت‌های «قابل پیش‌بینی» و «غیر قابل پیش‌بینی» در کتاب‌های انگلیسی در طی قرن بیستم [گوگل و آرم گوگل به عنوان نماد تجاری برای شرکت گوگل ثبت شده‌اند، استفاده شده با اجازه.]

عبارت دانش عمومی کشف نشده توسط سوانسون[1] (۱۹۸۶) برای توصیف دانشی ابداع شد که در دامنه عمومی است، اما به این دلیل یافت نشده که افزایش تخصصی علم منجر به گسستگی هر چه بیش‌تر آن شده است. ظرفیت ابزارهای خودکار برای کمک به شناسایی این دانش بیش‌تر و بیش‌تر تأیید شده است، درحالی‌که پژوهشگران شاید حق دسترسی به مقالات مجله را برای مطالعه داشته باشند، اما آنها نمی‌توانند برای شناسایی روابط جدید میلیون‌ها مقاله موجود را پیمایش کنند (ون نوردن ، ۲۰۱۳). با این وجود متن‌کاوی[2] همیشه اساس بررسی‌های زیادی را تشکیل داده است، مانند برخی بررسی‌ها که متن‌کاوی را با کتاب‌سنجی ترکیب کرده‌اند (مانند هی و همکارانش[3]، ۲۰۱۲؛ هوونگ[4]، ۲۰۱۲). ابزارهای متن کاوی در طیف‌های گسترده‌ای قابل دستیابی هستند، هم به صورت تجاری (مانند شرکت کاوش ساس[5] و لکسیمانسر[6]) و هم به صورت متن‌باز (مانند کنیمه[7]، و گیت[8]. درحالی‌که این ابزارها را بطور روزافزونی برای اهداف پژوهشی انتخاب می‌کنند، اما استفاده از آنها برای مجموعه‌ای از داده، با توجه به محدودیت‌های آنها، نحوه

1. Swanson
2. text mining
3. He et al.
4. Hung
5. SAS Enterprise Miner
6. Leximancer
7. KNIME (http://tech.knime.org/knimetextprocessing,and GATE,
8. GATE (http://gate.ac.uk)

استخراج مفاهیم آن‌ها، و اینکه آیا برای مجموعه داده‌های یک دامنه پژوهشی خاص آزمایش شده‌اند یا خیر، یک جهت‌گیری برای آنها وجود دارد.

متن کامل انتشارات‌های قدیمی نیز می‌توانند اساس پژوهشی فراتر از علم را تشکیل دهند، از طریق کاوش مقاله نشریات، و تحلیل‌های متن کامل آن کتاب‌هایی که توسط پروژه‌هایی چون کتاب‌های گوگل[1] فراهم می‌شوند. اگرچه ممکن است به کیفیت دیجیتال‌سازی[2]، محدودیت‌های او.سی.آر[3] و مسائل حقوقی مختلف انتقاداتی وارد شود، پروژه‌ای مانند کتاب‌های گوگل، با هدف دیجیتال‌سازی تمام کتاب‌ها تا سال ۲۰۲۰ (جکسون[4]، ۲۰۱۰)، بدون شک موجب امتیاز پژوهش دیجیتالی برای افرادی می‌شود که از مهارت‌های لازمه و روش‌شناختی‌ها برخوردارند (باتک[5]، ۲۰۱۰). از زمان راه‌اندازی برنامه انگرم کتاب‌های گوگل در سال ۲۰۱۰ تعداد روزافزونی از مطالعات از آن برنامه برای بررسی ظهور کلمه‌ها و عبارت‌های درون مجموعه کتاب‌های گوگل استفاده کرده‌اند. تونگ، کمپبل و جنتایل[6] (b۲۰۱۲) آن را برای بررسی نحوه استفاده ضمیرهای مذکر و مؤنث در کتاب‌های ایالات متحده استفاده کرده و وضعیت زنان را بین سال های ۱۹۹۰ و ۲۰۰۸ و نیز افزایش کلمه‌ها و عبارت‌های فردگرایانه را از سال ۱۹۶۰ نشان دادند (تونگ، کمپبل و جنتایل، a۲۰۱۲). اگنال[7] (۲۰۱۳) سیر تکاملی رمان آمریکایی را بررسی کرده، و برای سیر تکاملی رمان کلیدواژه‌های برنامه انگرم را با یک چارچوب پیشنهادی جدید مقایسه کردند. آسِربی و همکارانش[8] (۲۰۱۳) بیان احساسات را در کتاب‌های قرن بیستم بررسی کرده، و برای کشف تغییر در احساسات در طی قرن بیستم برنامه انگرم کتاب‌های گوگل را با اطلاعات پایگاه اطلاعات واژگانی وردنت[9] ترکیب کردند. مارینر و مورهانگ[10] (۲۰۱۳) از انگرم برای بررسی احیای نظریه بحران‌زدگی[11] استفاده کرده و دریافتند، که این موضوع به یک بیداری محیطی در دهه ۱۹۶۰ مرتبط

1. Google Books
2. digitization
3. optical character recognition
4. Jackson
5. Batke
6. Twenge, Campbell and Gentile
7. Egnal
8. Acerbi et al.
9. WordNet((http://wordnet.princeton.edu)
10. Marriner and Morhange
11. catastrophism

بوده است. کتاب‌های گوگل تنها منبع موجود نیست، و تعداد منابع روزافزون دیگری برای تحلیل وجود دارد. برای مثال، بُد[1] (۲۰۱۲) برای گردآوری اطلاعاتی در خصوص متون استرالیایی، به جای استفاده از متون متداول اندک از آستلیت[2] به عنوان یک منبع متون استرالیایی استفاده کرد.

بدون شک متن کامل برای حوزه‌های جدید پژوهش کتاب‌سنجی ارزشمند است چرا که امکان انجام طیف وسیعی از بررسی‌هایی که قبلاً امکان‌پذیر نبوده است را میسّر می‌سازد، و در مورد برنامه انگرم گوگل ابزار بسیار شهودی را برای پژوهشگران بسیاری از رشته‌ها مهیا می‌کند. اگرچه وب موجب تسهیلاتی در ترکیب موارد جانبی کتابشناختی و متن کامل شده است، فضایی را نیز برای توسعه یک زیست‌بوم بسیار غنی‌تری حول و حوش مواد کتابشناختی فراهم می‌سازد.

بستر بزرگ‌تر

کتاب‌سنجی سنتی اطلاعات محدودی از کاربرد و تأثیر موضوعات کتابشناختی فراهم کرده‌اند. اگرچه ممکن است این بحث پیش بیاید که استنادها مهم ترین مورد برای بررسی هستند چرا که آنها این موضوع را که ایده‌های سایرین زیر بنای ایده‌های یک فرد است را شدیداً تصدیق می‌کنند، اما این بحث نیز مطرح است که برای مدت زیادی استنادها مهم‌ترین مورد برای بررسی کتابشناختی بوده‌اند چرا که آن‌ها تنها شاخص برای بیش‌تر خروجی‌های علمی بودند که هر نوع تأثیری داشتند! هرچند اطلاعات فروش کتاب و تیراژ نشریات برای ناشران قابل دستیابی است، اما این موضوع چشم‌انداز بسیار محدودی را ایجاد می‌کند، چرا که تیراژ مجله در سطح مجله تجمیع می‌شوند نه در سطح مقاله، و کتابخانه‌ها نقش بسیار حیاتی در محیط علمی دارند، زیرا چندین کاربر به یک نسخه طی چندین بار دسترسی دارند.

اگرچه در حال حاضر یکپارچگی سیستم‌های کتابخانه‌ای موجب می‌شود اطلاعاتی در خصوص امانت کتاب‌ها و جستجوهای فهرست کتابخانه عرضه شود،

1. Bode 2. AustLit (www.austlit.edu.au)

اما آنها پدیده بسیار جدیدی در تاریخ انتشارات علمی هستند، و حتی از وقتی که اطلاعات قابل دستیابی بوده در انبارهای سازمانی میزبانی شده‌اند. هنگامی که مقالات مجله تنها به صورت نسخه‌های کاغذی در قفسه‌های کتابخانه وجود داشتند، هیچ چیزی غیر از تای گوشه صفحه‌ها نبود که نشان بدهد از یک مقاله مجله چند بار کپی شده یا خوانده شده است.

در مقابل، وب اطلاعات گسترده‌ای از تمامی انواع موضوعات کتابشناختی فراهم می‌کند، از اطلاعات استفاده و نگهداری موضوعات کتابشناختی گرفته، تا اشاره‌های موضوعات کتابشناختی در وب و وبگاه‌های شبکه‌های اجتماعی.

استفاده و موجودی‌ها

کتابخانه‌ها و ناشران هم‌چنان اطلاعات فراوانی درباره استفاده از منابع کتابشناختی در اختیار دارند، و از آنجایی که این منابع به صورت دیجیتالی در دنیای برخط موجود شده‌اند لذا طیف وسیع‌تری از داده‌ها در دسترس قرار گرفته است.

از آنجایی که فهرست کتابخانه ها و فهرست اتحادیه‌هایی چون ورلدکت[1] و کوپک[2] در وب قرار داده شده‌اند، آنها موجب دسترسی آسان به منابع اطلاعاتی در موجودی کتابخانه‌ها می شوند. استنادهای کتابخانه[3] در ارتباط با ظهور یک کتاب در فهرست اتحادیه ملی یا بین المللی ابداع شده است (وایت و همکارانش[4]، ۲۰۰۹). می‌توان به درستی استدلال کرد که استنادهای کتابخانه در مفهوم واقعی خود یک سنجه وبی نیستند. با این وجود، داده ها از وب جمع‌آوری نمی‌شود، ولی در عوض از پایگاه اطلاعاتی که واسطی در وب دارند جمع‌آوری می‌شود. البته استنادهای کتابخانه به اندازه استنادهای نشریات در جامعه کتابداری ریشه ندارند، و در دوران پیش از وب نیز در دسترس نبوده‌اند. استنادهای کتابخانه می‌تواند منبع ارزشمند فوق‌العاده‌ای باشد از آثاری که دیگر چاپ نمی‌شوند، و نیز برای آن‌هایی که نمی‌توانند به شکل قالب‌های فروشی اخیر درآیند.

1. WorldCat((www.worldcat.org)
2. Copac (http://copac.ac.uk)
3. Libcitations شمارشی از موارد در اختیار کتابخانه
4. White

خدمات اطلاعاتی و کتابخانه‌ای خودشان نیز مولد بزرگی برای اطلاعات کتاب‌سنجی هستند، و تعدادی پروژه وجود دارد که ظرفیت استفاده از قدرت اطلاعات نگهداری شده در فهرست‌های کتابخانه را نمایش می‌دهند. پروژه جیسک موسایک[1] در بریتانیا ظرفیت چنین داده‌ای را در حوزه علمی مورد تأکید قرار داده است. لازم به ذکر است که دانشگاه هدرسفیلد[2] در دوران کاهش امانت کتاب، عناوین تعداد امانت کتاب به ازای هر دانشجو، و تعداد امانت عناوین کتاب های منحصر به فرد را افزایش داد. با وجود اینکه قطعاً اطلاعات فراوانی است که می‌توان از داده های فهرست بدست آورد، اما مقدار حداکثری داده‌های موجود کتابخانه‌ها بسیار کم‌تر از مقدار در دسترس وبگاه‌هایی چون آمازون[3] است، و وضع به همین منوال خواهد ماند تا زمانی که بین کتابخانه‌ها اشتراک‌گذاری بیش‌تری از داده‌های کاربری صورت بگیرد. در آینده ای نزدیک مسائلی از قبیل محرمانگی کاربر احتمالاً مانعی برای دسترسی به مقدار وسیع داده های کتابخانه ای مؤسسات محسوب می‌شود. در نهایت، با این حال، هرچه مؤسسه‌ها در جمع‌آوری داده‌های کاربر ماهرتر می‌شوند، مانند گوگل ترندز رفتارهای فردی عینی و قابل تشخیص نیست، اما در نهایت این احتمال بنظر می آید که مقدار روزافزونی از امانت داده ها قابل دسترس خواهند شد. در سال ۲۰۱۲ کتابخانه هاروارد[4] اعلام کرد که دسترسی رایگان را به فراداده ۱۲ میلیون مورد در کتابخانه خود فراهم کرده است، و در آینده امانت داده‌ها را در دسترس قرار خواهد داد (هاردی[5]، ۲۰۱۲).

حرکت به سمت انتشارات دیجیتالی، به ناشران اطلاعات فراوان کاربرد را می‌دهد، این امر برای کتابخانه ها ضروری است مشروط به اینکه بخواهند بدانند از منابع الکترونیکی آن‌ها استفاده کارآمدی می‌شود یا خیر. این موضوع منجر به ایجاد اقداماتی چون کاربرد برخط منابع الکترونیکی شبکه‌ای[6] (کانتر) برای استاندارسازی

1. Jisc MOSAIC(http://ie-repository.jisc.ac.uk/466)
2. University of Huddersfield
3. Amazon
4. Harvard Library
5. Hardy
6. Counting Online Usage of Networked Electronic Resource (COUNTER)

نحوه گزارش استفاده داده ها و برداشت آماری کاربرد استاندارد شده[1] برای تسهیل برداشت این داده‌ها شده است (پسچ[2]، ۲۰۰۷). اغلب ناشران اصلی اکنون با کانتر موافق هستند، چراکه دسترسی به داده‌ها موجب می‌شود که از اطلاعات جدیدی از تأثیر محتوا بدست بیاورند، هرچند همانگونه که تلوال (۲۰۱۲) اشاره کرده است هنوز هم هیچ تضمینی نیست که آثار واقعاً مطالعه می‌شوند.

افزایش مصنوعی ارقام و اعداد کاربری می‌تواند، برای امانت دهی واژگان از جامعه بهینه‌سازی موتور جستجو، از طریق ابزارهای کلاه سیاه یا کلاه سفید اطلاعاتی تحقق یابد. اهمیت بهینه‌سازی موتور جستجو دانشگاهی مورد تأکید بل، گریپ و ویلد[3] (۲۰۱۰) بوده است، و در حالی که آنها استدلال می‌کنند که فریب موتورهای جستجو هدف این ایده نیست، بلکه کمک به آن‌ها است و راه مناسبی است، و عبور از آن مستلزم این است پژوهشگران به سادگی با بارگیری چندباره محتوای خود آمار و ارقام خود را بالا ببرند. درحالی‌که ادلمن و لارکین[4] (۲۰۰۹) شواهد محدودی مبنی بر فریب سیستم برای دلایل شغلی یافتند، اما آن را ناشی از مقایسه‌های اجتماعی دانستند، چر ا که پژوهشگران احتمالاً زمانی که شمارش بارگیری به وضوح افزایش می‌یابد می‌توانند سیستم را فریب دهند.

این احتمال که کاربران تلاش کنند تا سیستمی را که قطعاً راه سوءاستفاده از آن باز است را بازی دهند، لزوم برخورداری از یک مجموعه سنجه را به جای اتکا تنها بر یک یا دو سنجه یادآور می‌شود. این را می‌توان به وضوح در مقاله پلوس وان با عنوان «یک تحلیل عمیق از تکه‌ای آشغال»(کراف و همکارانش[5]، ۲۰۱۲) دید که در ۹ ماه نخست انتشار ۱۴۵۰۶ بازدیدکننده و ۱۰۵۱۶ اشتراک‌گذاری دریافت کرد. انکار اینکه عنوان و تصاویر همراه آن بر تعداد بارگیری‌ها تأثیراتی داشته مشکل خواهد بود، و اگر آن‌ها را جدا کنیم باعث تخمین بیش از حدی تأثیر علمی اثر می‌شود. با

1. Standardised Usage Statistics Harvesting Initiative (SUSHI)
2. Pesch
3. Beel, Gipp and Wilde
4. Edelman and Larkin
5. Krauth et al.

این وجود، سنجه‌های بیش تر درباره تعداد دفعات استناد شده و نشانه‌گذاری یک مقاله با انتظارات از مقالات مشابه همخوانی دارند.

رفتار کاربر و ردگیری‌های کاربر

علاوه بر اطلاعات کتابخانه‌ها و ناشران، وب امکان جمع‌آوری اطلاعات گسترده ای در خصوص کاربران وب را نیز میسّر می‌سازد. همانند بررسی‌های ارزیابانه تأثیر وب، بررسی‌های کتاب‌سنجی می‌تواند بر رفتار کاربر یا ردگیری‌های حقیقی که از افراد به صورت برخط باقی می‌ماند معطوف باشد.

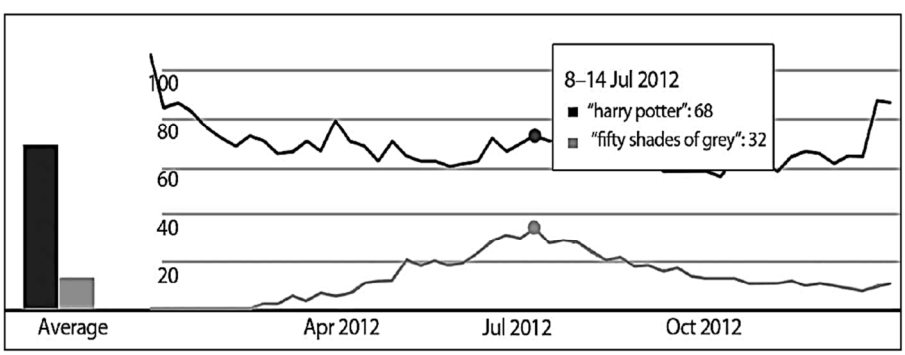
فروش کتاب – Amazon.co.uk

قطعی‌ترین نمونه از رفتار کاربری که به یک کتاب ابراز علاقه می‌کند خرید واقعی خود کتاب است، درحالی‌که فروشگاه‌های برخط کتاب ممکن است جمعیت‌شناسی کاربران را در دسترس قرار ندهند با این وجود، فهرست پرفروش‌ترین‌های آن‌ها، حوزه‌های مورد علاقه مردم را نشان می‌دهند. اندازه وسیع فروشگاه‌های برخط نه تنها نیازهای فراوان کاربر را برآورده می‌کند، که بخش قابل توجهی از فروش آنها ناشی از فروش میلیون‌ها جلد در تعداد ناچیزی است (اندرسون[1]، ۲۰۰۶)، بلکه فهرست بلندی از پرفروش‌ترین‌ها را نیز ارائه می‌کند. کتابداران محدود به پیگیری هر چه بیش تری چون پرفروش‌ترین کتاب‌ها، کتاب‌های مرجع، یا حتی کتاب‌های علوم اطلاعات و کتابداری نیستند؛ درعوض آن‌ها می‌توانند بر حوزه محدودی مثل کتابداری دیجیتالی یا مجموعه‌سازی تمرکز داشته باشند. شیوه مورد استفاده آمازون در تنظیم نمودارهای خود نشانگر این است که در حوزه‌هایی که فروش کم است نمودارها نسبتاً ناپایدار بوده، در حالی‌که رده‌بندی بسیاری از کتاب‌ها می‌تواند بی ربط باشد.

1. Anderson

گوگل ترندز

کتاب‌سنجی وب نباید تنها پیرامون فهرست پرفروش‌ترین‌ها باشد. پس از گذشت زمان های زیادی از خروج یک کتاب از فهرست پرفروش‌ترین‌ها نیز علاقه وافری به آن موضوع می‌تواند وجود داشته باشد، علاقه به موضوع الزاماً منجر به خرید یک کتاب نمی‌شود. کتاب‌ها همانند هر موضوع دیگری نیز می‌توانند در بین فعالیت‌های جستجوی افراد پیدا شوند، و گوگل ترندز می‌تواند در خصوص کتاب‌هایی که افراد در سراسر جهان از طریق موتور جستجوی گوگل آن‌ها را جستجو می‌کنند اطلاعاتی ارائه کند. اگرچه در سال ۲۰۱۲ کتاب *پنجاه سایه خاکستری*[1] ای. ال. جیمز[2] بسیار مورد توجه بوده، ۱۵ سال پس از چاپ اولین قسمت کتاب هری پاتر[3]، کتاب‌های هری پاتر هم‌چنان در صدر توجه آن سال بودند (شکل ۷-۲).



شکل ۷-۲. کتاب‌های هری پاتر در برابر کتاب پنجاه سایه خاکستری در گوگل ترندز

[گوگل و آرم گوگل به عنوان نماد تجاری برای شرکت گوگل ثبت شده‌اند، استفاده شده با اجازه.]

گوگل تأیید کرده است که گوگل ترندز می‌تواند اطلاعات موضوعاتی از قبیل کتاب‌ها و نویسندگان و در حال حاضر ۱۰ نمودار برتر را در طیف گسترده‌ای از موضوعات شامل کتاب‌ها و نویسندگان را بدهد (www.google.co.uk/trends/topcharts). هرچند در حال حاضر نمودارها فقط شامل داده‌های موجود در ایالات متحده است، بررسی نحوه

1. Fifty Shades of Grey 2. E. L. James 3. Harry Potter

جستجوی یک عبارت در اطراف دنیا امکان پذیر است. درحالی‌که به ندرت ممکن است انجیل در میان پرفروش‌ترین‌های آمازون قرار گیرد، اما از سال ۲۰۰۴ بر نمودار کتاب گوگل ترندز غالب شده و تنها یک بار اولین نبوده است.

هرچند زمانی که سخن از غالب آثار کتابشناختی می‌شود، که آن هم به ندرت جستجو می‌شوند، گوگل ترندز می‌تواند اطلاعات اندکی را فراهم کند. به همین دلیل بررسی ردگیری‌هایی که در وب باقی مانده‌اند ضروری است.

ارزیابی تأثیر وب

درحالی‌که گوگل ترندز تنها رفتار جستجوی جمعی را ارائه می‌کند، موتورهای جستجو برای دسترسی به صفحه‌های وبی طراحی شده‌اند هر چند که چند نتیجه با معیارها هم‌خوانی دارند. ارزیابی تأثیر وبی زمانی اجرا می‌شود که از یک موتور جستجو برای بررسی ظهور واژه‌ها و عبارت‌های خاصی استفاده شده، و تحلیل استنادی وب با استفاده از یک ارزیابی تأثیر وبی برای بررسی تأثیر مقالات دانشگاهی انجام شود (تلوال، ۲۰۰۹). می‌توان به این عنوان تصور کرد که این موضوع یک رویکرد قوی برای عادت شایع مردم است که نام خود یا شغل آنها را در گوگل جستجو کنند. برای مثال، ممکن است یک نفر بخواهد تأثیر نسبی انتشارات‌های پژوهشی دولت بریتانیا را بررسی کند. این پژوهش می‌تواند با جستجوی عبارتی در عناوین انتشارات از طریق یک موتور جستجو بررسی شود. جدول ۷-۱ (صفحه بعد) یافته‌های چنین پژوهشی را بر اساس پنج پژوهش نخست و تحلیل گزارش‌های منتشره در ژانویه ۲۰۱۲ در وبگاه GOV.UK[1] و بررسی در گوگل اسکالر و جستجوی بریتانیا گوگل نشان می‌دهد. هرچند گوگل اسکالر قلمرو انواع مدارک یک پایگاه اطلاعاتی کتابشناختی را بیش‌تر کرده است، اما این موضوعات خاص را نمایه نکرده است.

1. (www.gov.uk/government/publications)

جدول ۷-۱. تحلیل استنادی وب برای پنج انتشار دولتی بریتانیا در گوگل (جولای ۲۰۱۳)

عنوان گزارش	پژوهش گوگل	گوگل [site:gov.uk-]	گوگل filetype:pdf-] [site:gov.uk
«مشارکت در یادگیری پست-۱۶: آموزش کارآمد در مدارس»	گزارش یافت نشده	۴۸	۱
«نوآوری ناشی از پاداش برای توسعه»	گزارش یافت نشده	۵۷	۱
«همکاری نظارتی و تجارت بین‌المللی»	گزارش یافت نشده	۰	۰
«مورد اقتصادی برای HS2: ارزش برای بیانیه پول»	۳ جستجوی مرجع استنادی	۱۰۵	۶۳
«مورد اقتصادی برای HS2 ارزیابی روزآمد شده ارزیابی منافع کاربر حمل و نقل و منافع اقتصادی وسیع‌تر»	۳ جستجوی مرجع استنادی	۱۱۴	۸

این بدین معنا نیست که موضوعات در سایر اسناد ذکر نمی‌شوند، همچنانکه دو گزارش درباره HS2، در یکی خط راه آهن پر سرعت، در واقع هر دو در سه گزارش دیگر ذکر شده اند. گوگل اسکالر تنها یک زیر مجموعه از اسنادی است که موتور جستجوی گوگل آن را نمایه‌سازی کرده است، و در واقع اسناد استنادی بسیار بیش تری را می‌توانیم پیدا کنیم. دو جستجو در گوگل برای هر اثر انجام دادیم: ابتدا یک جستجوی عبارتی در رابطه با *وبگاه: gov.uk* برای استخراج آن صفحه‌ها در دامنه gov.uk انجام شد و دوم یک جستجوی عبارتی با *نوع فایل: pdf* برای محدودکردن جستجو به فایل‌هایی با پسوند pdf. اعمال محدودیت به نوع فایل بدین معنی است که جستجو فقط به آن اسنادی محدود است که احتمالاً کیفیت بالاتری دارند.

همانگونه که پیش‌تر هم گفته شد، تعداد تخمینی نتایج در هنگام بازدید از آخرین صفحه دستیابی شده دقیق‌تر از تعداد تخمینی نتایجی است که موتور جستجو شناسایی شده در صفحه اول نشان می‌دهد. جدول ۷-۱ یافته‌های یک تحلیل استنادی وبی مربوط به پنج انتشار پژوهشی دولتی بریتانیا را نشان می‌دهد که از طریق رفتن به

آخرین صفحه نتایج، و گنجاندن هر صفحه مشابهی که گوگل در ابتدا حذف کرده، بدست آمده است.

از تحلیل‌های اولیه به نظر واضح می‌رسد که مستندهای HS2 تاکنون بیش ترین تأثیر را داشته‌اند، این موضوع چندان غیر منتظره نیست چرا که این پروژه بسیار بحث‌انگیز است، هرچند که همکاری نظارتی و تجارتی بین المللی[1] هیچ تأثیری نداشته است. درحالی‌که ممکن است این موضوع برای تحلیل‌های ابتدایی که یافته‌های قوی‌تری لازم باشند کافی باشد، اما بررسی ماهیت استنادهای وبی از طریق یک تحلیل محتوا مورد نیاز است تا ماهیت استنادها مشخص شود(فصل ۴ را ملاحظه کنید).

وبگاه‌های شبکه‌های اجتماعی

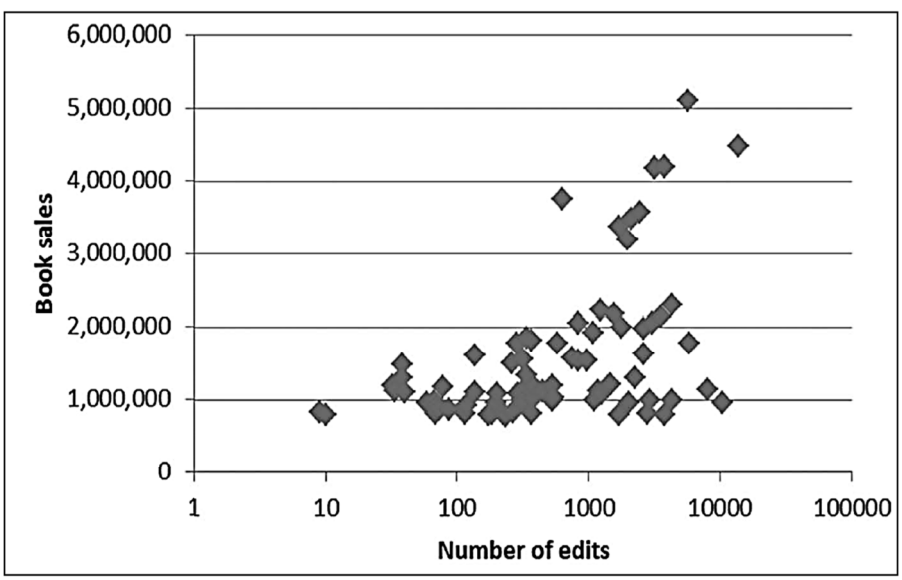
بیش‌تر مردم صفحه‌های وبی خود را خلق و میزبانی نمی‌کنند، بلکه از تعداد زیادی وبگاه‌های شبکه‌های اجتماعی که فرایند ایجاد محتوا را تسهیل می‌کنند استفاده می‌کنند. فرقی میان بررسی تأثیر یک کتاب و نظارت[2] بر تأثیر هر موضوع دیگری نیست. هر چند که، برای بررسی تأثیر منابع کتابشناختی، و ابزاری که برای جمع‌آوری اطلاعات از چندین وبگاه و خدمات ساخته شده‌اند، برخی وبگاه‌های شبکه‌های اجتماعی توجه بیش‌تری را بخود جلب می‌کنند.

درحالی‌که استنادات در درجه نخست برای جمع‌آوری اطلاعاتی در خصوص کارهای علمی مورد استفاده قرار می‌گیرند، وبگاه‌های شبکه‌های اجتماعی در حال حاضر می‌توانند اطلاعات همه نوع اثری را ارائه دهند. در حال حاضر تعداد زیادی از شبکه‌های اجتماعی متمرکز بر روی بحث کتاب (مانند Shelfari.com، LibraryThings و GoodReaders) به آمازون پیوسته اند و هزاران نقد در قالب یک عدد از ۱ تا ۵ وجود دارند. درحالی‌که چنین وبگاه‌های شبکه‌های اجتماعی کتاب محور می‌توانند اطلاعات برخی از کتاب‌های پرطرفدار را ارائه بدهند، برخی از وبگاه‌های شبکه های

1. Regulatory cooperation and international trade 2. Monitoring

اجتماعی عمومی‌تری مانند توییتر نیز قادر به انجام این کار هستند. حتی ممکن است برای آگاهی از دیدگاه مردم برای برخی از پرفروش‌ترین‌ها از ویکی‌پدیا استفاده شود.

شکل ۷-۳ به مقایسه تعداد ویرایش‌های ویکی‌پدیا (جمع آوری شده در ۵ آوریل ۲۰۱۳) با تعداد کتاب‌های فروخته شده ۸۲ کتاب از ۱۰۰ کتاب پرفروش‌ترین‌های کل تاریخ در بریتانیا که به عنوان دارنده صفحه ویکی‌پدیا خود را شناسایی کرده‌اند می پردازد، ۱۰۰ کتاب پرفروش از فهرست وب‌نوشت *گاردین*[1] استخراج شده است. شمار ویرایش‌های نشان داد شده روی مقیاس لگاریتمی به دلیل ویرایش توزیع قانون قدرت در شمار ویرایش‌هایی است که صفحات دریافت کردند. هنگام محاسبه ضریب همبستگی پیرسون بین تعداد ویرایش‌هایی که یک صفحه کتاب دریافت می‌کند و تعداد کتاب‌های فروخته شده به لحاظ آماری یک همبستگی معنی‌دار ($p<1\%$) دیده می‌شود.



شکل ۷-۳. مقایسه‌ای بین فروش کتاب‌ها و تعداد ویرایش‌های ویکی‌پدیا برای پرفروش‌ترین کتاب‌های بریتانیا، آوریل ۲۰۱۳

1. *Guardian* Datablog(www.guardian.co.uk/datablog)

اگرچه شکل ۷-۳ نشان می‌دهد که تعدادی کتاب با وجود فروش سطح پایین اما صفحه‌های ویکی‌پدیا آنها ویرایش نسبتاً زیادی دارد، و این موضوع می‌تواند انعکاسی از حامیان قویدر سراسر جهان باشد و با فیلم اخیر همخوانی دارد. برای مثال، *ارباب حلقه‌ها*[1] از جی. آر. آر. تولکین[2] دومین مورد دارای بیش‌ترین تعداد ویرایش است، علیرغم فروش نسبتاً اندک ۹۶۷۴۶۶ جلد آن در بریتانیا از زمان انتشار.

درحالی‌که وبگاه‌های شبکه‌های اجتماعی امکان ارائه اطلاعات اضافی همه نوع کتابی را دارد، امکان ارائه اطلاعات کارهای علمی که توجه بیش‌تری را جلب کرده است را دارد، با ماهیت ساختار یافته محتوای شبکه‌های اجتماعی بنیان دگرسنجه‌ها را شکل داده، و استفاده ماهیت فناوری‌های وب نسخه ۲/۰ برای ایجاد پالایش‌گرهای جایگزین و شاخص‌های تأثیر پژوهش(پرایم و همکارانش، ۲۰۱۰).

تاکنون در مطالعات بسیاری یک نوع پیوند بین تأثیر برخط و تأثیر سنتی نشان داده شده است همانگونه که اندازه‌گیری‌های استنادی قدیمی نشان می‌دهد. ایسنباچ (۲۰۱۱) دریافت که اشاره‌های توییتر شاخص اولیه استنادها هستند، شوای، پپ و بولن[3] (۲۰۱۲) توییتر را به عنوان شاخصی از بارگیری‌ها و استنادها یافتند، و بار-ایلن و همکارانش (۲۰۱۲) یک همبستگی را بین نشانه‌گذاری‌های شبکه های اجتماعی دانشگاهی مندلی یافتند. این بررسی‌ها در درجه نخست در مقیاس کوچکی بوده، و روی یک سنجه تمرکز داشتند. در یک بررسی بسیار بزرگ‌تر، که ۱۱ نوع مختلف دگرسنجی را بررسی کرد (شامل توییتر، لینکداین، گوگل پلاس و اشاره‌های ردیت[4])، تلوال و همکارانش (۲۰۱۳) دریافتند که به استثنای توییتر شواهد به اندازه‌ای غالب نبود که در عمل سودمند باشند، هرچند هنگامی که شواهد کافی موجود بود همبستگی بیش‌تری بین استنادها و دگرسنجه‌ها پیدا می‌شود.

اگرچه بعید است که در آینده نزدیک دگرسنجه ها نقش استنادها را بازی کند، اما بدون شک حوزه‌ای است که جذابیت فزاینده‌ای دارد، بطوری که احتمالاً اهمیت

1. *Lord of the Rings* 2. J. R. R. Tolkien 3. Shuai, Pepe and Bollen 4. Reddit mentions

آن‌ها با تغییر عادت‌های کاری پژوهشگران سیر صعودی می‌یابد، پژوهشگران می‌توانند با پروژه‌هایی مانند ORCID[1] آسان‌تر شناسایی شوند، و داده‌ها را می‌توان در آن واحد به سادگی از چندین وبگاه جمع‌آوری کرد (موضوعی که در فصل بعد به آن خواهیم پرداخت).

تاکنون چندین ابزار هستند که به بررسی دگرسنجه‌ها کمک می‌کنند، هرچند گران قیمت‌ترین آن‌ها پروژه‌های تجاری هستند (مانند، www.plumanalytics.com، www.altmetric.com). بسیاری از ناشران از این خدمات بیش از پیش برای ارتقا تأثیر مقاله‌های نشریات خود استفاده می‌کنند. ایمپکت استوری[2] خدمت رایگانی است، که به پژوهشگر امکان ساخت پروفایل برای خروجی‌های پژوهش، و جمع‌آوری اطلاعات تأثیر خروجی‌ها را فراهم می‌کند.

نتیجه‌گیری

محور بحث این فصل انتشارات علمی سنتی بود که هم‌چنان بخش مهمی از کار یک پژوهشگر به شمار می‌روند. همانگونه که بسیاری اذعان دارند، ماهیت نشر به سرعت در حال تغییر است و کتابداران آینده باید در مقایسه با اسناد متن خطی که طی صدها سال اساس خروجی دانشگاهی را تشکیل داده اند، برای ارزیابی تأثیر خود و طیف بسیار وسیع‌تری از خروجی‌ها راه‌های جدیدی را پیدا کنند.

برای آینده قابل پیش‌بینی است، هرچند که، انتشارات راه‌اندازی‌شده در درک تأثیر کار پژوهشگران نقشی اساسی را ایفا خواهند کرد، و این ایفای نقش قطعاً از طریق سنجه‌های جایگزین خواهد بود. برای بررسی تأثیر انواع مختلف منابع، به روش‌های مختلف و در جاهای متفاوت به مجموعه‌ای از سنجه‌ها نیاز خواهد بود.

1. (http://orcid.org) 2. Impact Story (http://impactstory.org)

سنجه‌های وبی و وب داده

مقدمه

تاکنون بیش‌تر بحث در خصوص ظرفیت سنجه‌های وبی برای کسب اطلاعات در وب اسناد بوده است: در وب به یک سند خاص چند بار اشاره شده است؟ چند بار از یک صفحه وب بازدید شده است؟ یک صفحه ویکی چند بار روزآمد شده است؟ هر چند، توصیه شده است که ما در حال حرکت از وب اسناد به سمت وب داده‌ها هستیم، جایی که داده‌ها در آن بیش از هر زمان دیگری به شکل خواندنی با ماشین[1] موجود هستند. این فصل با توضیحی درباره این «وب داده» و تمام جنبه‌های چندگانه آن آغاز می‌شود، پیش از اینکه پیامدهای روزافزون این داده‌های ساختار یافته در توسعه انواع مختلف سنجه‌های وبی گسترش یابد. در پایان این فصل به برخی از ابزارهای موجود برای بررسی وب داده‌ها پرداخته می‌شود.

وب داده

عبارت وب داده به داده‌های ساختاریافته‌ای که به شکل ماشین خوان در آمده و در وب صورت باز منتشر شده است اطلاق می‌شود(استورات، ۲۰۱۱). وب داده متفاوت از وب کنونی نیست بلکه به زیرمجموعه‌ای از آن اشاره داشته و شامل طیف وسیعی

1. machine-readable form

از فناوری‌های مختلف است، از یک صفحه گسترده گوگل ابری[1] گرفته تا یک صفحه گسترده اکسل که در انبار سازمانی جا داده شده؛ از رابط‌های برنامه‌نویسی که امکان دسترسی به داده ها را از وبگاه‌ها و خدمات وب ۲/۰ فراهم می‌کنند تا صفحه‌های وبی با ریزشکل[2]، ریزداده[3] یا چارچوب توصیف منابع در قسمت ویژگی‌ها[4]. برخی از این فناوری‌ها قبلاً در طول کتاب معرفی شده‌اند. در اینجا با جزئیاتی بیش‌تر به برخی از فناوری‌هایی که به وب داده ها کمک می‌کنند می‌پردازیم. قبل از اینکه وارد بحث پیامدهای وب داده‌ها در اشکال مختلف آن شویم، به منافع و زیان‌های فناوری‌های مختلف برای توسعه سنجه‌های وبی می‌پردازیم.

همانطور که در جای دیگری بحث شد، کتابداران سابقه طولانی دارند برای دسترس‌پذیرسازی اسناد، و برای تسهیل دسترسی به وب داده‌ها در مقیاس بسیار بزرگ‌تر ایده‌آل هستند (استوارت، ۲۰۱۱). تسهیل دسترسی به وب داده ها همچون تکامل سریع خدمات اطلاعات سنتی، شیوه جدیدی را برای توسعه خدمات اطلاعاتی فراهم می‌کند. درحالی‌که برای بسیاری از کتابداران دورنمای یادگیری برنامه‌نویسی دورنمایی رعب‌آور خواهد بود، دوره‌های برخط باز فراگیر[5] از وبگاه‌هایی چون اوداسیتی[6] و کورسرا[7] راه ساده‌ای را به کتابداران برای توسعه مهارت‌های فنی آنها ارائه می‌کنند.

همچنین یک انجمن پویا از اسکرپ‌های موجود در اسکرپرویکی وجود دارد[8]. این وبگاه یک بستر مبتنی بر وب ایجاد می‌کند که در آن برنامه‌نویس‌ها می‌توانند کدی ایجاد کنند که به صورت خودکار اجرا و داده ها ذخیره شوند. حتی در مواقعی که فردی متوجه شود داده ها هنوز اسکرپ نشده است و خود آنها مهارت‌های لازمه را برای نوشتن برنامه اسکرپ ندارند، گزینه درخواست داده‌ها نیز وجود دارد. می‌توانید هزینه کسی را پرداخت کنید تا داده‌ها را برای شما اسکرپ کند.

1. cloud
2. Microformats
3. Microdata
4. Resource Description Framework in Attributes (RDFa)
5. massive open online courses (MOOCs)
6. Udacity (www.udacity.com)
7. Coursera (www.coursera.org)
8. http://scraperwiki.com

از اسناد داده‌ها...

ساده‌ترین راه برای کمک به وب داده‌ها این است که داده‌ها را به شکل ماشین خوان قرار داده و در ابر بارگذاری شود یا در سرور وب قرار داده شود. اهمیت فرمت‌های ماشین خوان را هنگام ملاحظه تفاوت بین یک صفحه گسترده پی.دی.اف و یک صفحه گسترده اکسل می‌توان مشاهده کرد. پی.دی.اف برای شخصی که می‌خواهد مجموعه‌ای از داده‌ها را بخواند یا آن را چاپ کند فرمت ارزشمندی است، اما کاربرد کم‌تری دارد برای کسی که می‌خواهد از داده‌ها استفاده دوباره کند یا آن را مورد آزمایش قرار دهد. انتشار داده ها در اسناد ماشین خوان بیش از پیش به دو دلیل عمده متداول است: با این فرمت می‌توان به آسانی سند داده‌ای را به اشتراک گذاشت، و شناخت روزافزونی برای اهمیت داده‌های باز است.

برای فردی با یک فضای سرور و یک کاربر اف.تی.پی[1] انتشار یک صفحه گسترده به صورت برخط همواره فرایند نسبتاً ساده‌ای است: بارگذاری یک صفحه گسترده در یک اکسل یا فرمت CSV در یک سرور وبی دشوارتر از بارگذاری یک فایل اچ.تی.ام.ال یا یک تصویر نیست. تعداد زیادی از ابزارهایی وجود دارد که در صورت نیاز به فضای سرور و یک کاربر اف.تی.پی از گذشته آن را به وجود می‌آورند: یک فایل ممکن است در سرویسی مانند گوگل درایو[2] بارگذاری شود سپس برای هرکسی که پیوند را داشته باشد به صورت عمومی در دسترس قرار گیرد؛ یا داده‌ها را بتوان در وبگاهی مانند چشم‌های بسیاری (منی آیز[3]) بارگذاری کرد، که آن هم امکان دیداری‌سازی داده‌ها را فراهم می‌کند.

بزرگ‌ترین محرک انتشار داده ها الزاماً فنی نیست، بلکه افزایش شناخت ظرفیت داده های باز عمومی‌تر است. سازمان‌ها پی برده‌اند که داده های باز می‌توانند به خرد جمعی تلنگر بزنند؛ دولت‌ها مسئول پاسخگوی شفافیت بیش‌تر درخواست‌ها و شناسایی ظرفیت اقتصادی داده‌های خود هستند؛ و جامعه علمی ظرفیت داده های باز را برای

1. FTP client 2. Google Drive 3. Many Eyes (www958.ibm.com/so%ware/analytics/manyeyes)

کمک به پیشرفت علم به رسمیت می‌شناسد. مدارک داده‌ها شاید بهینه‌ترین راه برای انتشار این داده‌ها نباشند، اما یک شروع است و حاوی طیف وسیعی از نمونه‌های خلاقانه انتشار برخط داده‌ها است و استفاده مجدد از داده‌ای می‌کند که که در دسترس عموم قرار گرفته است. وب‌نوشت داده‌ای گاردین[1] مرتباً داده‌های مرتبط با رویدادهای خبری تحت پوشش خود را با استفاده از ابزاری چون صفحه‌های گسترده گوگل برای به اشتراک‌گذاری داده‌ها و منی‌آیز برای دیداری‌سازی آن‌ها منتشر می‌کند.

مانند هر نوع سند دیگری، یک مجموعه داده می‌تواند اطلاعات ارزشمندی درباره نحوه توسعه علم ارائه کرده، و پژوهش بررسی شده استفاده مجدد از مجموعه داده‌هایی که از طریق پایگاه داده‌های استنادی راه‌اندازی شده‌اند. برای نمونه پیووار، کارلسون و ویژن[2] (۲۰۱۱) شروع به ردگیری ۱۰۰۰ مجموعه داده از واسپارگاه‌های عمومی کرده‌اند تا نحوه اشاره به آن‌ها را در متون علمی مشاهده کنند، هرچند نویسندگان اذعان دارند که چنین رویکردی موجب می‌شود که سایر کاربردهای بالقوه داده‌ها، مثل استفاده مجدد سیاست‌گذاران، در آموزش، یا برای مطالعه شخصی نادیده گرفته شوند. از استنادهای وبی می‌توان برای عرضه درکی گسترده‌تر، یا اطلاعات کاربردی از یک چشم‌انداز متمایز استفاده کرد. همانطور که برخی واسپارگاه های داده های دانشگاهی می‌توانند آمارهای کاربردی ارائه کنند، دولت‌ها نیز در میان بزرگ‌ترین محرک‌های داده‌های باز بوده و وبگاه داده‌های دولت بریتانیا[3] اطلاعات بازدیدها و بارگیری‌های این وبگاه، ناشران مختلف (بخش‌های دولتی) و مجموعه داده‌های شخصی را ارائه می‌کند.

تحلیل کاربرد مجموعه داده‌ها ظرفیت واقعی وب داده‌ها برای سنجه‌های وبی نیست، هرچند بدون شک می‌تواند اطلاعات جذابی ارائه کند، اما تحلیل مقیاس بزرگ از خود داده‌ها است. دستیابی به این امر با حبس داده‌ها در انبارها ایجاد

1. Guardian Datablog (www.guardian.co.uk/news/datablog)
2. Carlson and vision
3. British Government (http://data.gov.uk/blog/siteusage)

نمی‌شود، بلکه از طریق اشتراک‌گذاری داده‌ها در وب برمبنای استانداردهای پذیرفته شده رایج تسهیل می‌گردد. این امر موجب می‌شود که داده ها از سراسر وب خوانده و درک شوند بدون اینکه لازم باشد که یک پژوهشگر خصوصیات را در ساختار هر یک از مجموعه داده‌های خاص ایجاد کند.

... تا وب معنایی

بیانیه وب معنایی منتشر شده در مجله ساینتیفیک امریکن[1] در ۲۰۰۱ وعده وبی را داده است که در آن بسیاری از کارهای محاسباتی روزانه مردم از طریق برنامه‌های رایانه‌ای به صورت خودکار انجام شوند (برنرس- لی، هندلر و لاسیلا[2]، ۲۰۰۱) مانند رزرو یک وقت ملاقات اضطراری با یک دندانپزشک؛ برنامه‌ریزی یک تعطیلات؛ شناسایی اسناد مرتبط؛ و بازیابی پاسخ‌های خاص برای پرسش‌های خاص. امروزه لازمه بقای چنین فعالیت هایی برخورداری از یک حد قابل توجهی از ورودی کاربر فراتر از شکل‌گیری جستجو است. حتی هنگام رزرو جا برای تعطیلات، که جمع‌کننده‌ها اطلاعات را جمع‌آوری کرده و همه را با هم در یک نقطه تجمیع می‌کنند، قبل از انتخاب هتل می‌توان از چندین جمع‌کننده مختلف بازدید کنیم.

بیش از یک دهه از زمانی که وب معنایی توجه‌های زیادی را به خود جلب کرده می‌گذرد، و به نظر می‌رسد موجب شده که بیش‌تر فعالیت‌های روزانه مردم کمی فرق کند. از آن زمان استانداردهای بیش‌تری که برای اجرای موفقیت آمیز وب معنایی لازم است (به اصطلاح دسته وب معنایی – http://en.wikipedia.org/wiki/Semantic_Web_Stack) راه‌اندازی شده‌اند، و در سال‌های اخیر تعداد روزافزونی از سازمان‌ها مجموعه داده‌هایی همچون داده‌های پیوندی[3]، از کاتالوگ موزه بریتانیا[4] گرفته تا سیستم طبقه‌بندی دهدهی دیویی و سرعنوان های موضوعی کتابخانه کنگره را ساخته‌اند. داده‌های پیوندی بر اهمیت انتشار داده ها در وب در سه‌گانه‌های چارچوب توصیف منابع، با استفاده از

1. *Scientific American*
2. Berners-Lee, Hendler and Lassila
3. Linked Data
4. British Museum

شناسه‌گر همگانی منابع[1] برای ارائه، و پیوند به سایر مجموعه داده‌ها (به جای دسترسی به داده ها به صورت جداگانه) تأکید می‌کند. با تصویب استانداردها، و ظهور لغت‌های تأیید شده در جوامع کاربری در سراسر وب، بررسی‌های سنجه وبی بسیار جامع‌تری میسّر خواهند شد.

ساخت وب معنایی

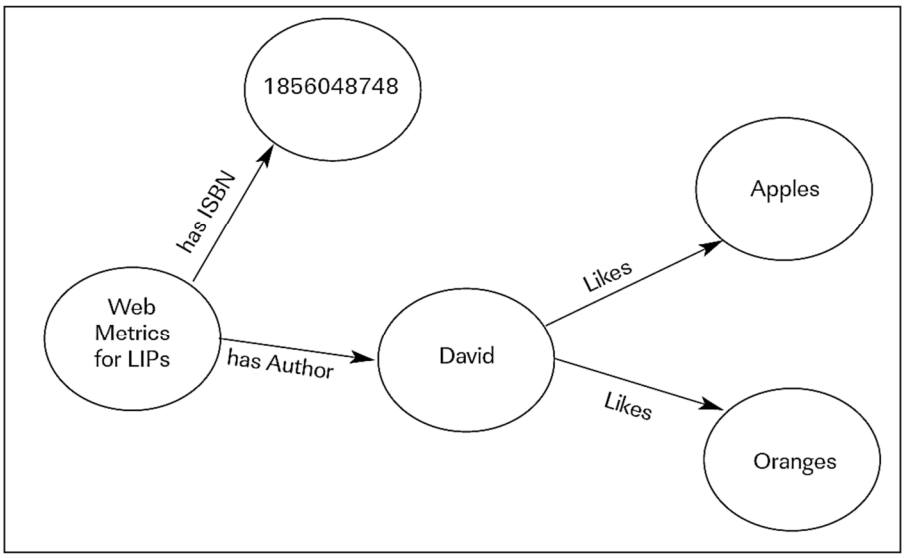
قطعه ساختمان اصلی وب معنایی سه‌گانه چارچوب توصیف منبع است، که بیانیه‌ای را در خصوص یک منبع به شکل سه گانه موضوع- مسند- هدف[2] را تشکیل می‌دهد. چنین سه‌گانه‌هایی خود به تنهایی داده‌های معنی‌دار را تشکیل می‌دهند، اما برای ایجاد یک نمودار بزرگ وبی نیز می‌توانند با یکدیگر ترکیب شوند. برای مثال، هر یک از سه‌گانه‌های زیر به تنهایی معنی‌دار است:

- دیوید [موضوع] - دوست دارد [مسند] - سیب [هدف]
- دیوید - دوست دارد - پرتغال.
- سنجه‌های وبی برایال.آی.پی‌ها - نویسنده دارد - دیوید.
- سنجه‌های وب برای ال.آی.پی ها - شابک دارد - ۱۸۵۶۰۴۸۷۴۸.

آن‌ها برای ساختن یک نمودار وب بزرگ‌تر نیز ترکیب می‌شوند (شکل ۸-۱را ملاحظه فرمایید).

رمزگذاری این سه‌گانه‌ها با توجه به شکل یکی از سریال‌سازی‌های چارچوب توصیف منبع (مانند ساختاردهی داده ها به فرمت شناخته شده ماشین خوان) موجب می‌شود که سایر رایانه‌ها بتوانند بیانیه‌ها را بخوانند.

1. Uniform resource identifiers (URIs)
2. subject-predicate-object



شکل ۸-۱. سه‌گانه‌های چارچوب توصیف منبع ترکیب شده به صورت یک نمودار داده

هرچند، این سه‌گانه‌ها دارای محدودیت‌هایی بوده، و قابل اصلاحند: شناسه‌گر همگانی منابع برای تمایز میان اشیاء دارای نام مشابه استفاده نشده است؛ موجودیت‌ها و نام آن ها ترکیب شده است؛ و واژگانی که مورد قبول همگان باشد استفاده نشده است.

شناسه‌گر همگانی منابع ردیف‌های منحصر به فرد نویسنده‌ها هستند که برای تمایز بین منابع در وب استفاده می‌شوند. شناخته‌شده‌ترین آن‌ها آدرس‌های اینترنتی[1] هستند که کاربران آن‌ها را در نوار آدرس مرورگر وب خود وارد می‌کنند تا صفحه‌ای مورد نظر را بارگیری کنند. در دنیای واقعی ممکن است چندین نفر یا چند شی دارای اسم مشابه باشند. برای مثال، اپل[2] ممکن است به میوه، شرکت رایانه‌ای، یا برچسب پیشینه اشاره داشته باشد. استفاده از نشانی‌های وبی از بروز هرگونه ابهامی جلوگیری می‌کند. در این مورد مفهوم اپل می‌تواند توسط یک شناسه‌گر همگانی منابع که هم‌اکنون برای نمایش میوه در دی‌بی‌پدیا[3] وجود دارد ارائه شود. هم‌چنین دی‌بی‌پدیا می‌تواند

1. Uniform Resource Identifier (URLs) 2. Apple 3. DBpedia, http://dbpedia.org/resource/Apple

یک شناسه‌گر همگانی منابع برای نمایش میوه پرتقال ارائه شود: http://dbpedia.org/resource/orange_(fruit). دی‌بی‌پدیا (Http://wiki.dbpedia.org) ساختار اطلاعات را از ویکی‌پدیا استخراج کرده و آن را به عنوان داده‌های پیوندی قابل دسترسی کرده است. شناسه‌گر همگانی منابع جایگزین نیز برای نمایش پیشینه های سیب در دسترس خواهند بود (http://dbpedia.org/resource/Apple_Records) و شرکت فناوری اپل (http://dbpedia.org/resource/Apple_Inc)، و همچنین رنگ نارنجی ((http://dbpedia.org/resource/Orange_(colour)) و شرکت تلفن ((http://dbpedia.org/resource/Orange_(telecommunication)). ماهیت متنوع محتوا در ویکی‌پدیا به معنای این است که دی‌بی‌پدیا برای طیف گسترده منابعی که می‌تواند مرجع افراد و سازمان‌ها باشند شناسه‌گرهای همگانی منابع دارد، زیرا نقش چشمگیری را در اتصال مجموعه داده‌های متنوعی در سراسر ابر داده‌های باز پیوندی ایفا می‌کند.

درحالی‌که دی‌بی‌پدیا برای طیف وسیعی از منابع شناسه‌گرهای همگانی منابع دارد نه همه چیز یا همه‌کس بلکه تنها آن‌هایی را در بر می‌گیرد که از ملاک های برجستگی ویکی‌پدیا برخوردارند: آدرس http://dbpedia.org/resource/David نویسنده این کتاب را نشان نمی‌دهد، بلکه شاه دیوید[1] کتاب مقدس را نشان می‌دهد که تصور می‌شود دارای رهنمون های ویکی‌پدیا برای برجستگی است. در عوض ممکن است من بخواهم یک پروفایل شخصی یک دوست از یک دوست[2] را در وبگاه شخصی ایجاد کنم تا نه تنها بتوانم جزئیات خودم را رمزگذاری کرده بلکه بتوانم خود را بشناسانم مانند www.davidstuart.co.uk/FOAF#me. بخش شناسه #me برای این افزوده شده است تا بتوان صفحه وبی و فردی که شناسه‌گر همگانی منابع آن را نمایش می‌دهد را از یکدیگر متمایز کرد.

اگرچه عنوان «سنجه‌های وب برای متخصصان کتابداری و اطلاع‌رسانی» برای یک کتاب می‌تواند عنوان کاملاً منحصر به فردی باشد، اما شاید (البته بسیار غیر

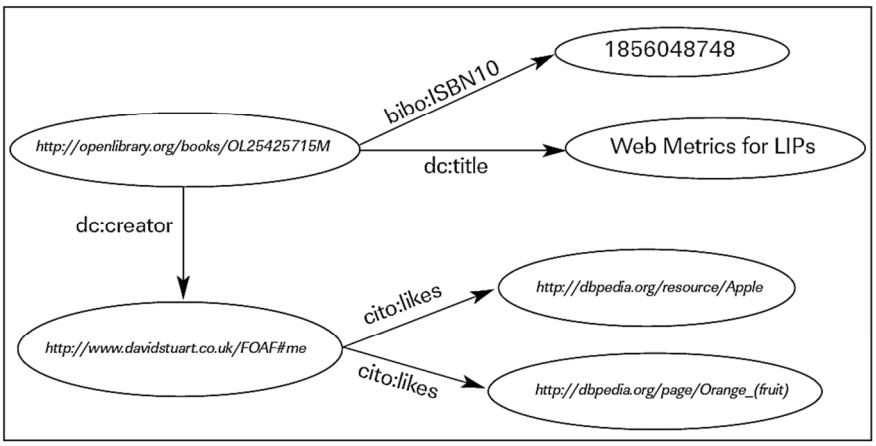
1. King David 2. Friend of a Friend (FOAF)

محتمل است) شخص دیگری هم بتواند کتاب مشابهی با همین نام تألیف کند. عنوان کتاب نیز با خود کتاب اشتباه گرفته می‌شود. از این رو ارائه کتاب با یک شناسه‌گر همگانی منابع مهم است. هدف اعلام شده کتابخانه باز[1] ایجاد یک صفحه وبی برای هر کتاب است و این اطلاعات را در یک با فرمت چارچوب توصیف منبع ارائه می‌کند زیرا می‌تواند از این طریق با دیگر سه‌گانه‌های آن پیوند برقرار کند. در داخل کتابخانه باز «سنجه‌های وبی برای متخصصان اطلاعات و کتابداری» توسط شناسه http://openlibrary.org/books/OL25425715M ارائه شده است. سپس عنوان ممکن است از کتاب تفکیک شود.

در نسخه اول، ارتباط بین موضوع‌ها و هدف‌های مختلف به انگلیسی ساده بیان شده است، و برای رمزگذاری انواع مشخصی از ارتباط طیف وسیعی از هستی‌شناسی‌ها[2] ایجاد شده است. برای نمونه، هستی‌شناسی FOAF (که در بالا اشاره مختصری به آن شده) یک هستی‌شناسی برای توصیف مردم و ارتباط بین آن‌ها است. انواع ارتباطات بیان شده بین موضوع ها و اهداف در این مثال غیر معمول نیستند، و تاکنون در هستی‌شناسی‌های موجود بیان شده‌اند: هسته دوبلین[3] از عناصر خالق[4] و عنوان[5] برخوردار است. هستی‌شناسی کتاب‌شناسی[6] ممکن است برای بیان شابک استفاده شود. درحالی‌که عبارت لایک[7] می‌تواند چندین معنا داشته باشد، معنی فیسبوکی لایکینگ[8] می‌تواند از طریق هستی‌شناسی چاپ بیان شود (CiTO؛ http://p نشانی وبی org/spar/cito.).

استفاده شناسه‌گرهای همگانی منابع و استفاده مجدد از هستی‌شناسی‌های موجود، موجب تولید یک نمودار وبی می‌شود که برای عوامل و وب‌پیماهای وب معنایی معنادار است (شکل ۸-۲).

1. Open Library (http://openlibrary.org)
2. ontologies
3. Dublin Core (http://dublincore.org)
4. Creator
5. title
6. Bibliographic Ontology (http://bibliontology.com)
7. like
8. liking



شکل ۸-۲.نمودار داده‌ها با استفاده ازنشانی‌های وبی و هستی‌شناسی‌های اشتراک گذاشته شده

بعد می‌توان این سه‌گانه‌ها را بر اساس یک تعدادی از سریال‌سازی‌های انتخابی گسترده در دسترس قرار داد. برای مثال، مجموعه‌ای از سه‌گانه‌های چارچوب توصیف منبع در خصوص یک کتاب می‌توانند به عنوان RDF/XML دردسترس قرار گیرند:

```
<rdf:RDF
   xmlns:rdf="www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:dc="http://purl.org/dc/elements/1.1/">
   <rdf:Description
rdf:about="http://openlibrary.org/books/OL25425715M">
      <dc:title>Web metrics for LIPs<dc:title>
      <dc:creator>David Stuart</dc:creator>
     </rdf:Description>
</rdf:RDF>
```

در فرمت لاک‌پشت[1] (TTL – زبان سه‌گانه کوتاه[2]) به این صورت:

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix rdf: <www.w3.org/1999/02/22-rdf-syntax-ns#>.
<http://openlibrary.org/books/OL25425715M> dc:creator
"David Stuart";
      dc:title "Web metrics for LIPs".
```

1. Turtle 2. Terse Triple Language

یا به عنوان N–سه‌گانه‌هایی، که از شناسه‌گرهای همگانی منابع فشرده استفاده نمی‌کنند (CURIE) و از این رو بسیار در دسترس خوانندگان انسانی هستند:

```
<http://openlibrary.org/books/OL25425715M> <http://purl.
org/dc/elements/1.1/title>"WebmetricsforLIPs".
<http://openlibrary.org/books/OL254257I5M> <http://purl.
org/dc/elements/1.1/creator>"DavidStuart".
```

سریال سازی N–سه‌گانه به صورت N–چهارگانه نیز گسترش یافته است، زیرا می‌توان اطلاعات سه‌گانه‌ها و بافت آن را جمع‌آوری نمود – جایی که سه‌گانه در آن پیدا شده است.

چنین سریال‌سازی‌هایی همانند درخواست ویژه‌ای برای ساخت یک ذخیره سه‌گانه می‌تواند روی این عنوان ایجاد شود (یک پایگاه اطلاعات طراحی شده مخصوص ذخیره سه‌گانه‌ها) یا در چارچوب توصیف منبع ممکن است در یک فایل متنی که در مکان خاصی ذخیره شده رمزگذاری شود.

داده‌های پیوندی به عنوان وب معنایی با شکلی منسجم توصیف می‌شود و انگیزه مضاعفی به وب معنایی داده است. پروژه‌های متعدد وب معنایی در سال‌های اخیر انجام شده است. البته لازم نیست که یک وب معنایی دیگر برای ایجاد پایگاه اطلاعاتی بزرگ یا محتوا یا رمزگذاری فایل‌های وب معنایی مجزا وجود داشته باشد. در واقع احتمال گنجاندن یک وب‌معنایی مقبول به صورت ضمیمه در وب فعلی بیش‌تر است.

معناهای درون کاشتی[1]

رویکردهای متفاوتی برای ادغام محتوای معنایی با اچ.تی.ام.ال[2] موجود وجود دارد، که در اینجا مختصراً به سه مورد از مهم‌ترین این رویکردها می‌پردازیم: RDFa، ریزشکل‌ها و ریزداده‌ها. این‌ها تنها تلاش‌های انجام شده برای درون کاشت[3] اطلاعات معنایی در درون صفحه‌های وبی نیستند. برای نمونه، eRFD روش

1. Embedded 2. HTML 3. embed

جـایگزینی (در حـال حاضـر منسـوخ شـده) بـرای RDF درون کاشـتی در درون اچ.تی.ام.ال است، درحالی‌که موضـوعات بـافتی در محـدوده‌هـا[1] روشـی بـرای درون کاشـت فـراداده‌هـای کتابشـناختی در درون اچ.تـی.ام.ال اسـت کـه از درون جامعـه کتابداری ظهور کرد. هرچند، ریزشکل‌ها و ریزداده‌ها و RDFa مهم‌تـرین آنهـا هسـتند: RDFa به دلیل روابط نزدیک با داده های پیوندی؛ ریزشکل ها به دلیل انتخـاب زیـاد؛ و ریزداده به دلیل اینکه نه تنها در ماهیـت html5 اسـت، بلکـه مـورد توجـه موتورهـای جستجو اصلی است.

RDFa

RDFa ابزارهای رمزگذاری RDF کامـل در درون صـفحه‌هـای وبـی اچ.تـی.ام.ال است. اساساً این ابزار به XHTML محـدود مـی‌شـد، هـر چنـد آخـرین نسـخه بـرای اچ.تـی.ام.ال (W3C، ۲۰۱۲a) مناسـب اسـت. بنظـر مـی‌رسـد ایجـاد محتـوا RDFa کاربرپسندانه‌تر از ایجاد RDF استاندارد نیست:

```
<div xmlns:dc="http://purl.org/dc/elements/1.1/"
  about=" http://openlibrary.org/books/OL25425715M">
  <span property="dc:title">Web metrics for LIPs</span>
  <span property="dc:creator">David Stuart</span>
</div>
```

تفاوت اصلی این است که الزامی برای ایجاد محتوای مجزا نبوده و سیسـتم‌هـای مدیریت محتوا می‌توانند برای افزایش نمونه خوانی به صفحه‌هـای وبـی معمـولی کـه تحت انتشار هستند ایجاد شوند.

ریزشکل‌ها و ریزداده‌ها

می‌توان استدلال کرد که سه‌گانه چارچوب توصیف منبع در بیش تر موارد بیش از حـد پیچیده است، و حتی خود چارچوب توصیف منبع کاربر پسند نیست. درک این نکتـه مهم است که با وجودی که داده های پیوندی در سال‌های اخیر بسیار مورد توجه بـوده

1. Context Objects in Spans (COinS)

است، تنها رویکرد ایجاد یک وب معنایی‌تر نیست. ریزشکل‌ها بسیار موفق بوده و مدعی بیش از ۷۰٪ داده‌های ساختاریافته در صفحه‌های وبی در استاندارد ریزشکل هستند (میکروفرمت‌ها، ۲۰۱۲). ریزشکل‌ها یک چهارچوب فراگیر طراحی شده نیستند تا مانند چارچوب توصیف منبع هر احتمالی را ترکیب کنند، بلکه بیش‌تر می‌خواهند مسائل خاص را با استفاده از قابلیت‌های فعلی اچ.تی.ام.ال، بر اساس این اصل حل کنند که نخست برای مردم طراحی شده‌اند و سپس برای ماشین‌ها.

ریزشکل‌ها تعداد محدودی فرمت‌های پیش‌نویس و ثابت را برای برخی از محتواهایی که بیش‌تر استفاده می‌شوند را فراهم می‌کنند. برای مثال، تقویمh[1] برای رمزگذاری اطلاعات رویدادها، کارتh[2] برای رمزگذاری اطلاعات تماس، و نقدh[3] برای نقد و بررسی‌ها و رتبه‌ها. یکی از اصول ریزشکل‌ها این است که برای ترغیب روزآمدسازی داده‌ها باید تمامی داده‌های رمزگذاری شده در یک صفحه وب قابل رویت باشند، و از تلاش مردم برای دستکاری داده‌های ساختار یافته برای ترغیب رفت و آمد بیش‌تر جلوگیری می‌کند. با تمرکز بر تعداد اندکی از استانداردها، ریزشکل‌ها دستاورد کم‌تر نادرست[4] یک وب معنایی را عرضه می‌کنند (سینگر[5]، ۲۰۰۹).

درحالی‌که RDFa قادر به ارائه کاملی از وب معنایی است، و امروزه اکثریت محتوای معنایی وب را ریزشکل‌ها ارائه می‌کنند، فرمت ریزداده‌ها به سرعت در حال ظهور است. ریزداده‌ها نسبت به HTML5 بومی هستند، بعلاوه در میان RDFa و ریزشکل‌ها ممکن است نادرست بنظر بیایند. هدف آن سهولت دسترسی ریزشکل‌ها از طریق استفاده از جفت‌های ارزش– نام است، در حالی که به تعداد کم واژگانی که مورد وفاق عمومی است محدود نمی‌شود.

پیامدهای وب داده‌ها برای سنجه‌های وبی

دقیقاً مشخص نیست که کدام نسخه وب داده‌ها غالب خواهد شد: ادامه انبار داده‌ها یا یک وب معنایی منسجم‌تر: یک وب معنایی درون کاشتی در اچ.تی.ام.ال یا وب

1. hCalendar 2. hCard 3. hReview 4. low lying fruit 5. Singer

داده ای که در کنار وب اسناد فعلی وجود داشته باشد؛ یعنی سه‌گانه چارچوب توصیف منبع یا جفت‌های نام- هدف. وب داده حتی در محدودترین شکل های خود برای راه‌اندازی سنجه‌های وبی با دسترسی بیش‌تر به داده‌های عمومی یک امتیاز می دهد تا این که بتواند با مجموعه سایر داده‌ها مقایسه و تحلیل شود. البته، به احتمال قوی بازهم تلاش چشمگیر پژوهشگر لازمه استفاده از داده‌های چنین اسنادی است. پیداست که آهسته به سمت یک وب معنایی هرچه بیش تر پیش می‌رویم، و هر شکلی به خود بگیرد پیامدهای مهمی متوجه توسعه سنجه‌های وبی جدید است.

فقدان استانداردهای مورد قبول همگان به این معنا است که تاکنون بیش‌تر بررسی‌های سنجه‌های وبی یا بر داده های بخش کوچکی از وب از قبیل داده های یک وبگاه شبکه های اجتماعی خاص تمرکز داشته‌اند، یا برای استنتاج نتایج معنای تأثیر زیاد وبی یا رابطه‌های بیان شده پیوندهای وبی از طریق همبستگی‌ها و تحلیل‌های محتوا بوده است. در مقایسه یک وب معنایی بیش از پیش قادر به بررسی محتوای ساختار یافته در چندین وبگاه است، در شرایطی که حتی دلایل جایگذاری پیوند نیز می‌توانند به صورت بالقوه آشکار شوند. یک وب معنایی می‌تواند اطلاعات تمامی حوزه‌های سنجه‌های وبی را ارائه دهد.

استانداردسازی به روشی که داده ها ساختاربندی شوند بیش از پیش درحال گسترش است، همانطور که بسیاری از هستی‌شناسی‌ها، که می‌توانند به صورت «خصوصیات صریح رسمی از یک مفهوم‌سازی مشترک» تعریف شده باشند (گروبر[1]، ۱۹۹۳). کتابداران در مرکز توسعه هستی‌شناسی‌های وب معنایی بسیاری بوده‌اند، و بنابراین جای تعجب ندارد اگر دریابیم که داده های کتابشناختی یکی از حوزه‌هایی است که بیش ترین هستی‌شناسی را راه‌اندازی کرده است. علاوه بر هسته دوبلین، که مدت‌هاست برای توصیف طیف وسیعی از منابع وبی طراحی شده است، هستی‌شناسی‌های تخصصی برای مواد کتابشناختی (مانند هستی‌شناسی کتابشناختی[2])

1. Gruber 2. The Bibliographic Oncology(http://bibliontology.com)

و هستی‌شناسی برای تجمیع منابع (مانند پیشگام بایگانی‌های آزاد[1] – تعویض و استفاده مجدد از موضوع[2]) وجود دارند. حتی یک هستی‌شناسی برای متداول‌ترین بخش تحلیل واحد کتابشناختی وجود دارد: استناد. سی‌تو[3] امکان بیان طیف گسترده دلایلی که یک مقاله می‌تواند مورد استناد مقاله دیگری باشد را میسّر می‌کند. برای نمونه، «موافق با[4]»، «استناد می‌کند همانند مطالعه‌های پیشنهاد شده»، «اختلاف‌ها[5]». قطعاً استفاده گسترده از هستی‌شناسی سی‌تو امکان تحلیل‌های استنادی خالص تری را میسّر می‌کند، و برخی از انتقادهای ایراد شده قبلی را درباره تحلیل استنادها رد می‌کند. هستی‌شناسی سی‌تو در وبگاه یو لایک[6] ثبت شده است، و خدمات نشانه گذاری برای مقاله‌های دانشگاهی، این امکان را به کاربران می‌دهد که سی‌تو روابط بین مقاله‌ها را نشانه‌گذاری کند. وب معنایی امکان نقد و بحث مقاله‌هایی که در متون علمی منعکس نشده را فراهم می‌آورد. هم چنین اغلب اظهار شده که مجلات تمایل ندارند که انتشار مقاله ها را با مجلاتی که نقد همان مقاله ها در آن بوده را پیگیری کنند، یک وب معنایی امکان شناسایی ساده چنین گفتمانی را میسّر می‌سازد. یک وب معنایی امکان بهبود بازیابی اطلاعات منابع را نیز فراهم می‌کند، واقعیتی که از Schema.org در پشتیبانی چهار موتور جستجو اصلی (گوگل، بینگ، یاهو و یاندکس[7]) منعکس شده است، که یک واژگان نشان‌گذاری مشترک است برای بیان محتوای وبی با استفاده از ریزداده‌ها یا RDFa.

یک وب معنایی امکان بررسی‌های مفصّل‌تر را نیز در خصوص پردازش‌های علمی می‌دهد، بویژه با افزایش میزان فعالیت‌های علمی و مباحثه‌های جایگذاری شده برخط. حتی چیزی به سادگی مکان‌یابی و یادداشت‌برداری در خصوص منابع برخط می‌توانند از هستی‌شناسی حاشیه‌نویسی باز[8] استفاده کند، درحالی‌که در منتهای دیگر برخی رشته‌های علمی یک چارچوب هستی‌شناسی کامل را توسعه داده‌اند. درک و

1. Open Archives Initiative
2. Object Reuse and Exchange(www.openarchives.org/ore/)
3. CiTO
4. agrees with
5. disputes
6. CiteULike (www.citeulike.org)
7. Yandex
8. Open Annotation Ontology((www.openannotation.org)

زمینه‌سازی پژوهش علاوه بر اینکه زمینه توسعه ابزارهای جدید را برای رفع نیازهـای پژوهشگران ایجاد می‌کنـد، امکـان سـاخت را بـر روی اثـر دیگـران فـراهم مـی‌کنـد. درحالیکه شاید پژوهشگران تمایل بـه اشـتراک‌گـذاری تمـام داده‌هـای خـود بـه‌ویـژه یاداشت‌هایشان را نداشته باشند، حتی حاشیه‌نویسی غیر شخصی متن می‌تواند موجـب ذخیره زمانی پژوهشگرانی شود که در جستجوی مرتبط‌ترین بخش‌های متن هستند.

به نظر می‌رسد هیچ حوزه‌ای برای تحقیق وب‌سنجی نیست که بتواند بـا انتخـاب گسترده هستی‌شناسی‌های وب معنایی ارتقا یابد. همانطور که تحلیل استنادی می‌توانـد از انتخاب گسترده یک هستی‌شناسی چون سی‌تو بهره ببرد، چنین هسـتی‌شناسـی (یـا یک نسخه تخصصی‌تر) برای تحلیل‌های ابرپیوند نیـز مـورد اسـتفاده قـرار مـی‌گیـرد. تحلیل شبکه‌های اجتماعی را می‌توان، بجای فرمت‌های خصوصـی غیـر متعـارف، بـا استفاده از هستی‌شناسی‌هایی از قبیل FOAF در چندین وبگاه انجام شود. در حـالی‌کـه یک هستی‌شناسی پیوندی گسترده می‌تواند بـرای تحلیـل وب‌سـنجی بـا مقیـاس بـالا ارزشمند باشد، به تحلیل‌های وبی کمک می‌کند، تـا روش سـریعی بـرای جمـع‌آوری بازخورد از وبگاه‌ها و خدمات فراهم شود.

با این وجود، در حال حاضر مشکل این است که هستی‌شناسی‌های بسیار انـدکی به صورت گسترده اجرا شده‌اند. بخشی از دلیـل ایـن موضـوع شـاید ایـن باشـد کـه شناسایی هستی‌شناسی‌های مناسب دشوار است، چیزی کـه بـه جـای انتخـاب علمـی هستی‌شناسی‌های مناسب به گمانه‌زنی متوسـل شـده اسـت (W3C، ۲۰۱۲a)؛ هرچنـد این امکـان هسـت کـه تعیـین کننـد آیـا یـک هسـتی‌شناسـی در زمـانی کـه ضـمیمه Schema.org بوده به اندازه کافی در جهت جریان اصلی قرار داشته است یـا خیـر، امـا این موضوع الزاماً موجب شناسایی مناسب‌ترین هستی‌شناسی بـرای یـک کـار خـاص نمی‌شود. احتمالاً الزامات افرادی که می‌خواهند صفحه‌های وبی خود را نشـانه‌گـذاری کنند تا دوستان آنها بتواننـد محتـوای بـرخط آن‌هـا را بیابنـد بـا الزامـات گـروه‌هـای پژوهشی که می‌خواهند شرح تفاصیل پژوهشگران و کارهای آنها را رمزگـذاری کننـد

فرق دارد؛ درحالی‌که هستی‌شناسی FOAF ممکن است برای یکی مناسب باشد، فرمت رایج اطلاعات پژوهشی اروپا[1] می‌تواند برای دیگری مناسب‌تر باشد. درک هستی‌شناسی‌های موجود در حوزه خاص و نحوه استفاده از آن‌ها می‌تواند بخش بسیار چشمگیری از کار کتابداران باشند. اگرچه در برخی حوزه‌ها وبگاه هایی هستند که کتابخانه‌های هستی‌شناسی بسیار توسعه یافته ای مانند بیو-پرتال[2] را برای جامعه پزشکی[3] ارائه می‌دهند، در بسیاری از زمینه‌ها شناسایی هستی‌شناسی‌های مناسب هم‌چنان به صورت فرایندی دشوار باقی مانده است.

بررسی وب داده‌های امروز

توسعه ابزارهای بررسی وب معنایی هنوز در مراحل بسیار ابتدایی است. بسیاری از این ابزارهای توسعه یافته از بخش دانشگاهی پدیدار شده‌اند، و اغلب با یک پروژه کوتاه مدت همراه بوده‌اند. با این وجود، قابلیت برخی از ابزارهای موجود بسیار بیش‌تر از قابلیت‌هایی است که در حال حاضر در موتورهای جستجوی وب قدیمی ارائه می‌شد، بعلاوه این مزیت را دارند که استفاده از کاربر خودکار را در نظر گرفته‌اند.

برای یک بررسی اسناد وبی، تناسب یک وسیله از وسیله‌ای دیگر شدیداً به ماهیت بررسی، و نوع داده‌ای که مورد توجه کتابداران است بستگی دارد. اگر پژوهشگر تنها به مجموعه داده های یک یا دو وبگاه علاقه داشته باشد، ممکن است اسپارکل[4] مناسب باشد، که امکان جستجوی ساده داده ها را میسّر کند. لازمه بررسی‌های بزرگتر می‌تواند استفاده از وب‌پیما یا خدمات گروه ثالث مانند سیندیک[5] باشد،که وب معنایی را نمایه‌سازی می‌کند.

اسپارکل

اسپارکل، که به صورت «sparkle» تلفظ می‌شود، یک زبان جستجو در چارچوب توصیف منبع است، که در آن کاربر الگوی نموداری که می‌خواهد بیابند را نمایش

1. Common European Research Information Format (CERIF)
2. BioPortal (http://bioportal.bioontology.org)
3. biomedical 4. SPARQL 5. Sindice

می‌دهد. الگوهای نموداری مانند سه‌گانه چارچوب توصیف منبع هستند، اما هر یک از بخش‌های سه‌گانه می‌توانند یک متغیر باشند، تا بعد ناشناخته سه‌گانه را نمایش دهند. برای نمونه، اگر کاربران بخواهند یک کتاب با شابک خاص را پیدا کنند، شاید تمایل داشته باشند که یک الگوی نموداری بیابند:

```
?book bibo:isbn13 "9781856047456"
```

در این مورد کتاب (book)? متغیری است که برای ارائه یک شناسه‌گر همگانی منابع کتاب استفاده شده است، که ناشناخته است، و "ISBN "9781856047456 یا شابک این کتاب است. مسند bibo:isbn13 مستلزم دانش هستی‌شناسی خاصی است، که در ذخیره[1] سه‌گانه استفاده می‌شوند، در این مورد هستی‌شناسی کتابشناختی قبلاً معرفی شد، و آن هم بخشی از مدل داده‌های نسخه داده‌های پیوندی سرگذشت‌نامه ملی بریتانیا[2] را تشکیل می‌دهد. با وارد کردن جستجوی مناسب در ویرایشگر اسپارکل کتابخانه بریتانیا می‌توان به جستجوی سرگذشت‌نامه ملی بریتانیا دست یافت: http://bnb.data.bl.uk?flint:

```
PREFIX bibo: <http://purl.org/ontology/bibo/>
SELECT ?book
WHERE {
   ?book bibo:isbn13 "9781856047456".
}
```

جستجوی بالا سه قسمت دارد: PREFIX امکان استفاده از اختصارها را در استفاده از شناسه‌گرهای همگانی منابع می‌دهد؛ SELECT به پایگاه داده‌ها اطلاع می‌دهد که کدام متغیرها باید بازگردانده شوند؛ و WHERE جزئیات الگویی که نمودار باید با آن منطبق باشد را نشان می‌دهد. از آنجایی که برای هر کتاب یک شابک مشخصی وجود دارد، وارد کردن جستجو موجب یافتن یک نتیجه book?

1. store 2. British National Biography

می‌شود، <http://bnb.data.bl.uk/id/resource/015855235>، که شناسه‌گر همگانی منابع کتاب است. برای اینکه جزئیات بیش‌تری بازیابی شود جستجوی گسترده‌تری مورد نیاز است:

```
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?authorname ?title

WHERE {
   ?book bibo:isbn13 "9781856047456";
         dct:creator ?authorURI;
         dct:title ?title.
   ?authorURI foaf:name ?authorname.
}
```

این جستجو نه فقط کتاب‌هایی را که دارای یک شابک مشخص هستند را تطبیق می‌دهد، بلکه هر جایی که کتاب یک عنوان یا یک خالق مشهوری دارد، را نیز مشخص می‌کند. اگر هر یک از این اطلاعات موجود نبود، آن‌وقت هیچ نتیجه بازگردانده نمی‌شود. هرچند، اسپارکل همچنین امکان تطبیق الگوی نمودار انتخابی[1] را نیز می‌دهد تا در صورت وجود اطلاعات اضافی بازگردانده شوند، اما در صورتی که داده‌ها موجود نبود از بازیابی نتایج ممانعت نمی‌کند.

هرچند برای جستجوهای بسیاری چندین نتیجه وجود دارد، اما در مثال بالا فقط یک نتیجه بدست آمد. برای مثال، امکان درخواست برای هر نام FOAF در یک ذخیره سه‌گانه با یک الگوی نمودار?subject foaf:name ?object یا حتی درخواست تمامی سه‌گانه‌ها بدون توجه به مسند وجود دارد:

1. optional

```
SELECT ?subject ?predicate ?object
WHERE {
  ?subject ?predicate ?object
}
LIMIT 50
```

در بسیاری از موارد ذخیره سه‌گانه حداکثر نتایج را ارائه می‌کند، یا در صورت گستردگی بیش از حد جستجو اخطار پایان می‌دهد. جایی که یک جستجوی گسترده استفاده می‌شود برای بررسی ماهیت واژه‌شناسی استفاده شده در ذخیره سه‌گانه، بهترین کار این است که تعداد نتایج را با افزودن یک حد[1] به انتهای جستجو محدود کرد.

همراه با علاقه روزافزون به داده های پیوندی در سال‌های اخیر، تعداد زیادی از مجموعه داده‌های پیوندی با نتایج پیشین اسپارکل منتشر می‌شوند تا کاربران امکان وارد کردن جستجوی خود از طریق یک مرورگر وب و نیز ارسال خودکار جستجوها را داشته باشند. علاوه بر سرگذشت‌نامه ملی بریتانیا کتابخانه بریتانیا (http://bnb.data.bl.uk/flint)، تعداد اندک دیگری از داده‌ها همراه با نتایج پیشین اسپارکل که می‌توانند مورد توجه کتابداران باشند ذخیره می‌شود شامل موارد زیر:

- دی‌بی‌پدیا[2]، که داده‌های ساختاریافته را از ویکی‌پدیا استخراج کرده، و از این رو داده‌ها شامل طیف گسترده‌ای از مباحث است.
- داده‌های مجموعه موزه بریتانیا (http://collection.britishmuseum.org/sparql)
- Data.gov (http://services.data.gov/sparql)
- Data.gov.uk (http://data.gov.uk/sparql)
- بستر داده‌های پیوندی طبیعی[3].

بسیاری از این وبگاه‌ها می توانند مجموعه‌های کمابیش بزرگی را با مالکیت کامل خودشان برای بررسی عرضه کنند: آیا در قرن گذشته سن انتشار اولین کتاب نویسندگان تغییر کرده است؟ زمینه‌های رشد مجموعه موزه بریتانیا چیست؟ هرچند

1. LIMIT 2. DBpedia (http://dppedia.org/sparql)
3. Nature Linked Data Platform (http://data.nature.com)

ارزش واقعی سنجه‌های وب معنایی، مانند سنجه‌های وبی بطور کلی، ناشی از تحلیل داده‌های چندین وبگاه است.

سیندیک

اگرچه موتورهای جستجوی وب معنایی و سایر ابزارهای نمایه‌سازی متعددی توسعه یافته‌اند، بیش‌تر آنها در جامعه دانشگاهی در رابطه با پروژه‌های نسبتاً کوتاه‌مدت توسعه یافته‌اند، و همیشه مشخص نیست که آنها تا چه حدی هم چنان درحال جمع‌آوری داده‌ها هستند یا آیا هنوز دارای کارآمدی کاملی هستند یا خیر. قاعدتاً سیندیک[1] گسترده‌ترین نمایه‌ساز وب معنایی فعال است که مرتباً رشد می‌یابد. سیندیک نه تنها چارچوب توصیف منبع، بلکه ریزشکل‌ها، ریزداده‌ها و RDFa را نیز نمایه‌سازی می‌کند. سیندیک در تابستان ۲۰۱۳ ادعا کرده است که ۷۰۸.۱۹ میلیون سند را نمایه‌سازی کرده است،که برابر با میلیاردها سه‌گانه است. مهم‌تر اینکه حاوی طیف وسیعی از واسط‌های کاربری برای جستجوی داده‌ها است. علاوه بر واسط‌های کاربری جستجوی ساده و پیشرفته، یک رابط برنامه کاربردی عرضه می‌کند و دارای یک نقطه دسترسی اسپارکل است، اخیراً یک وسیله تحلیلی برای ارائه اطلاعاتی در خصوص هستی‌شناسی‌های استفاده شده در دامنه‌های مشخص، و دامنه‌هایی که یک هستی‌شناسی در آن‌ها پرطرفدار است را ضمیمه کرده است.

همین‌طورکه بررسی‌های سنجه وبی پیشین از موتورهای جستجوی قدیمی استفاده می‌کنند، سیندیک همواره نمی‌تواند نتایج مورد انتظار، یا کارآمدی مورد نیاز را فراهم کند. برای مثال، استورات (۲۰۱۲) کاربرد هستی‌شناسی FOAF را در فضای وب دانشگاهی بریتانیا بررسی کرد. اگرچه از لحاظ نظری، این موضوع با ارسال چندین جستجو به نقطه پایانی اسپارکل امکان پذیر بوده است، در واقع نقطه انتهایی با زمان خروج ادامه یافت. از این رو جستجو برای کاربرد FOAF الزامی بود: مسند نام در دامنه uk.، و سپس پالایش نتایج بر اساس اینکه آیا نتایج شامل «ac » در نشانی

1. Sindice (http://sindice.com)

وبی نیز بوده یا خیر. این موضوع موجب عرضه مجموعه کوچکی خوبی از اسناد شده که بعدها می‌توان آنها را بارگیری کرده، و به صورت خودکار بررسی کرد که واقعاً از ac.uk. دوم یا دامنه سطح بالا هستند، و دیگر اینکه «ac» در جای دیگری از نشانی وبی نیامده باشد.

با این وجود، چنین سرویس‌هایی موجب دسترسی سریع به داده ها از چندین دامنه می‌شود، و کلاً بیش‌تر از داده ای که اغلب پژوهشگران می‌توانند برای خودشان جمع‌آوری کنند داده فراهم می‌کنند. برای مثال، جدول ۸-۱ جزئیات داده‌های ساختاریافته را در دامنه اصلی دانشگاه‌های گروه راسل[1] ارائه می‌کند همانگونه که در خدمات تحلیلی دیده شد.

جدول ۸-۱.محتوای ساختاریافته دانشگاه‌های گروه راسل که توسط سیندیک نمایه‌سازی شده (جولای ۲۰۱۳)

دانشگاه	نشانی وبی	صفحه‌های داده‌های ساختار یافته	تعداد کل سه‌گانه‌ها
دانشگاه بیرمنگام	www.birmingham.ac.uk/	۰	۰
دانشگاه بریستول	www.bristol.ac.uk/	۹	K 39/2
دانشگاه کمبریج	www.cam.ac.uk/	۵۰۵	K 61/7
دانشگاه کاردیف	www.cardiff.ac.uk/	۱	۱۰
دانشگاه دورهام	www.dur.ac.uk/	۶	۷۴
دانشگاه ادینبُرگ	www.ed.ac.uk/home	۶۸۵	K 24/14
دانشگاه اکسِتر	www.ex.ac.uk/	۰	۰
دانشگاه گلاسگو	www.gla.ac.uk/	۱۸	۱۲۳
امپریال کالج لندن	www3.imperial.ac.uk/	۰	۰
کینگز کالج لندن	www.kcl.ac.uk/	۶۲	۶۸۰
دانشگاه لیدز	www.leeds.ac.uk/	۲۸۶	K 26/2
دانشگاه لیورپول	www.liv.ac.uk/	۶	۱۷۳
مدرسه علوم اقتصادی و سیاسی لندن	www.lse.ac.uk/	۱۷۶	۸۴۸
دانشگاه منچستر	www.manchester.ac.uk/	۲۱۲	K 61/2
دانشگاه نیوکاسل	www.ncl.ac.uk/	۳۱۱	K 90/36
دانشگاه ناتینگهام	www.nottingham.ac.uk/	۱۶	۴۳۹

←

1. Russell Group (http://demo.sindice.net/dataset/)

→

دانشگاه	نشانی وبی	صفحه‌های داده‌های ساختار یافته	تعداد کل سه‌گانه‌ها
دانشگاه آکسفورد	www.ox.ac.uk/	K31/8	k 86/229
دانشگاه مری کویین لندن	www.qmul.ac.uk/	۰	۰
دانشگاه کوییز بلفاست	www.qub.ac.uk/	۰	۰
دانشگاه شفیلد	www.sheffield.ac.uk/	۰	۰
دانشگاه ساوسمپتون	www.soton.ac.uk/	K58/11	K 63/402
دانشگاه کالج لندن	www.ucl.ac.uk/	۱۰۴	K 90/7
دانشگاه وارویک	www2.warwick.ac.uk/	۳	۴۴۰
دانشگاه یورک	www.york.ac.uk/	K53/7	K 22/248

مثل روش رتبه‌بندی‌های دانشگاه‌ها و واسپارگاه هایWebometrics.info، ویژگی باز بودن را به عنوان خصوصیت کلیدی حضور وب سازمان در نظر می‌گیرند، و تعداد فایل‌های غنی (مانند pdf. و doc.) که گوگل اسکالر آنها را نمایه‌سازی کرده در رتبه‌بندی‌های خود وارد می‌کنند، میزان محتوای ساختار یافته نیز ممکن است خصوصیت بسیار قابل توجهی از حضور وبگاه باشد.

همانگونه که مشاهده می‌شود، تنوع زیادی در میزان محتوای ساختار یافته هر وبگاه هست. در یک طیف وبگاه‌های زیادی وجود دارند که هم چنان هیچ سه‌گانه‌ای ندارند، درحالی‌که در طیف دیگر دانشگاه ساوسمپتون بیش از ۴۰۰۰۰۰ سه‌گانه نمایه شده دارد. آنچه که در جمع‌آوری داده‌ها از طریق موتورهای جستجوی قدیمی اهمیت دارد، تمایز بین واقعیت و تصویر واقعیتی است که از طریق ابزارهای خاص عملیاتی شده.

اگرچه دقت بررسی ابزارهای سنجه وب معنایی به اندازه دقت ابزارهای سنجه وبی قدیمی نیست، محدودیت ابزارها را پیش از این می‌توان دید. برای نمونه، با اینکه سیندیک گزارش داده است که دانشگاه کوییز بلفاست هیچ صفحه با داده های ساختار یافته، و در نتیجه هیچ سه‌گانه‌ای ندارد، با این وجود گزارش می‌کند که اطلاعات متناقضی را نشان می‌دهد که پرطرفدارترین مقوله‌های استفاده شده در

qub.ac.uk مشخصات vCard هستند. هم چنین، اطلاعات طرفداری یک گروه در یک وبگاه خاص بنا به چشم‌انداز جستجو می تواند متفاوت باشد. برای مثال، در زمان نگارش، دامنه‌ای که از گروه چاپ الکترونیکی[1] بیش‌ترین استفاده را می‌کند دانشگاه ساوسمپتون با تعداد ۸/۹۰ هزار عدد واقعی[2] (مفهوم واقعی عدد) است، درحالی‌که پر استفاده‌ترین گروه در وبگاه دانشگاه ساوسمپتون گروه چاپ الکترونیکی با ۸/۷۱ هزار عدد واقعی است.

با این وجود، چنین ابزارهایی توان ارائه طیف وسیعی از شاخص‌های نیمه ارزیابانه را دارند. همانطور که با وب اسناد، حتی اگر برای وب‌پیمایی هر یک از داده‌ها و نیز دلایل اختلافات احتمالی در زمان جستجوی داده احتیاج به دانش بیش‌تری باشد، ممکن است پژوهشگران نیازمند جمع‌آوری داده برای خودشان باشند.

ال دی‌اسپایدر- یک وب‌پیما در چارچوب توصیف منبع

همان‌گونه که وب‌پیماهایی برای وب وجود دارند، برای وب داده‌ها نیز وب‌پیماهایی هستند. برخی صرفاً برای چارچوب توصیف منبع طراحی شده‌اند (مانند الدی‌اسپایدر[3])، درحالی‌که بقیه امکان استخراج طیف بسیار گسترده‌تری از داده‌ها را شامل RDFa، ریزشکل‌ها و ریزداده‌ها فراهم می‌کنند (مانند هرچیزی به سمت سه‌گانه‌ها[4] - وب‌پیمای پس وبگاه سیندیک).

از برخی لحاظ پیمایش وب داده‌ها بسیار ساده‌تر از پیمایش وب‌های قدیمی است. داده ها برای جمع‌آوری و تحلیل خودکار طراحی شده است، اگرچه این بدین معنا است که توسعه ابزارها به جای اینکه تنها برای کاربران عادی طراحی شده باشد در درجه نخست برای حفظ دانشمندان رایانه‌ای است، بررسی وب داده‌ها می‌تواند نیازمند جمع‌آوری تعدادی قطعات مختلف کد باشد، و واسط‌های کاربری نیز اغلب از نوع غیر نموداری هستند. با این وجود این ابزارها برای افرادی که دارای

1. EPrint((http://eprints.org/ontology/EPrint)
2. cardinality
3. LDSpider(http://code.google.com/p/ldspider)
4. Anything to Triples(http://any23.apache.org)

مهارت‌های فنی بیش‌تری هستند در دسترس است، بعلاوه برای جمع‌آوری سریع تعدادی ابزارهای متفاوت از میزان زیادی داده از وب معنایی نیز صبور هستند.

برای مثال، الدی‌اسپایدر ممکن است یا ضمیمه یک برنامه رایانه‌ای بزرگ باشد، یا پس از بارگیری آن از فرمان فوری[1] اجرا شده است. امکان ایجاد یک فهرست هسته[2] که باید پیمایش وب را از آن آغاز کرد، راهکارهای وب‌پیمایی مختلف (پهنای- بیش از همه[3] یا تعادل بار[4])، و دنبال کردن تنها مسندهای خاص را در شناسایی اسناد اضافی می‌دهد (مانند تنها دنبال کردن FOAF: رابطه‌ها را می‌داند). زمانی که فایل جار[5] الدی‌اسپایدر از وبگاه بارگیری شد، و با راهنمای فرمان فوری به مشابه فایل جار تغییر یافت، خط بعدی وب‌پیما با یک پیمایش پهنای- بیش از همه، و با برداشتن یک فهرست هسته از فایل seed.txt و خارج کردن نتایج به یک فایل متنی به نام breadth.nq در همان پوشه اجرا خواهد شد:

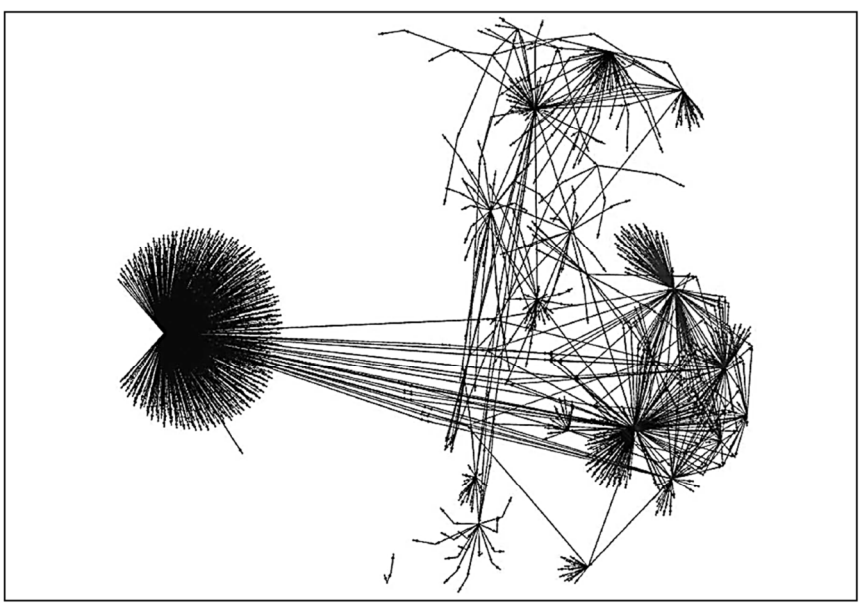
```
java -jar ldspider-1.1e.jar –b 50 10 –f http://xmlns.com/foaf/0.1/knows -s
seed.txt -o breadth.nq
```

استفاده از برنامه‌نویسی فایل جار موجب انتقال داده‌ها در تعدادی از سریال‌سازی‌ها می‌شود، بعلاوه فرمان فوری باعث خروج تنها RDF در سریال سازیZ-Quad می‌شود. هرچند این یکی از سریال سازی‌های RDF کاربر پسندانه‌تر است، چرا که هم سه‌گانه و هم نشانی وبی موجود در آن را نشان می‌دهد. N-Quads یکی از سریال‌سازی‌های جدیدتر هستند، بعلاوه تمامی نرم‌افزار تحلیلی نمی‌توانند آن‌ها را هدایت کنند. داده ها باید تغییر کند، چیزی که می‌توان آن را با استفاده از any23 یا از طریق بارگیری یک بخش اضافی نرم‌افزاری، مانند تبدیلRDF[6] انجام داد. زمانی که داده‌ها در سریال‌سازی RDF/XML است می‌توان با استفاده از طیف وسیعی از فایل‌ها تحلیل شوند. برای مثال، می‌تواند از طریق گراویتی RDF[7]

1. command prompt
2. seedlist
3. breadth-first
4. load balancing
5. jar file
6. RDFconvert((http://sourceforge.net/projects/rdfconvert)
7. RDF Gravity (http:// semweb.salzburgresearch.at/apps/rdfgravity)

بررسی شود، ابزار بصری‌سازی نموداری RDF، به نام توییِنکِل[1] را می‌توان برای بررسی‌های اسپارکل داده‌ها استفاده کرد، درحالی‌که پلاگ- این وب معنایی گِفی (http://wiki.gephi.org/ondex.php/SemanticWebImport) امکان جستجو و مشاهده داده‌ها را داده و در معرض ایجاد روش‌شناسی‌های تحلیل شبکه‌های اجتماعی قرار می‌گیرد.

شکل ۸-۳ شبکه‌های FOAF را بر اساس داده‌های جمع‌آوری شده از طریق الدی‌اسپایدر با استفاده از مشخصه‌های پیمایشی که در بالا ذکر شد و با استفاده از صفحه FOAF تیم برنرز-لی[2] به عنوان نقطه شروع نشان می‌دهد.



شکل ۸-۳. نمودار یک شبکه‌های FOAF که از طریق یک وب‌پیما الدی‌اسپایدر کشف شده

از آنجایی که وب‌پیما کل اسنادی را که توسط FOAF: نام، پیوند خورده را بارگیری می‌کند، و نه تنها سه‌گانه‌هایی که درون اسناد شامل مسند FOAF: نام هستند، وارد کردن یک جستجوی اسپارکل الزامی است چرا که فقط نمودار FOAF نمایش/تحلیل می‌شود:

1. Twinkle((www. ldodds.com/projects/twinkle)
2. Tim BernersLee's FOAF page (http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf)

```
construct {?x <http://xmlns.com/foaf/0.1/knows>  ?y}
where {?x <http://xmlns.com/foaf/0.1/knows> ?y}
```

آنگاه چنین نموداری می‌تواند در معرض انواع تحلیل‌های شبکه‌های اجتماعی چنانکه در فصل ششم بحث شد، مانند بررسی مفاهیمی از قبیل مرکزیت و خوشه‌بندی قرار گیرد.

نتیجه‌گیری

تنها زمان می‌تواند مشخص کند که فناوری‌های وب معنایی تا کجا استفاده خواهند شد، هرچند آن‌ها ظرفیت این را دارند که مزیتی برای بررسی‌های سنجه وبی محسوب شوند، انبارهای فعلی پژوهش را در هم شکسته، و امکان میزبانی یک تحلیل متقابل از وبگاه‌ها را با جزئیات فراهم می‌کنند. لازمه رخداد این اتفاق تلاش فراوان کتابداران است برای شناسایی هستی‌شناسی‌های موجود و توسعه هستی‌شناسی‌های جدید.

اگرچه فناوری‌ها آنقدر یکپارچه یا شهودی نیستند که پایه‌گذار تحلیل‌های شبکه‌های اجتماعی یا حتی تحلیل‌های پیوند وبی باشند، اما همچنان که در این فصل توضیح داده شد برای کسانی که تمایل دارند زمانی را برای یادگیری فناوری‌های جدید اختصاص دهند حجم عظیمی از داده‌ها وجود دارد.

۹

آینده سنجه‌های وبی و متخصصان کتابداری و اطلاع‌رسانی

هر گاه می‌توانید، بشمارید.
سر فرانسیس گالتون (۱۸۸۴)

در این گام از کتاب انتظار می‌رود که کتابداران در خصوص ظرفیت سنجه‌های وب در فعالیت‌های حرفه‌ای خود متقاعد شده باشند. حتی اگر آن‌ها آمادگی کاملی ندارند که برای یک خدمات سنجه وبی به کاربران خود تبلیغ کنند، حداقل به اهمیت پیگیری سنجه‌های خود پی برده و از اینکه بخواهند صرفاً به کامل‌ترین تعداد دسترسی داشته باشند منصرف شوند.

این فصل آخر به آینده بالقوه سنجه‌های وبی و رابطه آن با کتابداران می‌پردازد. همانگونه که در سراسر این کتاب تأکید شده است فناوری‌های وبی جدید و هم چنین ابزارهای موجود برای بررسی آن‌ها پی در پی پدیدار می‌شوند. تلاش برای پیش‌بینی ورود هر فناوری یا ابزار جدید خاصی، یا نفوذ آتی هر وبگاه به خصوصی، کار احمقانه‌ای است، اما منطقی است اگر انتظار داشته باشند که روندهای جاری ادامه یابد، از قبیل دیجیتالی کردن محتوای آنالوگ و افزایش حجم داده در زندگی روزمره. هم‌چنین می‌توان انتظار داشت که دسترسی بیش‌تر به این داده‌ها بیش از

پیش باز باشند، هرچند این موضوع به هیچ وجه قطعی نیست. درحالی‌که برخی پذیرای موقعیت‌های خدمات نوینی هستند که موجب اشتراک‌گذاری اطلاعات می‌شود، دیگران ظاهراً نگران از دست دادن حریم خصوصی هستند که قطعاً از دست خواهند داد. هر یک از این تغییرات پیامدهای احتمالی سنجه‌های وبی و کاربرد آن‌ها برای کتابداران را در پی خواهد داشت.

پیش از اینکه به سمت آینده برویم لازم است برای یک لحظه توقف کرده و درباره اینکه تا چه اندازه پیش آمده‌ایم، نه تنها در این کتاب، بلکه در سنجه‌های وبی تأمل کنیم.

تا کجا پیش آمده‌ایم؟

همانگونه که از یک کتاب با عنوان *سنجه‌های وبی برای متخصصان کتابداری و اطلاع‌رسانی* انتظار می‌رود، این کتاب به طیف وسیعی از ابزارها و منابع پرداخته است. این‌ها شامل: وب‌پیماها، وب‌پیماهای وب معنایی، وبگاه‌های شبکه‌های اجتماعی، سه‌گانه‌های RDF، رابط‌های برنامه‌نویسی، استخراج‌کننده های داده، اسکرپرهای داده، جمع‌کننده‌های داده ها و نرم‌افزار دیداری‌سازی نموداری. فهرست کامل فهرستی طولانی خواهد بود، و از آنجایی که حرکت سریع ما از یک فناوری به یک فناوری دیگر ضروری است، بسادگی می‌توان از مقیاس موضوعات مورد بحث چشم پوشی کرد. به هر حال، تشبیه توسعه وب‌پیماهای وب و خدماتی از قبیل گوگل ترندز به نمونه‌های اولیه ابزارهای علمی که اطلاعات دنیای فیزیکی را می‌دهند غیر منطقی نیست.

زمانی که *میکروگرفیا*[1] از رابرت هوکو[2] در سال ۱۶۶۵ منتشر شد او اولین کتاب پرفروش علمی دنیا را خلق کرد. میکروسکوپ اصلاح شده وی این امکان را برای او فراهم کرد که به ایجاد طرح‌های پیچیده اشیایی که در آن‌ها جزئیات ظریف‌تر به شدت نادیده گرفته شده‌اند بپردازد، هرچند این اشیا برای اشراف قرن هفدهم آشنا بودند.

1. *Micrographia* 2. Robert Hooke

طرح‌های هوک شامل کَک، شپش، نیش گزنه و حتی کریستال ادرار منجمد بودند، و او بخاطر علاقه بسیار فزاینده به میکروسکوپی اعتبار یافت. همانگونه که در این کتاب نشان داده شده است، ابزارهای سنجه وبی را نیز می‌توان استفاده کرد تا بینش‌های ناب از دنیایی که به روش‌های بسیاری برایمان آشنا است فراهم شود، اگرچه هنوز کتابی مانند میکروگرفیا نیامده است که بتواند تصورهای عمومی را به یک عرصه جلب کند. در واقع در دنیایی که در آن نوآوری مداوم طبیعی به نظر می‌رسد، ابزارها بندرت با یک نگاه لحظه‌ای مورد پذیرش قرار می‌گیرند چرا که جامعه انتظار دارد با ابزارهای درخشنده جدیدی که در شرف ظهور است ترغیب شوند. گوگل ترندز، خدمتی است که این امکان را فراهم می‌کند تا مردم در جریان افکار متراکم هزاران میلیون انسان در سراسر دنیا قرار گیرند، این امر نه تنها به صورت عادی مورد قبول است بلکه منتظر دریافت شکایت‌های مردم در خصوص عدم دسترسی خودکار به داده، یا اینکه عدم ارائه داده‌ها با جزئیات و در حد قابل قبول، باشد.

ابزارهای سنجه‌های وبی آتی بدون تردید بهتر خواهند بود، بعلاوه به نسبت تحلیل‌های فعلی امکان تحلیل طیف بسیار گسترده‌تری فراهم می‌شود، اما اینکه تا کجا پیش آمده‌ایم، و نیز ظرفیت ابزارهای موجود فعلی را نباید فراموش کنیم.

آینده سنجه‌های وبی

داده‌های بیش‌تر

هرگز چیزی در مورد آینده مگر استمرار تمدن قطعی نیست، اما این پیش‌بینی که وب به رشد خود ادامه می‌دهد اطمینان‌بخش است، بعلاوه به همراه آن ظرفیتی برای بررسی‌های سنجه وبی وجود دارد. در حال حاضر کمیّت داده ها پیوسته در حال رشد است که افراد و سازمان‌های بسیاری آن را به صورت برخط قابل دستیابی ساخته‌اند. برای افزایش داده های موجود دلایل زیادی وجود دارد که در اینجا به چهار مورد آن‌ها به اختصار می‌پردازیم: دیجیتالی‌سازی محتوای آنالوگ، ابری‌سازی، وب اشیا و توسعه ابزارهای علمی جدید.

می‌توان در این خصوص استدلال کرد که پروژه‌های دیجیتالی‌سازی بزرگ- مقیاس موجب جلب توجه بسیار زیادی می‌شوند در مقایسه با کمیّت اطلاعاتی که به صورت برخط قرار داده شده و نیز نسبت به اطلاعات مشابهی که واقعاً دیجیتالی شده‌اند. هرچند کتاب‌های گوگل بیش از ۳۰ میلیون نسخه را دیجیتالی‌سازی کرده اما تحت الشعاع مقدار محتوای رسانه‌های اجتماعی است که ایجاد شده و هم‌چنان می‌شود، درحالی‌که کتاب‌ها تنها بخش بسیار کوچکی از اطلاعاتی هستند که در کتابخانه‌ها، آرشیوها، موزه‌ها و نمایشگاه‌های سراسر جهان نگه‌داری می‌شوند. در بسیاری از موارد این اطلاعات هم چنان تنها به شکل فیزیکی وجود داشته و حتی نسخه‌های دیجیتالی فراداده ها در دسترس نیستند. با این وجود فرایند ممتد دیجیتالی‌سازی محتوای آنالوگ امکان انجام انواع کارهای برخط را تسهیل می‌کند.

تعداد بسیار زیادی از فعالیت‌ها در زندگی حرفه‌ای و شخصی ما به صورت برخط رخ می‌دهند، و به نظر می‌رسد همانطورکه خدمات تخصصی‌تر و زیرساخت های توسعه یافته‌ای برای انجام تعداد زیادی از کارها افزایش می یابد، تعداد فعالیت های حرفه ای که می‌توانند به صورت برخط انجام شوند نیز افزایش می‌یابد. همانگونه که مردم محتوا را ایجاد و منتشر کرده و این نوع خدمات و زیرساخت‌ها منابع اطلاعاتی جدیدی را عرضه می‌کنند، ظرفیتی برای طیف وسیعی از پارا داده[1] – زیر مجموعه فرا داده هایی که ثابت نیست- نیز وجود خواهد داشت از قبیل آمار کاربردی – گرچه به نظر می‌آید که احتمالاً پارا داده ها بسیار پیشرفته خواهند شد.

ایجاد داده‌ها به ایجاد محتوای عمدی توسط مردم محدود نمی‌شود، بلکه این کار با نزدیکی ما به وب اشیا به صورت خودکار می‌تواند ایجاد شود، و اشیا در دنیای واقعی متصل شده و اطلاعات را در وب به اشتراک می‌گذارند. اغلب مردم هم‌اکنون هر کجا که می‌روند گوشی تلفن همراه با خود حمل می‌کنند، که به این طریق امکان

1. paradata.

طیف گسترده‌ای از خدمات مبتنی بر موقعیت به وجود می‌آید، در حالی‌که سنجه‌های هوشمند[1] بیش‌تری نصب می‌شوند تا اطلاعات مصرف انرژی را به صورت خودکار به اشتراک بگذارند. این سنجه های هوشمند قادرند که طیف وسیعی از اطلاعات عادات مردم را به صورت بالقوه عرضه نمایند: اوقات فراغت خود را چگونه می‌گذرانند، هر چند وقت یک بار به سفر می‌روند، و چه زمانی مهمان دارند. هم‌اکنون طیف وسیعی از حسگرهای توانمند شده با وب وجود دارند، که اطلاعات کیفیت هوا را به اشتراک می‌گذارند (مانند http://airqualityedd.com)، به افراد یادآوری می‌کنند که به گیاه‌های خود آب بدهند (www.koubachi.com)، و همچنین برای نظارت بر فشار خون، ضربان قلب و سطح قند افراد حسگرهای زیستی وجود دارند. با افت قیمت‌های چنین محصولاتی می‌توان انتظار استفاده بیش‌تر آنها را داشت، و طیف محصولات موجود نیز افزایش می‌یابد. حسگرهای داخلی به هیچ وجه تنها ایجادکننده مقدار زیاد داده نیستند. با توسعه ابزارهای علمی هرچه پیشرفته‌تر مقدار زیادی داده جمع‌آوری می‌شوند. از برخورددهنده بزرگ هادرون[2] تا توالی‌های دی.ان.ای در حال حاضر دنیا لبریز از داده ها است. هنوز مشخص نیست چه مقدار از این داده‌ها در دسترس عموم خواهد بود.

باز بودن بیش‌تر

در سال‌های اخیر فشار عظیمی برای دسترسی عمومی به اطلاعات غیرشخصی بوجود آمده است. این موارد شامل به اشتراک‌گذاری بیش تر داده‌هایی است که دولت درباره نحوه خرج بودجه‌های عمومی، و از پژوهش بودجه عمومی، جمع‌آوری می‌کند. به اشتراک‌گذاری باز اغلب داده‌هایی که می‌توانند به اشتراک گذاشته شوند منوط به تصمیم‌های دولتی ها یا شرکت‌های بزرگ نیست بلکه متکی بر افراد است. درحالی‌که ممکن است شوراهای پژوهشی مقرر کنند که داده های تولید شده یک طرح تحقیقاتی در دسترس عموم قرار گیرد، بعید است که به عنوان یک پژوهشگر

1. smart meters 2. Large Hadron Collider

اصرار بر استفاده از منابع و ابزارهای برخط کنند، پارادادههای فعالیت آن‌ها نیز به اشتراک گذاشته می‌شود. هم چنین، درحالی‌که خانوادهها ممکن است تمایل داشته باشند که برای مشارکت در جمع‌آوری اطلاعات مصرف انرژی سنجه هوشمند نصب کنند، طرفداران حریم خصوصی نگران ترتیبی هستند که شرکت‌های انرژی این داده‌های بسیار حساس را با آن جمع‌آوری می‌کنند(که این امکان را برای آنها فراهم می‌کند که به ساخت دانش از رفتار افراد بپردازند) و پیشنهاد اشتراک‌گذاری عمومی چنین داده‌ای بعید است.

با این وجود، همانگونه که به کرات در بررسی داده‌های افراد به صورت برخط مشخص است، بخش های عظیمی از مردم مایلند که اطلاعات شخصی بسیاری را به اشتراک بگذارند. در واقع، برخی از افراد دچار اشتراک‌گذاری بیش از حد هستند زیرا در تمایز فضاهای عمومی و خصوصی سر در گم هستند (آگِگر[1]، ۲۰۱۲). چیزی که دیده شده است این است که اغلب مردم وقتی که نفعی برای آن‌ها داشته باشد مایلند اطلاعات شخصی را به اشترک بگذارند، مخصوصاً اگر فکر کنند که اشتراک‌گذاری عمومی اطلاعات به نفع آنها است.

درحالی‌که اشتراک‌گذاری عمومی اطلاعات پروفایل می تواند به یک کاربر کمک کند تا سرمایه اجتماعی جمع آوری کند، انگیزه‌های گروه‌های کاربری مختلف برای به اشتراک‌گذاری انواع متفاوت اطلاعات و جوامع مختلف کاربران احتمالاً متفاوت هستند. کلاً بعید است که اشتراک‌گذاری داده‌های سنجه هوشمند یک نفر برای سرمایه اجتماعی وی نفعی داشته باشد، اما اگر سایر افراد راه‌هایی را پیدا کنند که با به اشتراک‌گذاری اطلاعات می‌توانند پول ذخیره کنند، ممکن است شخصی بخواهد که این کار را انجام دهد. هم چنین، با وجودی‌که پژوهشگری ممکن است تمایل نداشته باشد که بیش‌تر محتوایی که به صورت برخط ایجاد کرده است را به اشتراک بگذارد، ممکن است اگر تشخیص بدهند که بر اساس این داده‌ها ابزارها و خدمات ارزشمندی می‌توانند ساخته شوند برای این کار تمایل بیش‌تری ابراز کند.

1. Agger

همگام با تصمیم‌گیری‌های به عمل آمده در دنیایی که همیشه وب در آن وجود داشته احتمال دارد که نگرش‌های فعلی به حریم خصوصی نیز تغییر کند، و همواره در آن اطلاعات خاصی که به صورت عمومی به اشتراک گذاشته می‌شود تغییر کند.

سنجه‌های وبی بیش‌تر

هرچند افزایش حجم داده‌های موجود برای تحلیل موجب عرضه ظرفیتی برای حجم وسیعی از مطالعه‌های سنجه‌های وبی جدید می شود، هم‌چنین می‌توان در این مورد استدلال کرد که ما در برهه‌ای هستیم که برای تحلیل باندازه کافی داده‌های رایگان داریم؛ ما اکنون به ابزارها، روش‌ها و افرادی که پرسش‌های صحیح را درخواست کنند نیاز داریم.

همانگونه که تاکنون دیده شده است، در طی سال‌ها برای جمع‌آوری داده‌های وبی ابزارهای زیادی توسعه یافته، اما برای توسعه ابزارهایی که بتوانند بهترین استفاده را از داده‌های موجود کنند هنوز باید کارهای زیادی انجام شود. برای اینکه بتوانیم از داده‌های دردسترس استفاده کنیم باید به توسعه ابزارهای بازتری بپردازیم، فقط نباید متکی به داده‌های موجودی که از طریق رابط‌های برنامه‌نویسی موتورهای جستجو و وبگاه‌های شبکه‌های اجتماعی بدست می‌آیند بود. بعلاوه به جای آنچه که در واقع مورد تقاضا باشد از کارآمدی که ارائه می‌شود استفاده کرد. در عصر داده‌های بزرگ برای نمایه‌سازی و تحلیل کمیّت‌های داده‌ای موجود، برای ابزارها و روش های مورد نیاز نیز چالش‌های جدیدی وجود دارد.

برای مواجه با هجوم انواع داده‌های جدیدی که می‌توانند کانون توجه بررسی‌ها شوند رویکردهای جدیدی نیز مورد نیاز هستند. برای مثال، همزمان با امکان دسترسی به مقدار عظیمی کد این امکان فراهم شود که تحلیل خودکار کد بتواند به شناسایی کدی که خوب نوشته شده کمک کند، و هم چنین برای به اشتراک‌گذاری کد اطلاعات یک رویکرد سرهم‌بندی را بدهد. شناسایی کد کیفیت به این سادگی نیست، اما در مورد دیکسترا[1] (۱۹۶۸) برخلاف بیانیه گو تو[2] در ساختار برنامه‌نویسی اولیه، برای تشخیص کیفیت کد ممکن است قوانین سختی تعبیه شود.

1. Dijkstra 2. GO TO

هر چند، احتمالاً بزرگ‌ترین تغییر، می‌تواند رشد چشمگیر تحلیل سنجه‌ها و داده ها باشد. در *نقد و بررسی کسب و کار هاروارد*[1] دانشمندان داده به عنوان دارندگان جذاب‌ترین شغل در قرن بیست و یکم توصیف شده‌اند (دونپورت و پاتیل[2]، ۲۰۱۲)، و درحالی‌که علم داده‌ها موجب گردآوری مهارت‌های هک کردن، مهارت‌های آماری و تخصص دامنه در تنظیم‌های کسب و کار برای کسب بینش‌های بدیع می‌شود، بعلاوه دانشمند الکترونیکی[3] کار مشابهی را در حوزه علمی انجام می‌دهد، سنجه‌های وبی روی این مهارت‌ها در مهم‌ترین منبع اطلاعاتی قرن بیست و یکم یعنی وب متمرکزند. در حال حاضر، علاقه وافری به ظرفیت سنجه‌ها در میزبانی زمینه‌هایی است که ممکن است قبلاً به احساسات درونی توجه بیش‌تری داشته‌اند، مشهورترین آن‌ها برای پرفروش‌ترین‌های *توپ پول*[4] (لوییس[5]، ۲۰۰۴) و *بازار فوتبال*[6] (کوپر و سیمنسکی[7]، ۲۰۱۲) برای دنیاهای بیس بال و فوتبال بازگو شده است.

آینده سنجه‌های وب و متخصصان کتابداری و اطلاع‌رسانی

در سراسر، این کتاب سعی شد که به کاربرد سنجه‌های وبی حوزه‌های فراوانی که در بسیاری از موقعیت‌ها می‌تواند مورد توجه کتابداران باشد پرداخته شود. پرسشی که به صورت طبیعی از این موضوع برمی‌آید این است که آیا باید انتظار داشت که کتابداران کاربرد سنجه‌های وبی را در نقش‌های فعلی خود در نظر بگیرند یا در صورت لزوم پست‌های مشخصی ایجاد کرده شوند.

مانند بسیاری از حوزه‌های تخصصی حرفه کتابداری، به نظر می‌آید که پاسخ این باشد که هرچند آشنایی گذرا با اساس سنجه‌های وبی برای تمام افراد ارزشمند است، اما برخورداری از یک حد مهارتی خاص نیز لازم است. این امر برای سایر حوزه‌های حرفه‌ای اطلاعات مانند کتاب‌سنجی، فهرست‌نویسی و رده‌بندی نیز صادق است. هرچند برای افرادی که در حرفه کتابداری هستند بهره مندی از شناخت کلی از

1. *Harvard Business Review*
2. Davenport and Patil
3. e-scientist
4. *Moneyball*
5. Lewis
6. *Soccernomics*
7. Kuper and Szymanski

برخی ابزارهای موجود و پژوهشی که می‌توان انجام داد ارزشمند است، اما در تمامی آن‌ها ضرورتی ندارد که برای جمع‌آوری و دستکاری داده‌ها از سطح بالای تخصصی برخوردار باشند، یا برای گرفتن استنتاج آزمون‌های آماری مورد نیاز را بدانند.

آمادگی حضور بیش از یک نفر متخصص حتی در تعدادی از حوزه های حیطه سنجه‌های وبی ارزشمند است؛ یک گزارش تحلیلی وب در خصوص تأثیر حضور رسانه های اجتماعی کتابخانه‌ای مستلزم مجموعه مهارت‌هایی است در مقایسه با بررسی وب معنایی هستی‌شناسی‌هایی که در یک زمینه خاص استفاده می‌شوند. در هر صورت نیاز به حضور متخصصان حوزه‌های خاص نه تنها منوط به اندازه کتابخانه است، بلکه به آینده حوزه‌های متفاوت سنجه‌های وبی نیز بستگی دارد.

تحلیل‌های وبی

تحلیل‌های وبی با سایر حوزه‌های سنجه‌های وبی مورد بحث این کتاب فرق دارد چرا که به جای تحلیل خدماتی که کتابداران می توانند به کاربران کتابخانه ارائه کنند، بر تأثیر خود کتابخانه می پردازند. تحقق شاخص‌های عملکرد در درون یک کتابخانه موضوع جدیدی نیست، و هم‌اکنون استانداردهای تأیید شده بین المللی از قبیل ISO 11620:2008 اطلاعات و مستندسازی[1] – شاخص‌های عملکرد کتابخانه[2] وجود دارند. با این وجود، زمانی که پای محتوای تحت وب در میان است اهمیت بسیار پیدا می‌کند زیرا در محیطی که به سرعت تغییر می‌کند آگاهی از بهترین روش استفاده از خدمات الزامی است، چراکه ممکن است یک خدمت برای کل کتابخانه مشکل آفرین شود.

ماهیت دقیق تحلیل‌های وبی قطعاً ظهور وبگاه‌ها و خدمات جدید، توسعه فناوری‌های طراحی شده برای ردگیری رفتار کاربر، و هم چنین تغییرات قوانین حریم شخصی اطلاعات را منعکس می‌کند. با این وجود این احتمال وجود دارد که اعتبار تحلیل‌های وبی منجر به شناسایی گسترده‌تری شده، و وضعیت تحلیل های وبی در سازمان‌های بزرگ مرکزی‌سازی شده یا به شرکت‌های اختصاصی تحلیل وب

1. Information and Documantation 2. Library Performance Indicators

برون‌سپاری شوند. با این وجود، لازم است که حداقل کتابداران علاقه خاص به سنجه‌های خود را حفظ کنند. اهداف و غایت رسانه‌های اجتماعی یا محتوای وب به این سادگی نمی‌تواند منجر به این شود که دریافت بیش‌ترین بازدیدها یا بیش‌ترین دنبال‌کننده‌ها کاهش پیدا کند، و این امر فقط در صورت شناخت بافت محتوایی خاصی است که سنجه‌ها واقعا معنی‌دار می‌شوند.

یکی از مهم‌ترین جنبه‌های تحلیل وبی این است که ابزار ارزشمندی محسوب می‌شوند برای ترویج گسترده‌تر سنجه‌ها. اگرچه اطلاع‌سنجی همواره این ظرفیت را داشته است که به بخش مفیدی از کار یک کتابدار تبدیل شود، اما از روش‌شناسی‌ها در عمل به صورت گسترده استفاده نشده است. تحلیل‌های وبی می‌توانند به کتابداران و کاربران کتابخانه ظرفیت سنجه‌ها را نشان دهند.

کتاب‌سنجی، علم‌سنجی و دگرسنجه‌ها

حوزه سنجه‌های وبی که بیش ترین هم‌سویی را با کارهای فعلی کتابداران دارند آن‌هایی هستند که به کتاب‌سنجی، علم‌سنجی و دگرسنجه‌ها وابسته هستند. از آنجایی که آثار موجود تکامل می‌یابند و گونه‌های جدید انتشاراتی ظاهر می‌شوند، بحث درباره کتاب‌سنجی یا علم‌سنجی خارج از طیف گسترده سایر سنجه‌ها بسیار بی‌معنی خواهد بود.

با انتشار بیش تر داده ها و برنامه‌های رایانه‌ای توسط پژوهشگرها، و ایجاد جریان ثابتی از خود–انتشاری‌ها[1] از طریق رسانه های اجتماعی، تمرکز بر یک نوع خروجی بسیار منسوخ بنظر می‌آید. از آنجا که گستره وسیع‌تر خروجی‌ها به عنوان بخش مهمی از روند آموزشی شناخته شده‌اند، به احتمال زیاد کتابداران در جذب داده‌ها نقش محوری‌تری دارند. درحالی‌که تمرکز علم‌سنجی‌های ارزیابانه و ارتباطی بر محاسبه‌های مبتنی بر استنادهایی است که آن‌ها را می‌توان با یکی از دو یا سه ابزار اندازه گرفت، از آنجا که یک دسته متنوع تری از راه های اطلاعاتی مهم تلقی می‌شوند

1. self-publications

طیف منابع وسیع‌تری برای جمع‌آوری داده‌ها لازم است، خواه بـرای اهـداف ارزیـابی باشد، خواه برای مشخص کردن منابع، یا فقط برای ترسیم یک محیط پـژوهش، بـرای شناخت محیط پژوهشی جدید پژوهش‌گران نیاز به کمک دارند. آن‌ها باید تغییـر نـوع محتوایی را که پژوهشگران در دسترس قرار داده‌انـد را درک کـرده و مقـدار کارهـای زیادی که به صورت برخط انجام می‌شود را بشناسند. این تغییر در شـیوه‌هـای کـاری پژوهشگران این امکان برای دانشمندان داده ها فراهم می‌کند که کارهـای آنهـا در حـد ارزیابانه مناسب‌تری بررسـی شـود: پژوهشـگران چگونـه بـا منـابع تعامـل دارنـد؟ از شیوه‌های کاری نویسندگان پر استناد و آن‌هایی که کم‌تر مورد اسـتناد قـرار مـی‌گیرنـد چه چیزی می‌توانیم بیاموزیم؟ قابلیت دسترسی داده ها نیز پرسش‌هـای جدیـدی را در این خصوص ایجاد می‌کند که پژوهشگران چگونه پـذیرای تیلوریزیشـن[1] شـیوه‌هـای کاری خود خواهند بود. فعالیت‌های پژوهشگران مانند شیوه‌های کـارگران کارخانـه‌ای قرن نوزدهم به طور نظام مندی خرد و تحلیل می‌شوند. همانگونـه کـه در اقتصـادهای رفتاری نشان داده شده است، زمانی که مردم می‌دانند کار آن‌هـا تحـت نظـارت اسـت رفتـار خـود را تغییـر مـی‌دهنـد (آریلـی و همکـارانش[2]، ۲۰۰۹)، بعـلاوه زمـانی کـه پژوهشگران متوجه نظارت شوند که کار آنهـا تحـت نظـارت اسـت، رفتـار آن‌هـا بـه روش‌های غیر مترقبه‌ای می‌تواند تغییر کنـد. ایـن درحـالی اسـت کـه ایـن اطلاعـات خاص ممکن است تا آن حد کامل نباشد.

از نقطه نظر کاربردی، به دلیل سرعت بالای تغییرات توسعه سنجه‌هایی که مـورد توافق همگان هستند عواقب قابل توجهی وجود دارد. صدها سال است که مقالـه‌هـای مجلات منتشر می‌شوند، با ایـن وجـود هنـوز هـم هـیچ تـوافقی بـر سـر مهـم‌تـرین سنجه‌های ارزیابی تأثیر وجود ندارد. در بازه یک دوره ده سـاله مـا مـی‌تـوانیم منتظـر تفوق خدمات جدید وبی برای یک جامعه باشیم تا این حد که مقالـه‌هـای مجـلات را در بررسی‌های علم‌سنجی و کتاب‌سنجی وارد نکنیم. این امر تنها برای آن‌هـایی اسـت که به سرعت از دور توجه خارج می‌شوند. آنچنان که ویرایشگری اخیـراً بـه موضـوع

1. Taylorization. 2. Ariely et al.

دگرسنجه‌ها در *مواد طبیعی*[1] اشاره می‌کند: «هر سنجه جدیدی که امروز معرفی شود شاید فرصتی برای اعتباریابی و کسب مقبولیت نداشته باشد» (مواد طبیعی، ۲۰۱۲، ۹۰). با این‌حال آنچه بسیار اهمیت دارد این است که خود را تنها به آن سنجه‌هایی که اعتباریابی شده‌اند محدود نکنیم، چرا که درمی‌یابیم که ما با تغییرات فناوری به سرعت پیش می‌رویم. به‌جای اشخاصی که فقط به سنجه‌های اعتباریابی شده دیگران اکتفا می‌کنند، کتابدارانی مورد نیاز مبرمند که بتوانند با توسعه سنجه‌های جدید تعامل موشکافانه‌ای داشته باشند.

وب‌سنجی

تغییرات اساسی در نحوه انتشار، و بازیابی اطلاعات در دو دهه گذشته بی شک منجر به تغییر نقش قدیمی و سنتی کتابداران شده است، و با انتشار منابع برخط بیش‌تر هم چنان به این کار ادامه خواهد داد، اشتراک‌ها با هم ادغام شده‌اند، کنسرسیوم‌ها معامله‌ها را در هم شکسته‌اند، و ابزارهای برخط جدیدی توسعه داده شده‌اند که به کشف محتوا کمک می‌کنند. حرفه اطلاعات حرفه‌ای است که در عین حالی‌که کتابداران به ارائه خدمات قدیمی کتابخانه مشغول هستند، متخصصان اطلاعات ضرورتاً باید برای ثبت جایگاه واقعی خود در محیط اطلاعاتی جدید تلاش کنند که از مهارت‌های فعلی آنها بهترین استفاده شود. این ایده که کتابداران موجب ارتقاء زنجیره پژوهش می‌شوند و بیش از پیش در کنار پژوهشگران قرار دارند ایده‌ای است که موجب جلب علاقه وافری شده است، و از آنجا که علم الکترونیکی و علم داده‌ها از همیشه مهم‌تر هستند به نظر می‌رسد که زمینه‌ای برای رشد باشد، و وب‌سنجی‌ها لزوماً بخشی از آن است.

همانگونه که تاکنون نیز ذکر شده است لازمه نقش دانشمند داده ها برخورداری از سه مجموعه مهارتی خاص می‌باشد: دانش دامنه، مهارت‌های رایانه‌ای و روش‌های آماری. کتابداران برای رسیدن به این سه مهارت در موقعیت بسیار خوبی قرار دارند.

1. *Nature Materials*

دانش دامنه کلاً به عنوان یکی از قابلیت‌های مورد نیاز کتابداران مدّ نظر است، کتابداران معمولاً در یک تخصص خاص دارای مدرک هستند و در علوم کتابداری و اطلاع رسانی نیز منزلتی دارند. مهارت‌های رخنه کردن[1] دلالت بر قابلیت جمع‌آوری و دستکاری چندین منبع‌های اطلاعاتی موجود می‌کند. درحالی‌که کتابداران در مورد چارچوب‌های فشرده داده‌ای در مقیاس بزرگ، لازمه تحلیل‌های داده‌های بزرگ (مانند هادووپ[2])، تجربه کمی دارند، در مورد وب‌سنجی بیش‌تر پردازش داده‌های بزرگ توسط خدمات بیرونی در سمت سرور[3] اتفاق می‌افتد. برای مثال، هنگام بررسی یک شبکه از صفحه‌های وبی پیوند شده توسط استنادهای نشانی وبی بینگ با مشکل جمع‌آوری و ذخیره یک رونوشت از وب مواجه هستیم درحالی‌که پژوهشگر تنها باید با تعداد انگشت‌شمار صفحه‌های موجود سر و کار داشته باشند. بعلاوه شناسایی منابع داده ای مناسب و ابزارهای تحلیل این داده‌ها نیز به اندازه «مهارت‌های رخنه کردن» مهم هستند، یعنی مهارت‌های که با مهارت‌های کتابداران هم‌سویی نزدیکی دارند. به هر حال، در حال حاضر روش‌های آماری مورد نیاز برای متخصصان کتاب‌سنجی نقش مهمی را ایفا می‌کنند، اما این موضوع برای بسیاری از کسانی که در حرفه کتابداری هستند نقش چشمگیری نداشته است، بلکه همان‌گونه که در این کتاب نشان داده شده است و همانگونه که سیلور (۲۰۱۲) در کار خود درباره علم داده‌ها تاکید کرده است، الزامی برای اعمال روش‌های بیش از حد پیچیده آماری نیست.

علاوه بر مهارت‌های عملی، کتابداران از توانمندی‌های دیگر برخوردارند که در پژوهش‌های وب‌سنجی نقش چشمگیری ایفا می‌کنند. در طی سال‌های متمادی اخلاق اطلاعات و حقوق مالکیت معنوی بخشی از کار دانشمندان اطلاعات بوده، و به راحتی برای مقابله با مسائل جدید که ناشی از دسترسی عمومی به اطلاعات است استفاده می‌شود، به روش‌هایی که بدواً برنامه ریزی نشده تجمیع و دستکاری شدند (اوبولر، ولش و کروز[4]، ۲۰۱۲؛ ویلینکسون و تلوال، ۲۰۱۱). شناخت این موضوع که

1. Hacking skills
2. Hadoop
3. server-side
4. Oboler, Welsh and Cruz

یک عنصر انسانی در وب‌سنجی وجود دارد نیز بسیار حائز اهمیت است: به منظور مطالبه پرسش‌های صحیح و کسب اطمینان از اینکه مردم ظرفیت داده های موجود را درک می‌کنند، ضروری است که منابع داده‌های موجود با کارآمدترین شیوه مورد تفحص قرار گیرند. کتابداران علاوه بر اینکه متخصص راه‌حل یابی مسائل خاص هستند، هم‌چنین به درک کاربران کمک می‌کنند تا در واقع چیزی که به دنبال آن هستند را به صورت مصاحبه مرجع درخواست کنند.

هرچند نقش بالقوه روشنی برای کتابداران در وب‌سنجی‌ها، یا حوزه‌های مشترک علوم داده‌ها و علوم الکترونیکی وجود دارد، اما به روشنی مشخص نیست که آن سمت [1] در کجا قرار خواهد گرفت. نقش کتابداران و کتابخانه مانند یک محیط دستخوش دوره‌ای از تحول است. این امر می‌تواند ناشی از این باشد که هرچندکه خدمات اطلاعاتی معین با هم در یک جا جمع می‌شوند (مانند واسپارگاه‌های داده)، اما وب‌سنج‌ها و دانشمندان داده‌ای با بخش‌های خاصی کار می‌کنند چرا که آنها دانش دامنه و نیازهای کاربران را مورد کند و کاو قرار می‌دهند.

برداشتن اولین گام‌ها

این طور به نظر می‌رسد که قرار است سنجه‌های وبی نقش بسیار چشمگیری را برای کتابداران ایفا کنند، حداقل برای کسانی که در حرفه‌ای هستند که دنبال این موضوع می باشند. این امر قطعاً این سؤال را به ذهن می‌آورد که علیرغم وجود چنین طیف گسترده از سنجه‌هایی که می‌توانند مورد استفاده قرار گیرند، کتابداران باید از کجا کار خود را شروع کنند؟

کتابداران در هنگام مواجهه با چنین محیط پیچیده‌ای نباید در حلقه تنوع گسترده مواردی که می‌توانند بررسی شوند باقی بمانند، بلکه باید بر یک مبنایی که به آسانی در دسترس باشد شروع کنند تا بتوانند بررسی‌های خود را با توجه به آن عملی سازند. به احتمال زیاد اساس سنجه‌های وبی تحلیل وبی خواهد بود، و بطور کلی

1. post

مناسب‌ترین مکان برای شروع حضور خود کتابدار یا مؤسسه آنها در شبکه‌های اجتماعی است. این کتاب به این منظور طراحی شده است که بجای تکیه بر یک کتابخانه یا تحلیل وبی خدمات اطلاعاتی بررسی طیف گسترده دادهای روی وب را ترغیب نماید، اما تصور اینکه چگونه نمودار یادگیری بررسی‌های سنجه وبی خاص می‌تواند دارای شیب بیش از حدی برای افراد باشد ساده است. به احتمال زیاد یک بررسی در خصوص اعتبار هستی‌شناسی‌های حوزه‌های پژوهشی خاصی بیش از تأثیر توییتر یک کتابدار مورد توجه است، اما از آنجا که بررسی هستی‌شناسی می‌تواند به طیف وسیعی از فناوری‌ها و توسعه روش‌شناسی‌ها و سنجه‌های جدیدی در این حوزه پژوهشی جدید نیاز داشته باشد، تحلیل یک حساب توییتر بر اساس مفاهیمی است که کتابداران آنها را می‌شناسند. امکان بررسی با سطوح بسیار پیچیده‌ای وجود دارد، برای نمونه:

۱. مقایسه تعداد دنبال‌کننده‌هایی که یک حساب کاربری دریافت می‌کند با تعداد دنبال‌کننده‌های مؤسسات مشابه.
۲. مقایسه تعداد دنبال‌کننده‌های دریافت شده با توجه به تعداد افرادی که دنبال می‌شوند و تعداد وضعیت روزآمدسازی شده با استفاده از تحلیل رگراسیون چندگانه.
۳. اجرای یک تحلیل محتوا از توییت‌های اخیر از حساب شخصی یک نفر و موفق‌ترین حساب‌های کاربری برای شناسایی تفاوت‌های بالقوه بین انواع وضعیت روزآمدسازی شده.
۴. تحلیل شبکه های ایجاد شده بین حساب‌های کاربری توییتر که بنیان بررسی‌ها را تشکیل می‌دهند.
۵. مقایسه شبکه‌های اجتماعی دوستان و دنبال‌کننده‌های دو حساب کاربری متفاوت. برای مثال، یک حساب کاربری موفق و یک حساب کاربری ناموفق، یا حساب‌های کاربری استفاده‌کننده از توییتر به صورت‌های مختلف.

برای مقایسه حساب‌های کاربری توییتر ممکن است دنبال‌کننده‌های توییتر سنجه معتبری نباشند، اما حتی این سنجه ساده احتمالاً بیش از چیزی است که کتابداران با استفاده از شیوه روشمند[1] تحلیل کرده‌اند(برعکس مقایسه گاه و بی‌گاه تعداد دنبال‌کننده‌های آن‌ها با تعداد دوستان یا همکارها به شیوه تصادفی).

در محاسبه سنجه‌ها برای هدف تحلیل وبی دلایل زیادی وجود ندارد، مگر اینکه آنها تفسیر شوند، و در صورت لزوم، برای آن‌ها اقدام شود. کتابداران چه بخواهند بر روی داده‌ای که به دست آورده‌اند اقدام کنند، و چند ماه بعد مسئولیت یک مطالعه پیگیری را بر عهده بگیرند، و یا اینکه بخواهند یافته‌ها را به صورت مستقیم به اشتراک بگذارند، در هر دو صورت باید بر اهمیت استفاده از اطلاعات در ساخت یک موضوع واقف باشند.

همانگونه که عنوان این بخش پایانی اشاره می‌کند، این‌ها تنها گام‌های نخست هستند. اگرچه برای سنجه‌های وبی داده‌ها و ابزارهای محدودی موجود هستند، اما محدودیت بزرگ‌تر تعداد و نوع پرسش‌هایی است که پرسیده می‌شوند، و کتابداران برای حل این مشکل پیشنهاداتی در اختیار دارند.

1. methodical

کتابنامه

Acerbi, A., Lampos, V., GarneW, P. and Bentley, R. A. (2013) The Expression of Emotions in 20th Century Books, PLOS ONE, 8 (3), e59030.

Achrekar, H., Gandhe, A., Lazarus, R., Ssu-Hsin, Y. and Benyuan, L. (2011) Predicting Flu Trends Using TwiWer Data. In IEEE InfoCom 2011 – IEE Conference on Computer Communications Worshop, 702-7.

Adamic, L. and Glance, N. (2005) The Political Blogosphere and the 2004 U.S. Election: divided they blog, hWp://nielsen-online.com/downloads/us/buzz/wp_PoliticalBlogosphere_Glance_2004.pdf.

Agger, B. (2012) Oversharing: presentations of self in the internet age, Routledge. Aguillo, I. (1997) Cybermetrics '97 (Jerusalem, Israel),

www.cindoc.csic.es/cybermetrics/cybermetrics97.html.

Aguillo, I. (2012) Is Google Scholar Useful for Bibliometrics? A webometric analysis,Scientometrics, 91 (3), 343–51.

Aksnes, D. W. and Rip, A. (2009) Researchers' Perceptions of Citations, Research Policy, 38 (6), 895-905.

Alexa (2013a) Frequently Asked Questions: how are Alexa's traffic rankings determined?, www.alexa.com/faqs/?p=134.

Alexa (2013b) Site Info: wikipedia.org, www.alexa.com/siteinfo/wikipedia.org. Almind, T. C. and Ingwersen, P. (1996) Informetric Analysis on the World Wide Web: a methodological approach to 'internetometrics', Centre for Informetric Studies, Royal School of Library and Information Science (CIS Report 2).

Almind, T. C. and Ingwersen, P. (1997) Informetric Analyses on the World Wide Web: methodological approaches to 'webometrics', Journal of Documentation,53 (4), 404–26.

Anderson, C. (2006) The Long Tail, Random House.

Angus, E., Thelwall, M. and Stuart, D. (2008) General PaWerns of Tag Usage Among University Groups in Flickr, Online Information Review, 32 (1), 89–101.

Ariely, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009) Large Stakes and Big Mistakes, Review of Economic Studies, 76, 451-69.

Arthur, C. (2013) Google Keep? It'll probably be with us until March 2017 – on average, Guardian,

www.guardian.co.uk/technology/2013/mar/22/google-keep-services-closed.

Banerjee, A. V. and Duflo, E. (2012) Poor Economics, Penguin.

Bar-Ilan, J. (2005) Expectations Versus Reality – search engine features needed for web research at mid 2005, Cybermetrics, 9 (paper 2), hWp://cybermetrics.cindoc.csic.es/articles/v9i1p2. html.

Bar-Ilan, J. (2012) JASIST@mendeley: altmetrics12, ACM Web Science Conference 2012Workshop, hWp://altmetrics.org/altmetrics12/bar-ilan/.

Bar-Ilan, J. and Azoulay, R. (2013) Map of Non-Profit Organization Websites in

Israel, Journal of the American Society for Information Science and Technology, 63 (6),1142–67.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H. and Terliesner, J. (2012) Beyond Citations: scholars' visibility on the social web, 17th International Conference on Science and Technology Indicators, Montreal, Canada, 5–8 Sept.

Batke, P. (2010) Google Books: Google Book Search and its critics, www.lulu.com/items/volume_68/8362000/8362807/5/print/pbedits.HC.4-26.pdf.

Beckmann, M. and von Wehrden, H. (2012) Where You Search Is What You Get: literature mining – Google Scholar versus Web of Science using a data set from a literature search in vegetation science, Journal of Vegetation Science, 23 (6), 1197–9.

Beel, J. and Gipp, B. (2009) Google Scholar's Ranking Algorithm: an introductory

overview. In Larsen, B. and Leta, J. (eds), Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), Rio de Janeiro, July, International Society for Scientometrics and Informetrics, 230–41.

Beel, J., Gipp, B. and Wilde, E. (2010) Academic Search Engine Optimization (ASEO): optimizing scholarly literature for Google Scholar and Co., www.sciplore.org/publications/2010-ASEO—preprint.pdf.

Behn, R. D. (2003) Why Measure Performance? Different purposes require different measures, Public Administrative Review, 63 (5), 586-606.

Benevenuto, F., Duarte, F., Rodrigues, T., Almeida, V., Almedia, J. and Ross, K. (2008) Understanding Video Interactions in YouTube. MM'08, October 26–31,Vancouver, British Columbia, Canada.

Bergman, E. M. L. (2012) Finding Citations to Social Work Literature: the relative benefits of using Web of Science, Scopus, or Google Scholar, Journal of Academic Librarianship, 38 (6), 370-9.

Berkowitz, D. (2009) 100 Ways to Measure Social Media, www.marketersstudio.com/2009/11/100-ways-to-measure-social-media-.html.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web, Scientific

American, May, 29-37.

Bertin, M. and Atanassova, I. (2012) Semantic Enrichment of Scientific Publications and Metadata: citation analysis through contextual and cognitive analysis, D-Lib Magazine, 18 (7/8), www.dlib.org/dlib/july12/bertin/07bertin. html.

Björneborn, L. and Ingwersen, P. (2004) Toward a Basic Framework for
Webometrics, Journal of the American Society for Information Science and Technology,55 (14), 1216–27.

Blood, R. (2000) Weblogs: a history and perspective, Rebecca's pocket, www.rebeccablood.net/essays/weblog_history. html.

Bode, K. (2012) Reading by Numbers: recalibrating the literary field, Anthem Press. Borgman, C. L. and Furner, J. (2002) Scholarly Communication and Bibliometrics. In Cronin, B. (ed.), Annual Review of Information Science and Technology, 36, 3-72.

Börner, K., Sanyal, S. and Vespignani, A. (2007) Network Science. In Cronin, B. (ed.), Annual Review of Information Science & Technology, 41, 537-607.

Bornmann, L., Mutz, R. and Daniel, H.-D. (2008) Are There BeWer Indices for
Evaluation Purposes than the h Index? A comparison of nine different variants of the h index using data from biomedicine, Journal of the American Society for Information Science and Technology, 59 (5), 830-7.

Bornoe, N. and Barkhuus, L. (2011) Privacy Management in a Connected World: students' perception of facebook privacy seWings, paper given at Workshop on Collaborative Privacy Practices in Social Media, part of the 2011 ACM Conference on Computer Supported Cooperative Work (CSCW '11), March 19–23, Hangzhou, China. Bossy, M. J. (1995) The Last of the Li،er: 'Netometrics', Solaris, 2,
hWp://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy. html.

Boudourides, M. A., Sigrist, B. and Alevizos, P. D. (1999) Webometrics and the Self-Organization of the European Information Society, draw report of the SOEIS project, hWp://hyperion.math.upatras.gr/webometrics/.

Boyd, D. M. and Ellison, N. B. (2007) Social Network Sites: definition, history, and scholarship, Journal of Computer-Mediated Communication, 13 (1), article 11,hWp://jcmc.indiana.edu/vol13/issue1/boyd.ellison. html.

Brin, S. and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Networks and ISDN Systems, 30 (1-7), 107–17.

Broadus, R. N. (1987) Early Approaches to Bibliometrics, Journal of Information
Science, 38 (2), 127–9.

Broder, A., Kumar, R., Maghoul, F., Raghaven, P., Rajagopalan, S., Stata, R., Tomkin, A. and Wiener, J. (2000) Graph Structure in the Web, Computer Networks, 33 (1-6),
309-20.

Bruns, A. and Highfield, T. (2013) Political Networks on TwiWer: tweeting the
Queensland state election, Information Communication & Society, 16 (5), 667-91.

Butler, D. (2013) When Google got Flu Wrong, Nature, 494, 155–6.

Chan, E. H., Sahai, V., Conrad, C. and Brownstein, J. S. (2011) Using Web Search Query Data to Monitor Dengue Epidemics: a new model for neglected tropical disease surveillance, PLoS Negl Trop, 5 (5), e1206.

Chen, C., Newman, J., Newman, R. and Rada, R. (1998) How did University

Departments Interweave the Web: a study of connectivity and underlying factors,Interacting with Computers, 10 (4), 353-73.

Chew, C. and Eysenbach, G. (2010) Pandemics in the Age of TwiWer: content analysis of tweets during the 2009 H1N1 Outbreak, PLOS ONE, 5 (11), e14118.

Choi, H. and Varian, H. (2012) Predicting the Present with Google Trends, Economic Record, special issue, June, 2-9.

Choi, S., Park, J. Y. and Park, H. W. (2012) Using Social Media Data to Explore

Communication Processes Within South Korean Online Communities,

Scientometrics, 90 (1), 43-56.

Chowdhury, S. A. and Makaroff, D. (2013) Popularity Growth PaWerns of YouTube Videos: a category-based study. In Krempels, K. H. and Stocker, A. (eds), Proceedings of WEBIST 2013: 8th International Conference on Web Information Systems and Technologies, 233-42.

Christakis, N. and Fowler, J. (2009) Connected: the amazing power of social networks and how they change lives, LiWle, Brown, and Company.

Chung, C. J. and Park, H. W. (2012) Web Visibility of Scholars in Media and

Communication Journals, Scientometrics, 93 (1), 207-15.

Ciesielski, K., Borkowski, P., Klopotek, M. A., Trojanowski, K. and Wysocki, K. (2012) Wikipedia-based Categorization, Security and Intelligent Information Systems, Lecture Notes in Computer Science, vol. 7053, 265-78.

Clauson, K. A., Polen, H. H., Boulos, M. N. K. and Dzenowagis, J. H. (2008) Scope, Completeness, and Accuracy of Drug Information in Wikipedia, Annals of Pharmacotherapy, 42, 1814–21.

Compete (2013) Our Data, www.compete.com/us/about/our-data/.

Cothey, V., Aguillo, I. and Arroyo, N. (2006) Operationalising 'Websites': lexically, semantically or topologically?, Cybermetrics, 10 (1), www.cindoc.csic.es/cybermetrics/articles/v10i1p3. html.

Cronin, B. (1984) The Citation Process, Taylor Graham.

Cunningham, J. A. (2012) Using TwiWer to Measure Behaviour PaWerns,

Epidemiology, 23 (5), 764-5.

Davenport, T. H. and Patil, D. J. (2012) Data Scientist: the sexiest job of the 21st century, Harvard Business Review, October.

De Bellis, N. (2009) Bibliometrics and Citation Analysis: from the Science Citation Index to Cybermetrics, Scarecrow Press.

Del Bosque, D., Leif, S. A. and Skarl, S. (2012) Libraries AtwiWer: trends in academic library tweeting, Reference Services Review, 40 (2), 199–213.

Didegah, F. and Erfanmanesh, M. A. (2010) The Study of Malaysian Public
Universities' Performance on the World Wide Web, Library Hi Tech News, 27 (3),7–11.

Dijkstra, E. W. (1968) A Case Against the GO TO Statement, Communications of the ACM, 11 (3), 147–8.

Du, R. Y. and Kamakura, W. A. (2012) Quantitative TrendspoWing, Journal of
Marketing Research, 49 (4), 514-36.

Dunbar, R. (2011) How Many Friends Does One Person Need? Dunbar's number and other evolutionary quirks, Faber & Faber.

Dzielinski, M. (2012) Measuring Economic Uncertainty and its Impact on the Stock Market, Finance Research Le،ers, 9 (3), 167–75.

Ebersbach, A., Glaser, M. and Heigl, R. (2008) Wiki: web collaboration, 2nd edn, Springer.

Edelman, B. and Larkin, I. (2009) Demographics, Career Concerns or Social Comparison: who games SSRN download counts?, Harvard Business School NOM Unit Working Paper No. 09-096, www.hbs.edu/research/pdf/09-096.pdf.

Egnal, M. (2013) Evolution of the Novel in the United States: the statistical evidence, Social Science History, 37 (2), 231-54.

Ennew, C., LockeW, A., Blackman, I. and Holland, C. P. (2005) Competition in
Internet Retail Markets: the impact of links on web site traffic, Long Range
Planning, 38 (4), 359-72.

EWredge, M., Gerdes, J. and Karuga, G. (2005) Using Web-Based Search Data to
Predict Macroeconomic Statistics, Communications of the ACM, 48 (11), 87-92.

Etzkowitz, H. and Leydesdorff, L. (1995) The Triple Helix: university-industry-government relations, EASST Review, 14 (1), www.easst.net/review/march1995/leydesdorff.

Eysenbach, G. (2006) Infodemiology: tracking flu-related searches on the web for syndromic surveillance, American Medical Informatics Association Annual
Symposium Proceedings, 244-8, www.ncbi.nlm.nih.gov/pmc/articles/PMC1839505/. Eysenbach, G. (2011) Can Tweets Predict Citations? Metrics of social impact based
on TwiWer and correlations with traditional metrics of scientific impact, Journal of Medical Internet Research, 13 (4), e123, www.jmir.org/2011/4/e123/. Facebook (2012) Newsroom – key facts, hWp://newsroom.k.com/Key-Facts.

Feldman, R. (2013) Techniques and Applications for Sentiment Analysis,
Communications of the ACM, 56 (4), 82-9.

Fields, E. (2010) A Unique TwiWer Use for Reference Services, Library Hi Tech News, 27 (6), 14–15.

Garfield, E. (1983) Citation Indexing – its theory and application in science, technology, and humanities, ISI Press.

Garfield, E. (2006) The History and Meaning of the Journal Impact Factor, JAMA,295 (1), 90-3.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., ScoW, P. and Trow, M.(1994) The New Production of Knowledge: the dynamics of science and research in contemporary societies, Sage Publications.

Giles, J. (2005) Internet Encyclopaedias Go Head to Head, Nature, 438, 900-1.

Giles, J. (2012) Making the Links, Nature, 448, 448–50.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, M. S., Smolinski and Brilliant, L. (2009) Detecting Influenza Epidemics Using Search Engine Query Data, Nature, 457, 1012–14.

Glynn, R. W., Kelly, J. C., Coffey, N., Sweeney, K. J. and Kerin, M. J. (2011) The Effect of Breast Cancer Awareness Month on Internet Search Activity – a comparison with awareness campaigns for lung and prostate cancer, BMC Cancer, 11, article 442.

Go, A., Bhayani, R. and Huang, L. (2009) Twi‿er Sentiment Classification Using Distant Supervision, technical report, Stanford, hWp://cs.stanford.edu/people/alecmgo/papers/TwiWerDistantSupervision09.pdf .

Godin, B. (2006) On the Origins of Bibliometrics, Scientometrics, 68 (1), 109–33. Golder, S. A. and Macy, M. W. (2011) Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures, Science, 333, 30 September,1878–81.

Google (2013) Google: Webmaster Tools – what file types can Google Index?, hWp://support.google.com/webmasters/bin/answer.py?hl=en&answer=35287.

GranoveWer, M. S. (1973) The Strength of Weak Ties, American Journal of Sociology,78 (6), 1360–80.

Gruber, T. R. (1993) A Translation Approach to Portable Ontology Specifications,Knowledge Acquisitions, 5 (2), 199–220.

Guardian (2012) University Guide 2013: university league table, Guardian, 21 May, www.guardian.co.uk/education/table/2012/may/21/university-league-table-2013.

Halpern, D. and Gibbs, J. (2013) Social Media as a Catalyst for Online Deliberation?

Exploring the affordances of Facebook and YouTube for political expression,

Computers in Human Behaviour, 29 (3), 1159–68.

Hand, C. and Judge, G. (2012) Searching for the Picture: forecasting UK cinema admissions using Google Trends data, Applied Economics Le‿ers, 19 (11), 1051–5.

Hardiman, S. J. and Katzir, L. (2013) Estimating Clustering Coefficients and Size of Social Networks Via Random Walk, WWW2013, May 13–17, Rio de Janeiro, Brazil.

Hardy, Q. (2012) Harvard Releases Big Data For Books, New York Times: Bits, hWp://bits.blogs.nytimes.com/2012/04/24/harvard-releases-big-data-for-books.

Harnad, S., Carr, L., Brody, T. and Oppenheim, C. (2003) Mandated Online RAE CVs Linked to University Eprint Archives: enhancing UK research impact and assessment, Ariadne, 35, www.ariadne.ac.uk/issue35/harnad.

Harter, S. P. and Ford, C. E. (2000) Web-based Analyses of E-Journal Impact:

approaches, problems, and issues, Journal of the American Society for Information

Science, 51 (13), 1159–76.

He, W., Chee, T., Chong, D. and Rasnick, E. (2012) Using Bibliometrics and Text Mining to Explore the Trends of E-Marketing Literature from 2001 to 2010, International Journal of Online Marketing, 2 (1), 16-24.

Himelboim, I., McCreery, S. and Smith, M. (2013) Birds of a Feather Tweet Together: integrating network and content analysis to examine cross-ideology exposure on TwiWer, Journal of Computer-Mediated Communication, 18 (2), 40-60.

Hirsch, J. E. (2005) An Index to Quantify an Individual's Scientific Research Output,Proceedings of the National Academy of Sciences of the United States of America,102 (46), 16569–72.

Holmberg, K. (2009) Webometric Network Analysis – mapping cooperation and geopolitical connections between local government administration on the Web, dissertation, Åbo: Åbo Akademi UP.

Hood, W. W. and Wilson, C. S. (2001) The Literature of Bibliometrics, Scientometrics, and Informetrics, Scientometrics, 52, 291-314.

Hrynaszkiewicz, I. (2013) Version Control for Scientific Research, BioMed Centralblog, hWp://blogs.biomedcentral.com/bmcblog/2013/02/28/version-control-for-scientific-research.

Hsu, C. L. and Park, H. W. (2012a) Korean and Chinese Webpage Content: who are talking about what and how?, Journal of Computer-Mediated Communication, 17 (2),202-15.

Hsu, C. L. and Park, H. W. (2012b) Mapping Online Social Networks of Korean Politicians, Government Information Quarterly, 29 (2), 169–81. Hubbard, D. W. (2011) Pulse, John Wiley & Sons.

Hung, J. (2012) Trends of E-Learning Research from 2000-2008: use of text mining and bibliometrics, British Journal of Educational Technology, 43 (1), 5–16. Information Commission Office (2012) The EU Cookie Law,

www.ico.gov.uk/for_organisations/privacy_and_electronic_communications/the_guide/cookies.aspx.

Ingwersen, P. (1998) The Calculation of Web Impact Factors, Journal of

Documentation, 54 (2), 236–43.Internet Archive (2012) 80 Terabytes of Archived Web Crawl Data Available for Research: internet archive blogs, hWp://blog.archive.org/2012/10/26/80-terabytes-

of-archived-web-crawl-data-available-for-research.

Jackson, J. (2010) Google: 129 million different books have been published, www.pcworld.com/article/202803/google_129_million_different_books_have_been_published. html.

Jacso, P. (2012) Google Scholar Metrics for Publications: the sowware and content features of a new open access bibliometric service, Online Information Review,36 (4), 604-19.

Journalism.org (2005) Seigenthaler and Wikipedia: a case study on the veracity of the 'wiki' concept, www.journalism.org/node/1672.

Kaplan, A. M. and Haenlein, M. (2010) Users of the World, Unite! The challenges and opportunities of social media, Business Horizons, 53, 59-68.

Kazienko, P., Szozda, N., Filipowski, T. and Blysz, W. (2013) New Business Client Acquisition Using Social Networking Sites, Electronic Markets, 23 (2), 93–103. Kepler, T. B., Marti-Renom, M. A., Maurer, S. M., Rai, A. K., Taylor, G. and Todd, M. H. (2006) Open Source Research – the power of us, Australian Journal of Chemistry, 59 (5), 291-4.

Kim, H. M., Abels, E. G. and Yang, C. C. (2012) Who Disseminates Academic Library Information on TwiWer, Proceedings of the American Society for Information Science and Technology, 49 (1), 1-4.

Kimmons, R. (2011) Understanding Collaboration in Wikipedia, First Monday,

16 (12), hWp://firstmonday.org/ojs/index.php/fm/article/view/3613/3117. Kousha, K. and Thelwall, M. (2006) Motivations for URL Citations to Open Access

Library and Information Science Articles, Scientometrics, 68 (3), 501-17.

Kousha, K. and Thelwall, M. (2008) Assessing the Impact of Disciplinary Research on Teaching: an automatic analysis of online syllabuses, Journal of the American Society for Information Science and Technology, 59 (13), 2060-9.

Kousha, K., Thelwall, M. and Rezaie, S. (2011) Assessing the Citation Impact of

Books: the role of Google Books, Google Scholar, and Scopus, Journal of the

American Society for Information Science and Technology, 62 (11), 2147-64.

Krauth, S. J., Coulibaly, J. T., Knopp, S., Traoré, M., N'Goran, E. K. and Utzinger, J. (2012) An In-Depth Analysis of a Piece of Shit: distribution of Schistosoma mansoni and hookworm eggs in human stool, PLoS Neglected Tropical Diseases, 6 (12), e1969.

Krippendorff, K. H. (2012) Content Analysis: an introduction to its methodology, 3rd edn, Sage Publications.

Kronick, D. A. (1990) Peer Review in 18th-century Scientific Journalism, Journal of the American Medical Association, 263 (10), 1321-2.

Kuper, S. and Szymanski, S. (2012) Soccernomics, HarperSport.

Kwan, Y. and Chan, L. M. (2009) Linking Folksonomy to Library of Congress Subject Headings: an exploratory study, Journal of Documentation, 65 (5), 872-900.

Lai, K. F. and Wang, D. (2013) Understanding the External Links of Video Sharing Sites: measurement and analysis, IEEE Transactions on Multimedia, 15 (1), 224-35. Larson, R. R. (1996) Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyber space. In Hardin, S. (ed.), Proceedings of the 59th Annual Meeting, ASIS, Learned Information Ltd, 71-9.

Lawrence, P. A. (2007) Popular Beat May Drown Out Genius, Times Higher Education, 24 August, www.timeshighereducation.co.uk/310247.article.

Lazarev, V. S. (1996) On Chaos in Bibliometric Terminology, Scientometrics, 35 (2), 271-7.

Lewis, M. (2004) Moneyball, W. W. Norton. Li, L. (2011) Social Network Site Comparison Between the United States and China:case study on Facebook and Renren network. In Proceedings of the 2011 International Conference on Business Management and Electronic Information, 825-7. Li, X., Thelwall, M., Musgrove, P. and Wilkinson, D. (2003) The Relationship Between the WIFs or Inlinks of Computer Science Departments in the UK and their RAE Ratings or Research Productivities in 2001, Scientometrics, 57 (2), 239-55..

LipseW, A. (2006) Metrics Will Kill Diversity Claim, Times Higher Education, 15 December, www.timeshighereducation.co.uk/news/metrics-will-kill-diversity-

claim/207149.article.

LipseW, A. (2007) Report: bibliometrics could distort research assessment, Guardian, 9 November, www.theguardian.com/education/2007/nov/09/highereducation. researchassessmentexercise.

Liu, B. (2012) Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers.

Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) Protein

Measurement with the Folin Phenol Reagent, Journal of Biological Chemistry,

193 (1), 265-75.

Lundvall, B. A. (ed.) (1992) National Systems of Innovation: towards a theory of innovation and interactive learning, Pinter.

MacRoberts, M. H. and MacRoberts, B. R. (1996) Problems of Citation Analysis, Scientometrics, 36 (3), 435-44.

Madden, K., Nan, X., Briones, R. and Waks, L. (2012) A Content Analysis of HPV Vaccine Information Online, Vaccine, 30 (25), 3741-6.

Marek, K. (2011) Using Web Analytics in the Library, Library Technology Reports, 47 (5).

Marriner, N. and Morhange, C. (2013) Data Mining the Intellectual Revival of

'Catastrophic' Mother Nature, Foundations of Science, 18 (2), 245-57.

Mendeley (2012) Global Research Report, www.mendeley.com/global-research-report.Microformats (2012) Microformats at 7,hWp://microformats.org/2012/06/25/microformats-org-at-7. Microsow (2012) Microsow Academic Search API,

hWp://academic.research.microsow.com/about/Microsow%20Academic%20
Search%20API%20User%20Manual.pdf.

Milstein, S. (2009) TwiWer for Libraries and Librarians, Computers in Libraries, 29 (5),17–18.

Minguillo, D. and Thelwall, M. (2012) Mapping the Network Structure of Science Parks: an exploratory study of cross-sectoral interactions reflected on the web, Aslib Proceedings, 64 (4), 332-57.

Moed, H. (2005) Citation Analysis in Research Evaluation, Springer. Mola-Velasco, S. M. and Rosso, P. (2011) Wikipedia Vandalism Detection. In Proceedings of the 20th International Conference Companion on World Wide Web (WWW), 391-6.

Mortensen, P. S. (2011) Patentometrics as Performance Indicators for Allocating Funding to Universities, hWp://pure.au.dk/portal/files/39714215/WP2011_1_Patentometrics_
as_performance_indicators.pdf.

Mukewar, S., Mani, P., Wu, X. R., Lopez, R. and Shen, B. (2013) YouTube and
Inflammatory Bowel Disease, Journal of Crohns & Colitis, **7** (5), 392-402. Narin, F. (1994) Patent Bibliometrics, Scientometrics, 30 (1), 147–55.

Nature Materials (2012) Alternative Metrics: editorial, Nature Materials, 11, 907, htmwww.nature.com/nmat/journal/v11/n11/full/nmat3485.html.

Noyons, E. (2001) Bibliometric Analysis of Science in a Science Policy Context,
Scientometrics, 50 (1), 83–98.

Oboler, A., Welsh, K. and Cruz, L. (2012) The Danger of Big Data: social media as computational social science, First Monday, 17 (7), hWp://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3993/3269.

OED (2001) Metric, Oxford English Dictionary, www.oed.com/view/Entry/117657. Orduña-Malea, E. and Ontalba-Ruiperez, J. A. (2013) Selective Linking from Social Platforms to University Websites: a case study of the Spanish academic system,Scientometrics, 95 (2), 593-614.

Orduña-Malea, E. and Regazzi, J. J. (2013) US Academic Libraries: understanding their web presence and their relationship with economic indicators,Scientometrics,
hWp://link.springer.com/article/10.1007%2Fs11192-013-1001-0. O'Reilly, T. (2005) What is Web 2.0: design pa ِerns and business models for the next
generation of so2ware, hWp://oreilly.com/pub/a/web2/archive/what-is-web-20.html. Ortega, J. L. and Aguillo, I. F. (2008) Visualization of the Nordic Academic Web: link
analysis using social network tools, Information Processing & Management, 44 (4), 1624–33.

Ortega, J. L. and Aguillo, I. F. (2012) Science is All in the Eye of the Beholder: keyword maps in Google Scholar citations, Journal of the American Society for Information Science and Technology, 62 (12), 2370-7.

Ortega, J. L. and Aguillo, I.F. (2013) Institutional and Country Collaboration in an Online Service of Scientific Profiles: Google Scholar citations, Journal of Informetrics, **7** (2), 394-403.

Ortiz, J. R., Zhou, H., Shay, D. K., Neuzil, K. M. and Goss, C. H. (2010) Does Google Influenza Tracking Correlate with Laboratory Tests Positive for Influenza?, American Journal of Respiratory Critical Care Medicine, 181, A2626.

OWe, E. and Rousseau, R. (2002) Social Network Analysis: a powerful strategy, also for the information sciences, Journal of Information Sciences, 28 (6), 441-53.

Oyelude, A. A. and Bamigbola, A. A. (2012) Libraries as the Gate: 'ways' and

'keepers' in the knowledge environment, Library Hi Tech News, 29 (8), 7–10.

Ozel, B. and Park, H. W. (2012) Online Image Content Analysis of Political Figures:an exploratory study, Quality and Quantity, 46 (4), 1013–24.

Paltoglou, G., Thelwall M. and Buckley K. (2010) Online Textual Communications Annotated with Grades of Emotion Strength. In Proceedings of the 3rd International Workshop of Emotion: corpora for research on Emotion and Affect, 25-31.

Pantelidis, I. S. (2010) Electronic Meal Experience: a content analysis of online restaurant comments, Cornell Hospitality Quarterly, 51 (4), 483-91.

Park, H. W., BarneW, G. A. and Nam, I.-Y. (2002) Hyperlink-Affiliation Network Structure of Top Web Sites: examining affiliates with hyperlink in Korea, Journal of the American Society for Information Science and Technology, 53 (7), 592-601.

Paul, J. A., Baker, H. M. and Cochran, J. D. (2012) Effect of Online Social Networking on Student Academic Performance, Computers in Human Behaviour, 28 (6),2117-27.

Pesch, O. (2007) Usage Statistics, Information Service & Use, 207, 207-13.

Piwowar, H. (2013) Value All Research Products, Nature, 493, 159.

Piwowar, H. A., Carlson, J. D. and Vision, T. J. (2011) Beginning to Track 1000

Datasets from Public Repositories into the Published Literature, Proceedings of the American Society for Information Science and Technology, 48 (1), 1-4. Polanyi, M. (1966) The Tacit Dimension, University of Chicago Press.

Ponce, B. A., Determann, J. R., Boohaker, H. A., Sheppard, E., McGwin, G. and Theiss, S. (2013) Social Networking Profiles and Professionalism Issues in Residency Applicants: an original study-cohort study, Journal of Surgical Education, 70 (4), 502-7.

Popegrutch (2011) Marketing Academic Libraries Through Social Media: resources and examples, hWp://popegrutch.wordpress.com/2011/02/23/marketing-academic-

libraries-through-social-media-resources-examples.

Price, A. and Grann, V. R. (2012) Portrayal of Complementary and Alternative Medicine for Cancer by Top Online News Sites, Journal of Alternative and Complementary Medicine, 18 (5), 487–93.

Price, D. J. S. (1963) Li ِle Science, Big Science, Columbia University Press.

Priem, J., Taraborelli, D., Groth, P. and Neylon, C. (2010) Altmetrics: a manifesto, hWp://altmetrics.org/manifesto.

Pries, T., Moat, H. S., Stanley, H. E. and Bishop, S. R. (2012) Quantifying the Advantage of Looking Forward, Scientific Reports, www.ncbi.nlm.nih.gov/pmc/articles/PMC3320057.

Pritchard, A. (1969) Statistical Bibliography or Bibliometrics?, Journal of

Documentation, **25** (4), 348-9.Quantcast (2013) How We Do It, www.quantcast.com/how-we-do-it/methodology.

Quint, B. (1998) A 'Giw of the Web' for the Library of Congress from Alexa Internet,Information Today, hWp://newsbreaks.infotoday.com/nbreader.asp?ArticleID=17893.

Ranganathan, S. R. (1931) Five Laws of Library Science, Madras Library Association. Raymond, E. S. (2001) The Cathedral and the Bazaar: musings on the Linux and Open Source by an accidental revolutionary, O'Reilly Media.

Rector, L. H. (2008) Comparison of Wikipedia and Other Encyclopedias for

Accuracy, Breadth, and Depth in Historical Articles, Reference Services Review,36 (1), 7-22.

Rolnick, A. J. and Weber, W. E. (1986) Gresham's Law or Gresham's Fallacy, Journal of Political Economy, 94 (1), 185–99.

Rorick, W. C. (1987) Discometrics – a system for acquiring scores and sound recordings, Library Journal, 112 (19), 45–7.

Rousseau, R. (1997) Sitations: an exploratory study, Cybermetrics, 1 (1), hWp://cybermetrics.cindoc.csic.es/pruebas/v1i1p1.htm.

Russell, A. L. (2005) Standardization in History: a review essay with an eye to the future. In Bolin, S. (ed.), The Standards Edge: future generations, Sheridan Press,247–60.

Schreiber, M., Malesios, C. C. and Psarakis, S. (2012) Exploratory Factor Analysis for the Hirsch Index, 17 h-type variants and some traditional bibliometric indicators, Journal of Informetrics, 6 (3), 347–58.

Schubert, A. (2012) Jazz Discometrics – a network approach, Journal of Informetrics,6 (4), 480–4.

Schuster, N. M., Rogers, M. A. M. and McMahon, L. F. (2010) Using Search Engine Query Data to Track Pharmaceutical Utilization: a study of statins, American Journal of Managed Care, 16 (8), E215–19.

Seiwer, A., Schwarzwalder, A., Geis, K. and AucoW, J. (2010) The Utility of 'Google Trends' for Epidemiological Research: Lyme disease as an example, Geospatial Health, 4 (2), 135–7.

Sewell, R. R. (2013) Who Is Following Us? Data mining a library's TwiWer followers,Library Hi Tech, 31 (1), 160–70.

Shapiro, F. R. (1992) Origins of Bibliometrics, Citation Indexing, and Citation Analysis: the neglected legal literature, Journal of the American Society for Information Science, 43 (5), 337–9.

Shuai, X., Pepe, A. and Bollen, J. (2012) How the Scientific Community Reacts to Newly SubmiWed Preprints: article downloads, TwiWer mentions, and citations, PLOS ONE, **7** (11), e47523.

Silver, N. (2012) The Signal and the Noise: the art and science of prediction, Allen Lane. Simpson, C. (2008) Five Laws, Library Media Connection, April/May, 6.

Singer, R. (2009) Content Sources and Mashing Them Up. In Engard, N. (ed.),

Library Mashups: exploring new ways to deliver library data, Facet Publishing. Skolnik, H. (1979) Historical Development of Abstracting, Journal of Chemical Information and Computer Science, 19 (4), 215-18.

Smith, A. (1999) ANZAC Webometrics: exploring Australasian webstrctures, hWp://conferences.alia.org.au/online1999/proceedings99/203b.html

Sterne, J. (2010) Social Media Metrics, John Wiley & Sons.

Stonier, T. (1999) Information as a Basic Property of the Universe, Biosystems,

38 (2-3), 135–40.

Strathern, M. (1997) 'Improving Ratings': audit in the British University system,European Review, 5 (3), 305–21.

Strychowsky, J. E., Nayan, S., Farrokhyar, F., Maclean, J. (2013) YouTube: a good source of information on pediatric tonsillectomy?, International Journal of Pediatric Otorhinolaryngology, 77 (6), 972–5.

Stuart, D. (2010) What Are Libraries Doing on TwiWer?, Online, 34 (1), 45-7.

Stuart, D. (2011) Facilitating Access to the Web of Data: a guide for librarians, Facet

Publishing.

Stuart, D. (2012) FOAF Within UK Academic Web Space: a webometric analysis of the semantic web. In Widén, G. and Holmberg, K. (eds), Social Information Research, Emerald Publishing, 173–91.

Stuart, D. and Thelwall, M. (2006) Investigating Triple Helix Relationships Using URLCitations: a case study of the UK West Midlands automobile industry, Research Evaluation, 15 (2), 97–106.

Stuart, D. and Thelwall, M. (2007) University-Industry-Government Relationships

Manifested Through MSN Reciprocal Links. In Torres-Salinas, D. and Moed, H.F. (eds), Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics, vol. 2, Madrid: CINDOC, 731–5.

Stuart, D., Thelwall, M. and Harries, G. (2007) UK Academic Web Links and

Collaboration – an exploratory study, Journal of Information Science, 33 (2), 231-46. Sueki, H. (2011) Does the Volume of Internet Searches Using Suicide-Related Search Terms Influence the Suicide Death Rate: data from 2004 to 2009 in Japan,Psychiatry and Clinical Neurosciences, 65 (4), 392-4.

Sugimoto, C. R., Thelwall, M., Lariviere, V., Tsou, A., Mongeon, P. and Macaluso, B. (2013) Scientists Popularising Science: characteristics and impact of TED talk, PLOS ONE, 8 (4), e62403.

Sullivan, D. (2012) Google: 100 billion searches per month, search to integrate Gmail,launching enhanced search app for iOS, hWp://searchengineland.com/google-search-press-129925.

Sullivan, D. (2013) Marketing Land: Chrome to gain search encryption, following similar moves by Firefox & Mobile Safari, hWp://marketingland.com/chrome-to-gain-encryption-31085.

Swanson, D. R. (1986) Fish Oil, Raynaud's Syndrome, and Undiscovered Public

Knowledge, Perspectives in Biology and Medicine, 30 (1), 7–18.

Tague-Sutcliffe, J. (1992) An Introduction to Informetrics, Information Processing & Management, 38 (1), 1-3.

Tancer, B. (2009) Click: what we do online and why it ma¸ers, Harper Collins. Thelwall, M. (2001a) Exploring the Link Structure of the Web with Network Diagrams, Journal of Information Science, 27 (6), 393–401.

Thelwall, M. (2001b) Results from a Web Impact Factor Crawler, Journal of

Documentation, 57 (2), 177–91.

Thelwall, M. (2002) Conceptualizing the Documentation on the Web: an evaluation of different heuristic-based models for counting links between university web sites, Journal of the American Society for Information Science and Technology, 53 (12),995–1005.

Thelwall, M. (2004a) Link Analysis: an information science approach, Elsevier Academic Press.

Thelwall, M. (2004b) Weak Benchmarking Indicators for Formative and Semi-

Evaluative Assessment of Research, Research Evaluation, 13 (1), 63-8.

Thelwall, M. (2008) Social Networks, Gender, and Friending: an analysis of MySpace member profiles, Journal of the American Society for Information Science and Technology, 59 (8), 1321–30.

Thelwall, M. (2009) Introduction to Webometrics: quantitative web research for the social sciences, Morgan & Claypool.

Thelwall, M. (2012) Journal Impact Evaluation: a webometric perspective,

Scientometrics, 92 (2), 429–41.

Thelwall, M. and Price, L. (2003) Disciplinary Differences in Academic Web

Presence: a statistical study of the UK, Libri, 53 (4), 242–53.

Thelwall, M. and Stuart, D. (2006) Web Crawling Ethics Revisited: cost, privacy, and denial of service, Journal of the American Society of Information Science and Technology, 57 (13), 1771–9.

Thelwall, M. and Stuart, D. (2010) Social Network Sites: an exploration of features and diversity. In Zaphiris, P. and Ang, C. S. (eds), Social Computing and Virtual Communities, CRC, 265-84.

Thelwall, M. and Sud, P. (2011) A Comparison of Methods for Collecting Web Citation Data for Academic Organizations, Journal of the American Society for Information Science and Technology, 62 (8), 1488–97.

Thelwall, M. and Sud, P. (2012) Webometric Research with Bing Search API 2.0, Journal of Informetrics, 6 (1), 44–52.

Thelwall, M. and Wilkinson, D. (2004) Finding Similar Academic Web Sites with Links, Bibliometric Couplings and Colinks, Information Processing & Management, 40 (3), 515–26.

Thelwall, M., Harries, G. and Wilkinson, D. (2003) Why do Web Sites from Different Academic Subjects Interlink?, Journal of Information Science, 29 (6), 453–71. Thelwall, M., Haustein, S., Larivière, V. and Sugimoto, C. R. (2013) Do Altmetrics Work? TwiWer and ten other social web services, PLOS ONE, 8 (5): e64841.

Thornton, E. (2012) Is Your Academic Library Pinning? Academic libraries and Pinterest, Journal of Web Librarianship, 6 (3), 164–75.

Torres-Salinas, D., Cabezas-Clavijo, A. and Ruiz-Perez, R. (2011) State of the Library and Information Science Blogosphere Awer Social Networks Boom: a metric approach, Library & Information Science Research, 33 (2), 168–74.

Turow, J. (2011) The Daily You: how the advertising industry is defining your identity and worth, Yale University Press.

Twenge, J. M., Campbell, W. K. and Gentile, B. (2012a) Increases in Individualistic Words and Phrases in American Books, 1960-2008, PLOS ONE, **7** (7), e40181. Twenge, J. M., Campbell, W. K. and Gentile, B. (2012b) Male and Female Pronoun Use in US Books Reflects Women's Status, 1900-2008, Sex Roles, **67** (9–10), 488–93. TwiWer (2013) Celebrating #TwiWer7,

hWps://blog.twiWer.com/2013/celebrating-twiWer7.

UNESCO (1985) Recommendation Concerning the International Standardization of

Statistics Relating to Book Production and Periodicals, hWp://portal.unesco.org/en/ev.php- URL _ID=13068&URL _DO=DO_TOPIC&URL_SECTION=201.html.

US Postal Service (2012) Books: general: definition of a book, hWp://pe.usps.com/text/pub2/pub2c6_002.html.

Van Noorden, R. (2012) What Were the Top Papers of 2012 on Social Media?, Nature News Blog, hWp://blogs.nature.com/news/2012/12/what-were-the-top-papers-of-

2012-on-social-media.html.

Van Noorden, R. (2013) Tensions Grow as Data-Mining Discussions Fall Apart,

Nature, 498, 14–15.

Vaughan, L. (2006) Visualizing Linguistic and Cultural Differences Using Web

Co-Link Data, Journal of the American Society for Information Science and Technology, 57 (9), 1178–93.

Vaughan, L. (2012) An Alternative Data Source for Web Hyperlink Analysis: 'Sites Linking In' at Alexa Internet, COLLNET Journal of Scientometrics and Information Management, 6 (4), 31–42.

Vaughan, L. and Romero-Frías, E. (2012) Exploring Web Keyword Analysis as an Alternative to Link Analysis: a multi-industry case, Scientometrics, 93 (1), 217–32.

Vaughan, L. and Yang, R. (2012) Web Data as Academic and Business Quality Estimates: a comparison of three data sources, Journal of the American Society for Information Science and Technology, 63 (10), 1960–72.

Vaughan, L., Gao, Y. and Kipp, M. (2006) Why are Hyperlinks to Business Websites Created? A content analysis, Scientometrics, 67 (2), 291–300.

Vera, C. R. T. (2009) Open Notebook Science: an exploratory study on motivations and perceived challenges. In Shuhua, H. and Thota, H. (eds), Proceedings of the 4th International Conference on Product Information Management, 1779–85.

Vosen, S. and Schmidt, T. (2012) A Monthly Consumption Indicator for Germany Based on Internet Search Query, Applied Economics Le‚ers, 19 (7), 683-7.

Voß, J. (2005) Measuring Wikipedia. In P. Ingwersen and B. Larsen (eds), Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics, vol. 1, Stockholm: Karolinska University Press, 221-31.

Vucovich, L. E., Gordon, V. S., Mitchell, N. and Ennis, L. A. (2013) Is the Time and Effort Worth It? One library's evaluation of using social networking tools for outreach, Medical Reference Services Quarterly, 32 (1), 12-25. W3C (2012a) Ontology Dowsing,

www.w3.org/wiki/Ontology_Dowsing. W3C (2012b) RDFa 1.1 Primer, www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/#html-vs.-x.html.

W3Techs (2013) Usage Statistics and Market Share of Google Analytics for websites, hWp://w3techs.com/technologies/details/ta-googleanalytics/all/all.

Wagstaff, A. and Culyer, A. J. (2012) Four Decades of Health Economics Through a Bibliometric Lens, Journal of Health Economics, 31 (2), 406-39.

WalcoW, B. P., Nahed, B. V., Kahle, K. T., Redjal, N. and Coumans, J. V. (2011) Determination of Geographic Variance in Stroke Prevalence Using Internet Search Engine Analytics, Neurosurgical Focus, 30 (6), E19.

White, H. D. and McCain, K. W. (1989) Bibliometrics. In Williams, M. E. (ed.),Annual Review of Information Science and Technology, 24, 119–86.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, S. and Cole, F. T. H. (2009) Libcitations: a measure for comparative assessment of book publications in the humanities and social sciences, Journal of the American Society for Information Science and Technology, 60 (6), 1083–96.

Wikimedia Foundation (2011) Editor Survey Report, hWp://commons.wikimedia.org/

wiki/File%3AEditor_Survey_Report_-_April_2011.pdf.

Wikipedia (2013a) List of Social Networking Sites, hWp://en.wikipedia.org/wiki/List_of_social_networking_websits

Wikipedia (2013b) Wikipedia: Wikipedia Loves Libraries, hWp://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Loves_Libraries.

Wikipedia (2013c) Wikipedians, hWp://en.wikipedia.org/wiki/Wikipedia:Wikipedians.

Wilkinson, D. and Huberman, B. (2007) Assessing the Value of Cooperation in

Wikipedia, First Monday, 12 (4), hWp://firstmonday.org/ojs/index.php/fm/article/view/1763/1643.

Wilkinson, D. and Thelwall, M. (2011) Researching Personal Information on thePublic Web: methods and ethics, Social Science Computer Review, **29** (4), 387–401. Wilkinson, D. and Thelwall, M. (2012) Trending Topics in English: an international comparison, Journal of the American Society for Information Science and Technology,63 (8), 1631–46.

Wong, V. S. S., Stevenson, M. and Selwa, L. (2013) The Presentation of Seizures and Epilepsy in YouTube Videos, Epilepsy & Behaviour, **27** (1), 247–50.

Xiang, X. (2012) Linguistic and Cultural Characteristics of Domain Names of the Top Fiwy Most-Visited Web Sites in the US and China: a cross-linguistic study of domain names and e-branding, Names – A Journal of Onomastics, 60 (4), 210–19.

Yan, E., Ding, Y. and Zhu, Q. (2010) Mapping Library and Information Science in China: a coauthorship network analysis, Scientometrics, 83 (1), 115-31.

Yeo, G. (2010) Introduction to Archives: principles and practices, Facet Publishing.

Zhan, Y. and Yan, Y. (2011) Construction and Optimization of Recruitment Websites in China. In Fourth International Conference on Business Intelligence and Financial Engineering (BIFE), 143–6.

Zhang, X., Fuehres, H. and Gloor, P. A. (2012) Predicting Asset Value through

TwiWer Buzz, Advances in Collective Intelligence 2011, 23–34.

Zhang, Y., He, D. and Sang, Y. (2013) Facebook as a Platform for Health Information and Communication: a case study of a diabetes group, Journal of Medical Systems,37 (3), article 9942.

Ziman, J. M. (1969) Information, Communication, Knowledge, Nature, 224, 318–24.

واژه‌نامه انگلیسی به فارسی

Academic search engines	موتورهای جستجوی دانشگاهی
Advertising	تبلیغ
Bibliographic search engines	موتورهای جستجوی کتابشناختی
Big data	داده‌های بزرگ
Blog	وب‌نوشت
Blogger	وب‌نویس
Blogosphere	فضای وب‌نوشت‌ها
Blogs	وبی‌نوشت‌ها
Bonferroni correction	تصحیح بونفرونی
Bookmarking	مکان‌یابی
Bow tie	پاپیون
Categorization	دسته‌بندی
Centrality	مرکزیت
Chi-square	کای‌دو
Citation	استناد
Citation analysis	تحلیل استنادی
Click	کلیک
Cluster	خوشه
Content analysis	تحلیل محتوا
Cookies	شناسه شناسایی کاربر
Correlation	همبستگی
Criticisms	انتقادها
Cybermetrics	سایبرسنجی
Data	داده

Data extraction	استخراج داده
Data science	علم داده
Data scientist	دانشمند داده
Definition	تعریف
Digitization	دیجیتال‌سازی
Dublin Core	هسته دوبلین
Dunbar number	تعداد دانبار
e-books	کتاب–الکترونیکی
e-scientist	دانشمند الکترونیکی
Economic indicators	شاخص‌های اقتصادی
European Commission	کمسیون اروپا
Evaluative	سنجشی
Facebook	فیسبوک
Flickr	فلیکر
Foursquare	فورسکوئر
Freedom of information	آزادی اطلاعات
g-index	شاخص جی
Goodhart's law	قانون گودهارت
Google	گوگل
Google Flu Trends	روندهای سرماخوردگی گوگل
Google public Data	داده‌های عمومی گوگل
Google Trends	روندهای گوگل
Graph density	تراکم نمودار
Grey literature	متون خاکستری
Guardian	گاردین
h-index	شاخص اچ
Harvard Business Review	نقد و بررسی کسب و کار هاروار
Harvard Library	کتابخانه هاروارد
Impact story	ایمپکت استوری
Indicators	شاخص‌ها
Information Commissioner's	دفتر کمیسیون اطلاعات
Informetrics	اطلاع‌سنجی

Instagram	اینتساگرام
Intellectual property rights	حقوق مالکیت معنوی
Internet Archive	آرشیو اینترنت
Internet Corporation for Assigned Names and Numbers	سازمان اینترنتی برای نام‌ها و شماره‌های تخصیصی
Internetometrics	اینترنت‌سنجی
Journal impact factor	ضریب تأثیر مجله
Journal of Digital Humanities	مجله علوم انسانی دیجیتال
Journal of Visualized Experiments	مجله تجربه‌های بصری
Libcitations	استنادهای کتابخانه
Library catalogues	فهرست کتابخانه
Library of Alexandria	کتابخانه الکساندریا
Library of Congress Subject Headings	سرعنوان‌های موضوعی کتابخانه کنگره
LinkedIn	لینکداین
Little Science, Big Science	علم کوچک، علم بزرگ
Marketing	بازاریابی
Metadata	فراداده
Microdata	میکروداده
Microformats	میکروفرمت
Micrographia	میکرونمودار یا
Microsoft Academic Search	جستجوی دانشگاهی مایکروسافت
Microsoft Word	مایکروسافت وورد
Modularity	مادوله
MOOCs	دوره‌های برخط باز فراگیر
Multiple regression analysis	تجزیه و تحلیل رگرسیون چندگانه
MySpace	مای اسپیس
National Central Library of Florence	کتابخانه ملی مرکزی فلورانس
National Central Library of Rome	کتابخانه ملی مرکزی رم
National Central Library of France	کتابخانه ملی فرانسه
National Central Library of Spain	کتابخانه ملی اسپانیا
Natural language processing	پردازش زبان طبیعی
Nature	طبیعت

Nature Linked Data Platform	بستر داده‌های پیوندی طبیعی
Nature Materials	مواد طبیعی
Netometrics	تعامل علمی واسطه‌گری اینترنت
NodeXL	نودکسل
Ontologies	هستی‌شناسی‌ها
Open access	دسترسی باز
Open Annotation Ontology	هستی‌شناسی حاشیه‌نویسی باز
Open data	داده‌های باز
OpenDAOR	اوپن دووار
Openly Local	اوپنلی لوکال
Open notebook science	علوم دفتر باز
Oxford English Dictionary	لغت‌نامه انگلیسی آکسفورد
Page tagging	برچسب‌زنی صفحه
Page view	بازدیدهای صفحه
Patentometrics	حق امتیازسنجی
Pearson Rank	رتبه پیرسون
Peer review	هم‌ترازخوانی
Plum analytics	پلوس وان
Podcast	پادکست
Privacy	حریم خصوصی
Project Gutenberg	پروژه گوتنبرگ
Public engagement	مشارکت عمومی
Public health investigations	بررسی سلامت عمومی
Publish or Perish	انتشار یا نابودی
Ranganathan's Laws of Library	قوانین علوم کتابداری رانگاناتان
RDF	چارچوب توصیف منابع
Reference managers	مدیریت منابع
Relational	ارتباطی
Research Excellence Framework	چارچوب ارتقای پژوهش
ResearchGate	ریسرچ گیت
Retweet Rank	رتبه ریتوییت‌رنک
Robots exclusion standard	استاندارد حذف ربات‌ها

San Francisco Declaration Research Assessment	اعلان ارزیابی پژوهش سان فرانسیسکو
Science Citation Index	نمایه استنادی علوم
Scientometrics	علم‌سنجی
Scopus	اسکوپوس
ScraperWiki	اسکرپرویکی
Search engine optimization	بهینه‌سازی موتور جستجو
Search engines	موتورهای جستجو
Semantic web	وب معنایی
Sentiment140	سنتیمنت۱۴۰
Sentiment analysis	تحلیل معنایی
Sindice	سیندیک
SildeShare	اسلایدشیر
Smart meters	سنجه‌های هوشمند
Soccernomics	بازار فوتبال
Social media	رسانه‌های اجتماعی
Social network analysis	تحلیل شبکه‌های اجتماعی
Spam bots	پیام نگار ناشناس
SPARQL	اسپارکل
Spearman's rank	رتبه اسپیرمن
Standardization	استانداردسازی
Statistical Cybermetrics Research Group	گروه پژوهشی سایبرسنجی آماری
Structured content	محتوا ساختاریافته
Tacit knowledge	دانش تلویحی
Terminology	واژگان
Text mining	متن کاوی
Tweet Beep	توییت‌بیپ
Tweet Grader	رتبه‌دهنده توییت
Twitter	توییتر
Undiscovered public knowledge	دانش عمومی کشف نشده
Union catalogue	فهرست کتابخانه‌ها
Universal library	کتابخانه جهانی

URL anatomy	آناتومی نشانی وبی
URL citations	استنادهای نشانی وبی
URL shorteners	کوتاه‌کننده‌های نشانی وبی
US National Science Foundation	مؤسسه علوم ملی ایالات متحده
Web 2.0	وب ۲/۰
Web analytics	تحلیل وب
Web Analytics Association	انجمن تحلیل دیجیتال
Web bibliometrics	کتاب‌سنجی وب
Web citation analysis	تحلیل استناد وب
Web connectivity model	مدل ارتباطی وب
Web crawling	وب پیمایی
Web impact assessment	ارزیابی تأثیر وب
Web impact factor	عامل تاثیر وب
Web metrics	سنجه‌های وب
Web of data	وب داده
Web of Knowledge	وب دانش
Wen of Science	وب علم
Webometrics Analyst	تحلیلگر وب‌سنجی
Webometrics	وب‌سنجی
Web scientometrics	علم‌سنجی وب
Wikipedia	ویکی‌پدیا
Wisdom of the crowd	خرد جمعی
Wordnet	وردنت
WorldCat	ورلدکت
Yahoo!	یاهو
YouTube	یوتیوب

واژه‌نامه فارسی به انگلیسی

ارتباطی	Relational
ارزیابی تأثیر وب	Web impact assessment
اسپارکل	SPARQL
استاندارد حذف ربات	Robots exclusion standard
استانداردسازی	Standardization
استخراج داده	Data extraction
استناد	Citation
استنادهای نشانی وبی	URL citations
استنادهای کتابخانه	Libcitations
اسکرپرویکی	ScraperWiki
اسکوپوس	Scopus
اسلایدشیر	SildeShare
اطلاع‌سنجی	Informetrics
اعلان ارزیابی پژوهش سان فرانسیسکو	San Francisco Declaration Research Assessment
انتشار یا نابودی	Publish or Perish
انتقادها	Criticisms
انجمن تحلیل دیجیتال	Web Analytics Association
انجمن کتابخانه آمریکا	American Library Association
اوپن دووار	OpenDAOR
اوپنلی لوکال	Openly Local
ایمپکت استوری	Impact story
اینتساگرام	Instagram
اینترنت‌سنجی	Internetometrics
آرشیو اینترنت	Internet Archive

Freedom of information	آزادی اطلاعات
URL anatomy	آناتومی نشانی وبی
Soccernomics	بازار فوتبال
Marketing	بازاریابی
Page view	بازدیدهای صفحه
Page tagging	برچسب‌زنی صفحه
Public health investigations	بررسی سلامت عمومی
Nature Linked Data Platform	بستر داده‌ها پیوندی طبیعی
Search engine optimization	بهینه‌سازی موتور جستجو
Bow tie	پاپیون
Podcast	پادکست
Natural language processing	پردازش زبان طبیعی
Project Gutenberg	پروژه گوتنبرگ
Plum analytics	پلوس وان
Spam bots	پیام نگار ناشناس
History	تاریخچه
Advertising	تبلیغ
Multiple regression analysis	تجزیه و تحلیل رگرسیون چندگانه
Citation analysis	تحلیل استنادی
Web citation analysis	تحلیل استنادی وب
Social network analysis	تحلیل شبکه‌های اجتماعی
Content analysis	تحلیل محتوا
Sentiment analysis	تحلیل معنایی
Web analytics	تحلیل وب
Webometrics Analyst	تحلیلگر وب‌سنجی
Graph density	تراکم نمودار
Bonferroni correction	تصحیح بونفرونی
Netometrics	تعامل علمی واسطه‌گری اینترنت
Dunbar number	تعداد دانبار
Definition	تعریف
Tweet Beep	توییت‌بیپ
Twitter	توییتر

جستجوی دانشگاهی مایکروسافت	Microsoft Academic Search
چارچوب ارتقای پژوهش	Research Excellence Framework
چارچوب توصیف منابع	RDF
حریم خصوصی	Privacy
حق امتیازسنجی	Patentometrics
حقوق مالکیت معنوی	Intellectual property rights
خدمات پستی ایالات متحده	US Postal Service
خرد جمعی	Wisdom of the crowd
خوشه	Cluster
داده	Data
داده‌های باز	Open data
داده‌های بزرگ	Big data
داده‌های عمومی گوگل	Google public Data
دانش تلویحی	Tacit knowledge
دانش عمومی کشف نشده	Undiscovered public knowledge
دانشمند الکترونیکی	e-scientist
دانشمند داده	Data scientist
دسترسی باز	Open access
دسته‌بندی	Categorization
دفتر کمیسیون اطلاعات	Information Commissioner's
دگرسنجه‌ها	Altmetrics
دوره‌های برخط باز فراگیر	MOOCs
دیجیتال‌سازی	Digitization
رتبه‌دهنده توییت	Tweet Grader
رتبه اسپیرمن	Spearman's rank
رتبه پیرسون	Pearson Rank
رسانه‌های اجتماعی	Social media
روندهای سرماخوردگی گوگل	Google Flu Trends
روندهای گوگل	Google Trends
ریتوییت‌رنک	Retweet Rank
ریسرچ گیت	ResearchGate
سازمان اینترنتی برای نام‌ها و شماره‌های تخصیصی	

Internet Corporation for Assigned Names and Numbers	
Cybermetrics	سایبرسنجی
Library of Congress Subject Headings	سرعنوان‌های موضوعی کتابخانه کنگره
Sentiment140	سنتیمنت۱۴۰
Evaluative	سنجشی
Web metrics	سنجه‌های وب
Smart meters	سنجه‌های هوشمند
Sindice	سیندیک
Indicators	شاخص‌ها
Economic indicators	شاخص‌های اقتصادی
h-index	شاخص اچ
g-index	شاخص جی
Cookies	شناسه شناسایی کاربر
Journal impact factor	ضریب تأثیر مجله
Nature	طبیعت
Web impact factor	عامل تاثیر وب
Web scientometrics	علم‌سنجی وب
Data science	علم داده
Little Science, Big Science	علم کوچک، علم بزرگ
scientometrics	علم‌سنجی
Open notebook science	علوم دفتر باز
Metadata	فرادادە
Blogosphere	فضای وب‌نوشت‌ها
Flickr	فلیکر
Foursquare	فورسکوئر
Union catalogue	فهرست کتابخانه‌ها
Library catalogues	فهرست کتابخانه
Facebook	فیسبوک
Bradford'd law	قانون برادفورد
Goodhart's law	قانون گودهارت
Ranganathan's Laws of Library	قوانین علوم کتابداری رانگاناتان
Chi-square	کای‌دو

کتاب‌سنجی وب	Web bibliometrics
کتاب–الکترونیکی	e-books
کتابخانه جهانی	Universal library
کتابخانه ملی اسپانیا	National Central Library of Spain
کتابخانه ملی فرانسه	National Central Library of France
کتابخانه ملی مرکزی رم	National Central Library of Rome
کتابخانه ملی مرکزی فلورانس	National Central Library of Florence
کتابخانه هاروارد	Harvard Library
کلیک	Click
کمسیون اروپا	European Commission
کوتاه‌کننده‌های نشانی وبی	URL shorteners
کتابخانه الکساندریا	Library of Alexandria
گاردین	Guardian
گروه پژوهشی سایبرسنجی آماری	Statistical Cybermetrics Research Group
گوگل	Google
لغت‌نامه انگلیسی آکسفورد	Oxford English Dictionary
لینکداین	LinkedIn
مادوله	Modularity
مای اسپیس	MySpace
مایکروسافت وورد	Microsoft Word
متن کاوی	Text mining
متون خاکستری	Grey literature
محتوا ساختاریافته	Structured content
مدل ارتباطی وب	Web connectivity model
مدیریت منابع	Reference managers
مرکزیت	Centrality
مشارکت عمومی	Public engagement
مکان‌یابی	Bookmarking
مواد طبیعی	Nature Materials
موتورهای جستجو	Search engines
موتورهای جستجوی کتابشناختی	Bibliographic search engines
مؤسسه علوم ملی ایالات متحده	US National Science Foundation

Microdata	میکروداده
Microformats	میکروفرمت
Micrographia	میکرونمودار‌یا
Journal of Visualized Experiments	مجله تجربه‌های بصری
Journal of Digital Humanities	مجله علوم انسانی دیجیتال
Harvard Business Review	نقد و بررسی کسب و کار هاروار
Science Citation Index	نمایه استنادی علوم
NodeXL	نودکسل
Terminology	واژگان
Webometrics	وب‌سنجی
Blog	وب‌نوشت
Blogs	وب‌نوشت‌ها
Blogger	وب‌نویس
Web 2.0	وب ۲/۰
Web crawling	وب پیمایی
Web of data	وب داده
Web of Knowledge	وب دانش
Wen of Science	وب علم
Semantic web	وب معنایی
Wordnet	وردنت
worldCat	ورلدکت
Wikipedia	ویکی‌پدیا
Dublin Core	هسته دوبلین
Ontologies	هستی‌شناسی‌ها
Open Annotation Ontology	هستی‌شناسی حاشیه‌نویسی باز
Bibliographic Ontology	هستی‌شناسی کتاب‌شناسی
Correlation	همبستگی
Peer review	هم‌ترازخوانی
Yahoo!	یاهو
YouTube	یوتیوب

نمایه

ج

چ

ح

خ

Web Metrics

for Library and Information Professionals

By:
David Stuart

Translated by:
Maryam Khosravi
Fatemeh Bahmani

Edited by:
Hamid Keshavarz

Iranian Research Institute for Information Science and Technology
(IRANDOC)

Chapar Publication

2018