

مقایسه میزان دقت راهنماهای موضوعی وب در بازیابی اطلاعات فنی-مهندسی

مهردخت وزیرپور کشمیری | استادیار گروه کتابداری و اطلاع‌رسانی، دانشگاه آزاد تهران شمال
بهاره فزونی* | کارشناس ارشد کتابداری و اطلاع‌رسانی

دریافت: ۱۳۸۹/۰۶/۲۹ | پذیرش: ۱۳۸۹/۱۰/۰۱

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۲۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS، LISA و ISC
http://ijim.irandoc.ac.ir
دوره ۲۷ | شماره ۳ | ص ص ۵۹۷-۶۱۶
بهار ۱۳۹۱
نوع مقاله: پژوهشی

* b.fozooni@gmail.com

چکیده: هدف این مقاله، مقایسه میزان دقت راهنماهای موضوعی وب در بازیابی اطلاعات فنی-مهندسی است. گردآوری اطلاعات براساس روش اسنادی و وب‌سنجی بوده است. کلیدواژه‌های تخصصی حوزه فنی-مهندسی از سایت تخصصی IEEE (Institute of Electrical and Electronics Engineers) و مجلات واقع در سایت ساینس دایرکت (science direct) انتخاب شدند. این کلیدواژه‌ها در راهنماهای موضوعی یاهو، گوگل، اینفوماین، اینتیوت، و دی‌موز که پرستفاده‌ترین راهنماهای موضوعی بودند، جستجو شدند. با توجه به اینکه به‌طور معمول، نتایج اولیه هر بازیابی، در ابزارهای کاوش ارتباط بیشتری با کلیدواژه مورد نظر دارند، بنابراین در هر جستجو ۱۰ نتیجه اول ارزیابی شد. این ارزیابی‌ها براساس میزان دقت، میزان خطا و نسبت اقلام فنی به کل اقلام بودند. معیارهای تعیین میزان دقت که براساس استانداردهای پرستفاده در مدارک مختلف بودند شامل: حضور کلیدواژه‌ها در عنوان، هم‌جواری کلیدواژه‌ها، بسامد کلیدواژه‌ها در صفحات وب بازیابی شده، رده موضوعی، نشانی اینترنتی، و توصیف سایت است. تجزیه و تحلیل اطلاعات براساس آزمون کروسکال والیس و L.S.D فیشر انجام شد. نتایج نشان داد که اختلاف معنی‌داری در میزان دقت راهنماهای موضوعی وب وجود دارد و فرضیه اول تأیید گردید. راهنماهای موضوعی وب از نظر دقت به این ترتیب رتبه‌بندی شدند: گوگل، یاهو، اینتیوت، دی‌موز، و اینفوماین. از نظر میزان خطای مشاهده شده، راهنماهای موضوعی یاهو، دی‌موز، اینتیوت، گوگل، و اینفوماین به ترتیب کمترین میزان خطا را داشتند. در این پژوهش، مقایسه‌ای بین نسبت اقلام بازیابی شده در رده فنی-مهندسی با اقلام بازیابی شده از کلیدواژه‌ها در همه رده‌های موضوعی انجام شد و فرضیه دوم یعنی وجود اختلاف معنی‌دار بین این راهنماها تأیید گردید. همچنین، راهنماهای اینتیوت، اینفوماین، یاهو، دی‌موز، و گوگل به ترتیب بیشترین نسبت اقلام فنی به کل را دارا بودند.

کلیدواژه‌ها: راهنماهای موضوعی وب، دقت، بازیابی اطلاعات، مقایسه، فنی-مهندسی

۱. مقدمه

با ظهور اینترنت به عنوان منبع عظیمی از اطلاعات، تغییرات شگرفی در عرصه علم اطلاع‌رسانی به وجود آمده است. در مدت زمان کوتاه، میلیاردها صفحه وب در اختیار جامعه بشری قرار گرفته است و هر لحظه نیز بر تعداد این صفحات افزوده می‌شود. بر همین اساس، به‌طور معمول جستجو و یافتن اطلاعات مفید در اینترنت مشکل بوده است و حتی کتابداران نیز از اوایل کار این شبکه از این مسأله لب به شکایت گشوده‌اند (جو کار و جو کار ۱۳۸۶).

ابزارهای جستجو انواع مختلفی دارند و هر دسته از آنها در مواردی چون عمق نمایه‌سازی صفحات وب، امکانات جستجو، رتبه‌بندی نتایج، و میزان اطلاعات توصیفی برای هر مدخل بازیابی شده، با هم متفاوت هستند (رضوانی ۱۳۸۳). این تفاوت‌ها نشان‌دهنده قابلیت‌ها و توانایی‌های متفاوت انواع ابزارهای جستجو هستند.

یکی از مهمترین خدمات وب بازیابی اطلاعات است که به وسیله انواع ابزارهای جستجو فراهم می‌شود. به علاوه، برای فراهم کردن ابزارهای پرسش و تعیین موقعیت اطلاعات خاص، ابزارهای جستجو، راهنماهای وبی را به کاربران پیشنهاد می‌کنند تا محتواهای سازماندهی شده را مشاهده کنند. در یک مقایسه، راهنماهای وب قابل قیاس با فهرست مندرجات یک کتاب هستند، چون منابع اطلاعاتی را مطابق با موضوعات شان تقسیم‌بندی می‌کنند. در این زمینه، رابط کاربری کانال مهمی برای انتقال اطلاعات است. یک رابط انعطاف‌پذیر باید با نیازهای اساسی کاربر مطابقت داشته باشد، به طوری که محیط راحتی را برای کاربران ایجاد نماید تا آنها بتوانند محتوای مناسب خود را به آسانی تشخیص دهند، یعنی آزادانه در اطراف جستجو کنند و نتایج را به دقت بررسی نمایند. مطالعه جستجوی اطلاعات نشان داده است که رابط کاربری عاملی است که بر توان عملیاتی کاربران به ویژه سرعت و دقت در پیدا کردن اطلاعاتی خاص، تأثیر بسزایی دارد (Chen 2005).

به کارگیری ابزارهای جستجو در اینترنت در زمینه‌های موضوعی مختلف بدون داشتن آگاهی از توانایی‌های این ابزارها چندان سودمند نیست. تعدادی از پژوهشگران با انجام پژوهش‌هایی در این زمینه، اقدام به شناسایی و معرفی برترین ابزارها به کاربران کرده‌اند. پژوهش حاضر نیز قصد دارد تا این هدف را در رابطه با راهنماهای موضوعی وب دنبال کند. این پژوهش به طور خاص در زمینه موضوعی فنی-مهندسی است که به دلیل کاربرد زیاد جستجو در میان کاربران این رشته می‌تواند مفید باشد.

۲. بیان مسأله

رشد و گسترش روزافزون شبکه جهانی وب، تغییر و تحول شگرفی در فرایند اطلاع‌یابی به‌وجود آورده است. اطلاعات از طریق صفحات وب منتشر می‌شوند و قابل دسترس هستند. برای دسترسی به این اطلاعات باید از محل استقرار آنها آگاه بود. برای حل این مشکل، ابزارهای جستجو در اینترنت پدید آمده‌اند که مجموعه‌ای از اطلاعات منظم‌شده صفحات وب را در بردارند. ابزارهای جستجو در اینترنت به چند گروه عمده تقسیم می‌شوند: راهنماهای موضوعی^۱ و موتورهای کاوش^۲ و ابرموتورهای کاوش^۳.

متخصصان فهرست‌نویسی و علوم کتابداری بر اهمیت طبقه‌بندی مختلف و به‌کارگیری آنها برای فهرست‌نویسی سایت‌های اینترنتی تأکید دارند. خوانندگان مختلفی در یک کتابخانه یا مرکز منابع اسنادی وجود دارند که نیازهای آنها می‌تواند تعیین اولویت و مشخص شود و طبقه‌بندی سایت‌ها براساس این نیازها صورت پذیرد. دو محیط برای توسعه بیشتر راهنماهای وب که با خواسته‌های موجود جایگزین شده‌اند، ایجاد شده است: راهنماهای منطقه‌ای که به‌وسیله بخشی از فعالیت، دامنه جغرافیایی یا فرهنگی، خاص شده‌اند و دیگری نظام‌هایی که براساس ارزیابی گروهی، جمعی و فهرست‌نویسی وب‌سایت‌هایی در بین جوامع علاقه‌مند وجود دارند (Assadi and Bauvisage 2002).

استفاده‌کنندگان این نظام‌ها، آگاهانه یا غیر آگاهانه، نظام را برحسب اینکه تا چه حد نیازهای اطلاعاتی آنان را تأمین می‌کند، ارزیابی می‌کنند. استفاده‌کنندگان به‌طور معمول به چگونگی کارکرد نظام علاقه‌مند نیستند، بلکه فقط به نتایج نهایی توجه دارند.

از آنجایی که تعداد زیادی از این کاربرها را پژوهشگران و استادان و دانشجویان رشته‌های فنی-مهندسی تشکیل می‌دهند و با توجه به نقش محوری این حوزه در تولید دانش و توسعه ناشی از آن، بازایی اطلاعات و دسترسی پژوهشگران به دانش و مطالب روز که در این زمینه در مراکز علمی و فنی سراسر دنیا تولید و ارائه می‌شود، از اهمیت ویژه‌ای برخوردار است. بنابراین، در این پژوهش سعی شده است از طریق بررسی راهنماهای موضوعی وب، با درج کلیدواژه‌های تخصصی پرسامد که از طریق سایت IEEE و مقالات تخصصی موجود در سایت ساینس دایرکت^۴ و مشورت با متخصصان این امر انتخاب شده است و رعایت اصول جستجو در هریک از آنها، میزان دقت و یافتن مرتبط‌ترین اطلاعات در این راهنماها مورد ارزیابی قرار گیرد. بدین طریق

1. Web directories
2. Search engines
3. Meta search engines
4. Science direct

می‌توان راهنماهای موضوعی مورد مطالعه را رتبه‌بندی کرد و به این کاربران در انتخاب راهنماهای مناسب برای دسترسی هرچه سریع‌تر به منابع مورد نظر و دریافت دقیق‌ترین نتیجه یاری رساند.

۳. اهمیت و ضرورت انجام پژوهش

با انجام این پژوهش مشخص خواهد شد که کدام‌یک از راهنماهای موضوعی وب در حوزه فنی- مهندسی دقت بیشتری دارند و منابع مرتبط‌تری را بازیابی می‌کنند. بر این اساس، راهنماهای موضوعی وب رتبه‌بندی خواهند شد و متخصصان این رشته به بهترین ابزار جستجو مراجعه می‌کنند و دقیق‌ترین نتایج را دریافت خواهند کرد. این کار به صرفه‌جویی در وقت برای جستجو کمک زیادی خواهد کرد.

۴. اهداف مشخص پژوهش

هدف اصلی این پژوهش بررسی توانمندی راهنماهای موضوعی وب در بازیابی اطلاعات حوزه فنی- مهندسی است.

اهداف فرعی:

- ۱) تعیین بهترین راهنمای موضوعی وب از نظر دقت در بازیابی اطلاعات فنی- مهندسی؛
- ۲) تعیین بهترین راهنماهای موضوعی وب از نظر کمترین میزان خطا در بازیابی اطلاعات فنی- مهندسی؛
- ۳) تعیین بهترین راهنمای موضوعی وب از نظر بازیابی اطلاعات در رده خاص فنی- مهندسی؛ و
- ۴) صرفه‌جویی در وقت کاربران برای دسترسی هرچه سریع‌تر به منابع مرتبط تر.

۵. پرسش‌های اساسی

- ۱- آیا از لحاظ دقت در بازیابی مدارک، بین راهنماهای موضوعی وب، اختلاف معنی‌داری وجود دارد؟
- ۲- کدام‌یک از راهنماهای موضوعی وب در زمینه فنی- مهندسی اطلاعات دقیق‌تری را بازیابی می‌کنند؟
- ۳- کدام‌یک از راهنماهای موضوعی وب کمترین میزان خطا را در بازیابی اطلاعات دارند؟
- ۴- آیا اقلام بازیابی شده از کلیدواژه‌ها، در همه رده‌های موضوعی راهنماهای مورد نظر، به‌طور دقیق منطبق با اقلام بازیابی شده در رده فنی- مهندسی است؟

۶. فرضیه پژوهش

- ۱) از لحاظ میزان دقت راهنماهای موضوعی وب در بازیابی اطلاعات فنی- مهندسی، اختلاف معنی داری وجود دارد.
- ۲) بین اقلام بازیابی شده در رده فنی-مهندسی با اقلام بازیابی شده در همه رده های راهنماهای موضوعی مورد نظر اختلاف معنی داری وجود دارد.

۷. متغیرهای مورد بررسی

متغیر مستقل: راهنماهای موضوعی وب
متغیر وابسته: میزان دقت بازیابی اطلاعات در حوزه فنی- مهندسی

۸. روش پژوهش

در این پژوهش، از دو روش متفاوت استفاده شده است: برای پیدا کردن راهنماهای موضوعی وب که کاربرد بیشتری دارند و نیز معیارهایی که راهنماهای موضوعی وب از آن استفاده می کنند تا دقت خود را در بازیابی مدارک بالا ببرند از روش سندی یا کتابخانه ای و برای سنجش میزان دقت راهنماهای موضوعی وب از روش وب سنجی استفاده شده است. ضریب اعتبار این آزمون که بر پایه فرمول کلی آلفای کرونباخ محاسبه گردید، برابر ۰/۷۵۶ است. این مقدار نشان می دهد معیارهای مورد بررسی از هماهنگی درونی کافی برخوردار است. در نتیجه، ضریب پایایی در این پژوهش متوسط و قابل قبول بوده است. همچنین، پایایی آن را شاخص ها و آزمون های به کار رفته در این پژوهش که کارآمد و مؤثر و قابل استفاده در دیگر حوزه های موضوعی هستند، تأمین می کند.

روشی که برای تعیین روایی این پژوهش انجام شد، روش روایی محتوی است. بدین منظور در این پژوهش، کلیدواژه ها و معیارهای انتخاب شده جهت ارزیابی، توسط متخصصان امر مورد بررسی قرار گرفتند. به علت انجام ارزیابی دقیق، پیش آزمون صورت گرفت و نظرات متخصصان، اساتید و مشاوران آماری منجر به بهتر شدن روند پژوهش گشت، در نتیجه دارای اعتبار لازم است. در پژوهش حاضر، با پیمایش جامعه راهنماهای موضوعی وب تلاش شد به بررسی و مقایسه این راهنماها از نظر میزان دقت در بازیابی اطلاعات فنی-مهندسی پرداخته شود. این میزان را معیارهای مطرح شده در پژوهش و مشورت با متخصصان امر تعیین می کند.

۹. جامعه آماری

جامعه آماری این پژوهش شامل راهنماهای موضوعی وب یا هو، اینفوماین، گوگل،

اینترنت، و دی موز است.

Yahoo (<http://www.dir.Yahoo.com>)
 Google (<http://www.directory.google.com>)
 Dmoz (<http://www.dmoz.org>)
 Infomine (<http://infomine.ucr.edu/>)
 Intute (<http://www.intute.ac.uk>)

این ابزارهای کاوش در دسته راهنماهای موضوعی، جزء پر استفاده ترین آنهاست. برای تعیین راهنماهای موضوعی وب پر استفاده، پیشینه پژوهش بررسی شد و ۵ راهنمای موضوعی که بیشترین حضور را داشتند، برای این پژوهش انتخاب شدند.

برای واقعی بودن کلیدواژه‌های به کار رفته در جستجو، جامعه مورد مطالعه در این بخش را مقالات تخصصی رشته‌های فنی- مهندسی موجود در سایت (IEEE) Institute of Electrical and Electronics Engineers و Sciencedirect تشکیل می‌دهد. کلیدواژه‌های این جامعه شامل ۲۰ کلیدواژه در حوزه فنی- مهندسی است. با توجه به پیشینه‌های مورد مطالعه و پژوهش‌های صورت گرفته در آنها مانند کشاورزی که دارای گرایش‌های مختلفی است و نیز کتابداری و اطلاع‌رسانی و سایر رشته‌ها به‌طور معمول ۱۰ کلیدواژه مورد بررسی قرار می‌گرفته است. همچنین، بررسی ۲۰ کلیدواژه در ۵ راهنمای موضوعی و تحلیل ۱۰ نتیجه اولیه بازیابی شده در هر کلیدواژه، منجر به بررسی ۱۰۰۰ نتیجه شد. انتخاب این ۲۰ کلیدواژه با مشورت با اساتید مربوط انجام شد.

۱	Interconnected + power + systems
۲	Spectrum sensing
۳	"Electromagnetic radiation" + effects
۴	"millimeter wave" + switches
۵	"VHF radio communication"
۶	"Computational Fluid Dynamics" or "CFD"
۷	"Structural Health Monitoring"
۸	hybrid + electric + vehicle
۹	"heat transfer" + "coefficient"
۱۰	Micro pump
۱۱	"case hardening"
۱۲	heat + treatment
۱۳	"Internet Communication"
۱۴	Internet + Content Management
۱۵	Network + Security
۱۶	Communications Protocols
۱۷	Design + irregular buildings
۱۸	High performance concrete
۱۹	"Satellite Navigation Systems"
۲۰	Inorganic + substances

۱۰. روش تجزیه و تحلیل اطلاعات

برای پاسخ به پرسش‌های اساسی این پژوهش از آزمون کروسکال والیس^۱ استفاده شده است. این آزمون برای مقایسه سه یا بیش از سه گروه مستقل که دست کم در سطح رتبه‌ای اندازه‌گیری می‌شوند، مورد استفاده قرار می‌گیرد (میرزایی ۱۳۸۸). برای رتبه‌بندی راهنماها نیز از آزمون کمترین تفاوت معنی‌داری (L.S.D) فیشر استفاده شده است. همچنین، برای تجزیه و تحلیل داده‌ها از نرم‌افزار Stat Graph و SPSS استفاده شده است.

۱۱. پیشینه پژوهش

در داخل کشور بر روی میزان دقت موتورهای کاوش و ابرموتورهای کاوش کارهایی صورت گرفته است، اما کار منسجم و متمرکزی درباره راهنمای موضوعی دیده نمی‌شود. البته در رابطه با معیارهای محتوی و کارایی و توانمندی و همچنین، مقایسه چند راهنما کارهای بسیار کمی صورت گرفته است. در خارج از کشور نیز بیشتر در رابطه با ساختار راهنماهای موضوعی و نحوه نمایه‌سازی آنها پژوهش‌هایی انجام شده است.

۱-۱۱. پژوهش در ایران

کميجانی (۱۳۸۱) در پایان‌نامه خود با عنوان "مقایسه کارایی موتورهای کاوش وب در بازیابی اطلاعات کشاورزی" هفت موتور کاوش عمومی وب شامل آلتاویستا، اسکایت، هات‌بات، اینفوسیک، گوگل، لایکاس، و یاهو و دو موتور کاوش تخصصی کشاورزی وب شامل آگری‌سرف^۲ و وب‌آگری^۳ را در بازیابی اطلاعات تخصصی حوزه کشاورزی با هم مقایسه کرده است. این مقایسه نشان داد که در معیار ارتباط و اعتبار، سه موتور کاوش گوگل و لایکاس و یاهو به ترتیب رتبه اول تا سوم را داشتند و آخرین رتبه‌ها به موتورهای کاوش تخصصی کشاورزی تعلق داشت. این پژوهش نشان داد که موتورهای کاوش عمومی در هر دو معیار ارتباط و اعتبار که به نظر می‌رسد به ترتیب مهمترین معیارهای مورد توجه از طرف کاربران متخصص کشاورزی باشند، با اختلاف چشمگیری بر موتورهای کاوش تخصصی کشاورزی برتری دارند.

آزادی (۱۳۸۲) در پایان‌نامه خود با عنوان "میزان دقت موتورهای کاوش وب در بازیابی اطلاعات کتابداری و اطلاع‌رسانی" هفت موتور کاوش وب را از لحاظ دقت و میزان استفاده از برنشانه کلیدواژه در صفحات وب مورد مقایسه قرار داده است. براساس یافته‌های او، موتور

1. Kruskal-Wallis Test
2. Agrisurf
3. Web-agri

کاوش اینفوسیک^۱ و وب کراولر^۲ به ترتیب بالاترین و پایین ترین میزان دقت در بازیابی اطلاعات مرتبط با این حوزه را نشان داده اند. همچنین، از لحاظ میزان استفاده از ابرنشانه کلیدواژه، موتور کاوش هاتبات^۳ و وب کراولر به ترتیب بیشترین و کمترین میزان استفاده از این ابرنشانه را در صفحات وب داشته اند.

مهراد و بیش (۱۳۸۴) در مقاله خود با عنوان "ارزیابی سایت های وب بازیابی شده توسط دو راهنمای شبکه وب اوپن دایرکتوری و یاهو (Yahoo و ODP) در حوزه علوم کتابداری و اطلاع رسانی براساس معیارهای محتوی و کارایی" به بررسی ۴۲ شاخص در زیر گروه های صحت، اعتبار، جدید بودن، کیفیت نوشته، کاربرپسندی، قابلیت جستجو، و قابلیت مرور، براساس سیاهه های واری در دسترس پرداختند و سپس، کیفیت محتوی و کارایی سایت های وب بازیابی شده در حوزه علوم کتابداری و اطلاع رسانی را در چهار مقوله بازیابی اطلاعات، مدیریت دانش، راهنماها، و مراکز پژوهشی در دو راهنمای یاهو و طرح راهنمای باز^۴ براساس این سیاهه واری مورد مقایسه قرار دادند. نمونه مورد پژوهش شامل ۱۴ سایت از راهنمای یاهو و ۲۸ سایت از راهنمای طرح راهنمای باز است که به روش تصادفی نظام یافته انتخاب گردیده است. روش مورد استفاده برای انجام این پژوهش که توصیفی و از نوع کاربردی است، تحلیل محتوی بوده است. این پژوهش نشان داد که به جز در مورد شاخص های "زمان ایجاد سایت"، "روزآمد بودن پیوندها"، "وجود HELP"، "امکان ایجاد تغییر در اندازه حروف متن"، "مشخص بودن جعبه جستجو"، و "امکان جستجو در سایت" در مورد سایر شاخص ها در راهنمای یاهو دارای وضعیت بهتری نسبت به راهنمای اوپن دایرکتوری است. اما، اختلاف معنی داری بین این دو راهنما وجود ندارد. در خاتمه نیز سیاهه واری تدوین شده در این پژوهش که به عنوان ابزار گردآوری داده ها مورد استفاده قرار گرفته، برای طراحی و نیز ارزیابی محتوی و کارایی سایت ها توصیه شده است.

جوکار و جوکار (۱۳۸۶) در مقاله خود با عنوان "مقایسه راهنماهای یاهو^۵، اینفوماین^۶، و لی^۷ از نظر ویژگی های اقلام بازیابی شده در حوزه علوم تربیتی و روان شناسی" به مقایسه بازیابی اقلام مرتبط این سه راهنما پرداخته اند. نتایج حاصل نشان دهنده این است که گرچه راهنمای یاهو تعداد اقلام بیشتری را نسبت به دو راهنمای دیگر بازیابی می کند، اقلام بازیابی شده توسط دو راهنمای اینفوماین و لی به ترتیب از نظر ربط موضوعی در جایگاه بهتری نسبت به مدارک بازیابی شده توسط یاهو قرار دارد.

1. Infoseek
2. Webcrawler
3. Hotbot
4. ODP (Open Directory Project)
5. Yahoo
6. Infomine
7. LIJ= Librarians Index to Internet

۱۱-۲. پیشینه در خارج

اسدی و بوویژاژ در مقاله خود با عنوان "مطالعه تطبیقی در شش راهنمای وب به زبان فرانسه"^۱ به مقایسه راهنماهای وب ام‌اس‌ان^۲، نماد^۳، اوپن دایرکتوری^۴، وویلا^۵، وویلاپیج پرسو^۶، و یاهو پرداخته‌اند. این مطالعه بر روی کیفیت (تحلیل دستی راهنماها و تحلیل‌های انتقادی از روش‌های سازماندهی) و کمیت (براساس فرایندهای آماری و رسمی) جنبه‌هایی از سازماندهی آن راهنماها و روش‌هایی که آن سایت‌ها را تشریح می‌کنند، متمرکز شده است. این مقاله انواع گسترده‌ای از ساختارها و محتوی و اصول سازماندهی را آشکار می‌کند، اما به نظریه کلاسیک طبقه‌بندی پاسخ نمی‌دهد، بلکه مشکلات طرح ساختاریافته محتوای نامتجانس وب را مورد بررسی قرار می‌دهد. این پژوهش، ابتدا یک بسته نرم‌افزاری خاص را به منظور جستجوی ساختاری و محتوایی راهنماها بررسی می‌کند (پیوندهای سلسله مراتبی و ارجاع‌های متقابل بین طبقات، آدرس‌های فهرست‌شده و توصیف‌های سایت). سپس براساس این اطلاعات، یک تحلیل کیفی از سازماندهی هر راهنما ارائه می‌دهد. در این پژوهش و از نظر ساختار درختی، ارجاع متقابل، همپوشانی بین راهنماهای مختلف یاهو بیشترین امتیاز را داراست (Assadi and Bauvisage 2002).

کریکاس، استامو، و کگسیس، در مقاله خود با عنوان "رتبه راهنما: نظم و ترتیب صفحات در راهنماهای وب"^۷ به ارائه چارچوب منظمی از چگونگی آموزندگی صفحات میان موضوعی داده‌شده، پرداختند. این پژوهش در سه مرحله انجام شده است: ابتدا صفحات وب را میان یک موضوع، به منظور ساختار دقیق که زنجیره‌های منطقی نامیده می‌شوند، پیش می‌برد، سپس به میزان آموزندگی یک صفحه در یک موضوع خاص می‌پردازد. سرانجام دو مرحله، برای مرتب کردن صفحه‌های یک موضوع قبل از ارائه به استفاده‌کنندگان، با هم ترکیب می‌شوند. این پژوهش اندازه‌گیری رتبه راهنماها را به منظور یک جفت از صفحات که در ۱۵۶ موضوع در راهنمای گوگل فهرست شده است، به کار می‌برد. آنها رتبه‌بندی ایجادشده رتبه راهنما را با رتبه صفحه که در راهنمای گوگل برای صفحه‌ها و موضوع‌های داده شده است، مقایسه کردند. بررسی در مورد ۱۵ استفاده‌کننده‌ای که از آن استفاده می‌کردند، تکیه داشت. نتایج به‌دست آمده نشان داد که استفاده‌کنندگان زیادی رتبه راهنما را برای نظم دادن صفحات وب داخل موضوع‌های راهنما بر رتبه صفحه ترجیح می‌دهند (Krikos, Stamou, and Kokosis 2005).

1. A Comparative Study of Six French-Language Web Directories
2. MSN
3. Nomade
4. Open Directory
5. Voila
6. Voila Pages Perso
7. Directory Rank: Ordering Pages in Web Directories

ماسمان و رام در مقاله خود با عنوان "ارزیابی تطبیقی راهنماهای موردی وب"^۱ به ارزیابی میزان فناوری تطبیقی که می‌تواند برای تعیین خودکار راهنماها مورد استفاده قرار گیرد، پرداختند. هدف آنها یافتن روش‌های بهتر فنون خاص تطبیق براساس مورد، با استفاده از نشانی سایت و نام و توصیف وب‌سایت‌های دسته‌بندی شده است. روش‌هایی بر مبنای موارد خاص برای مقیاس‌های مشابه مختلف ارزیابی کرده و ترکیب آنها را از طریق فراداده مورد مطالعه قرار داده‌اند. مقایسه‌ای را در چهار راهنمای وب در فروش اینترنتی که عبارت بودند از دی‌موز^۲، گوگل، وب، و یاهو انجام داده‌اند. با وجود شباهت در حد متوسط و همپوشانی وب‌سایت راهنماها، مشکلات تطبیق می‌تواند در مقیاس بالایی حل شود (Massmann and Rahm 2008).

کپتین و کمپس در مقاله خود با عنوان "راهنماهای وب به‌عنوان محتوای موضوعی"^۳ به بررسی اینکه آیا می‌توان از راهنمای دی‌موز برای طبقه‌بندی پرسش‌ها از رده‌بندی‌های موضوعی در سطوح مختلف استفاده کرد و آیا می‌توان از این چهارچوب موضوعی برای بهبود عملکرد بازیابی استفاده کرد، پرداختند. در این مقاله، به کاربرهای مورد آزمون این امکان داده شد تا به‌وضوح پرسش‌های رده‌های موضوعی را رده‌بندی کنند. رده‌ها از دی‌موز یا از فهرست پیشنهادهایی که توسط فنون رده‌بندی موضوعی به‌طور خودکار ایجاد شده‌اند، انتخاب می‌شدند. نتایج این مطالعه نشان داد که رده‌بندی‌های دی‌موز برای طبقه‌بندی‌های موضوعی مناسب هستند. جستجوهای آزاد و همچنین ارزیابی فهرست پیشنهادها برای ایجاد چارچوب موضوعی به کار گرفته شد. زمانی که رده‌های موضوعی انتخاب شده از طریق جستجوی آزاد به‌عنوان چارچوب موضوعی به کار گرفته شد، بهبود نتایج بازیابی شده مشهود گردید (Kaptien and Kamps 2009).

۱۲. تجزیه و تحلیل اطلاعات

برای جوابگویی به پرسش‌های اساسی پژوهش و در مجموع رسیدن به هدف پژوهش باید مجموع امتیازاتی که راهنماهای موضوعی وب در بازیابی اطلاعات فنی - مهندسی طی ۱۰ جستجو به‌دست آورده‌اند، با هم مقایسه شوند. در این پژوهش، در هر یک از ۲۰ کلیدواژه مورد جستجو در پنج راهنمای موضوعی ۱۰ نتیجه اول مورد بررسی قرار گرفت (از آنجایی که نتایج اولیه هر بازیابی، در ابزارهای کاوش ارتباط بیشتری با کلیدواژه مورد نظر دارند، در هر جستجو ۱۰ نتیجه اول ارزیابی شد). این امتیازات از حضور کلیدواژه‌ها در عنوان مدرک، هم‌جواری کلیدواژه‌های مورد جستجو، و میزان حضور یا بسامد کلیدواژه‌ها در متن مدرک، رده موضوعی، توصیف سایت، و نشانی اینترنتی تشکیل شده‌اند.

1. Evaluation Instance-based Matching of Web Directories
2. Dmoz
3. Web Directories as Topical Context

برای محاسبه امتیاز دقت پنج راهنمای موضوعی وب در بازیابی اطلاعات با توجه به داده‌های به‌دست‌آمده در ستون‌های بسامد، حضور و هم‌جواری و اختلاف فاحش آنها با بقیه داده‌ها از نظر مقدار، این اطلاعات کدگذاری شده است.

۱۳. پاسخ به پرسش‌های اساسی پژوهش و آزمون فرضیه‌ها

پرسش اساسی ۱: آیا از لحاظ دقت در بازیابی مدارک، بین راهنماهای موضوعی وب، اختلاف معنی‌داری وجود دارد؟

شیوه امتیازدهی به این ترتیب بود که ابتدا کلیدواژه‌های انتخابی با به‌کارگیری امکانات جستجو و عملگرهای مختلف در هریک از راهنماهای موضوعی جستجو شد. ۱۰ نتیجه اول از کل نتایج بازیابی‌شده توسط هر راهنمای موضوعی برای ارزیابی و مقایسه در نظر گرفته شد. ابتدا، عنوان مدرک بازیابی‌شده، بررسی شد. اگر کلیدواژه‌ای که برای جستجو انتخاب شده بود در عنوان مدرک موجود بود به تعداد حضور آنها یک امتیاز تعلق می‌گرفت. بعضی از کلیدواژه‌ها به‌صورت عبارتی و ترکیبی بودند. اگر کلیدواژه‌ها در عنوان مدرک به‌همان صورت ترکیبی یا عبارتی ظاهر می‌شد گذشته از امتیاز حضور در عنوان، یک امتیاز هم‌جواری نیز دریافت می‌کرد. برای اینکه تراکم کلیدواژه‌ها در مدرک مشخص‌گردد، تعداد کلیدواژه‌ها در صفحات آن مدرک شمارش می‌شد. در قسمت رده موضوعی، هنگام جستجوی کلیدواژه‌ها، در بعضی از راهنماها مراحل رده‌بندی موضوعات نشان داده می‌شد که وجود این رده در هریک از این راهنماها امتیاز یک در غیر این صورت امتیاز صفر دریافت می‌کرد. در قسمت توصیف سایت، توضیحی از سایت بازیابی‌شده درج شده است. درواقع، دامنه توصیف که ابزارهای کاوش در بیشتر موارد هنگام بازیابی سایت‌های مورد نظر بیان می‌کنند، دسترسی به توصیف‌های بیشتر در کنار هر مدرک بازیابی‌شده، امتیاز سه، توصیفات کوتاه، امتیاز دو؛ خیلی کوتاه، امتیاز یک، و بدون توصیف، امتیاز صفر را دریافت می‌کرد (۰ و ۱ و ۲ و ۳). در قسمت نشانی اینترنتی که نمایش نشانی سایت در هر کدام از مدارک بازیابی‌شده است، وجود نشانی، امتیاز یک (۱)؛ در غیر این صورت، صفر را دریافت کرد (۰ و ۱). امتیازاتی که این راهنماهای موضوعی (در قسمت تراکم کلیدواژه‌ها، تعداد آنها در صفحات نیز بررسی شد، یعنی در مورد هر کلیدواژه به صفحات ۱۰ نتیجه اول رجوع شد) در موارد مختلف به‌دست می‌آوردند با یکدیگر جمع شدند و مجموع امتیازاتی که بدین طریق حاصل شد، مبنای مقایسه قرار گرفت.

جهت پاسخگویی به این پرسش اساسی، پس از کدگذاری داده‌ها، همگن‌سازی آنها (به‌علت اینکه داده خام در قسمت حضور کلیدواژه در عنوان، هم‌جواری کلیدواژه‌ها، تراکم کلیدواژه در

صفحه با رده موضوعی، توصیف و نشانی سایت یکسان نبودند) صورت گرفت؛ برای همگن سازی آنها از کدگذاری استفاده شد. به عنوان مثال، (۰) همان صفر را دریافت کرد، ۱-۵، یک (۱)، ۶-۱۰ دو (۲)، ۱۱-۲۰ سه (۳)، بالاتر از ۲۱ چهار (۴) را دریافت کرد (البته تمام این کارها با صلاحدید متخصص آمار انجام گرفت). در این پژوهش، از آزمون کروسکال والیس جهت یافتن اختلاف معنی دار از نظر دقت بین پنج راهنمای موضوعی وب استفاده شد. آزمون کروسکال والیس یک آزمون ناپارامتری (توزیع آزاد) است که برای مقایسه سه یا بیش از سه گروه مستقل که دست کم در سطح رتبه‌ای اندازه گیری می‌شوند، مورد استفاده قرار می‌گیرد. این آزمون از طریق نرم افزار SPSS انجام پذیرفت (جدول ۱).

جدول ۱. نتایج آزمون کروسکال والیس

آماره توزیع خی دو	۹/۵۰۴
درجه آزادی	۴
احتمال معنی داری (P)	۰/۰۴۶۹۶

در این آزمون، داده‌ها از تمامی سطوح در ابتدا ترکیب و از کوچکتر به بزرگتر مرتب می‌شوند. سپس، برای داده‌ها در هر سطح رتبه میانگین محاسبه می‌شود. از آنجایی که مقدار احتمال (P) کوچکتر یا مساوی ۰/۰۵ است، اختلاف معنی داری در درجه اطمینان ۹۵٪ وجود دارد. به بیان دیگر، از نظر دقت در بازیابی مدارک بین پنج راهنمای موضوعی وب اختلاف معنی داری وجود دارد.

جدول ۲. جدول میانگین رتبه داده‌ها در آزمون کروسکال والیس

رتبه میانگین	اندازه نمونه	راهنمای موضوعی
۶۵/۴۷۵	۲۰	گوگل
۵۴/۰۵	۲۰	ياهو
۴۱/۸۷۵	۲۰	اینفوماين
۴۷/۷	۲۰	اینتیوت
۴۳/۴	۲۰	دی‌موز

با استناد به جدول‌های ۱ و ۲ و توضیحات مربوط به پرسش اساسی ۱، فرضیه ۱ تأیید می‌شود. به عبارتی، از نظر دقت در بازیابی مدارک بین پنج راهنمای موضوعی وب اختلاف معنی داری وجود دارد و این مطابق فرضیه ۱ پژوهش است.

پرسش اساسی ۲: کدام یک از راهنماهای موضوعی وب در زمینه فنی- مهندسی اطلاعات دقیق تری را بازیابی می کنند؟

با توجه به پاسخ پرسش ۱، از نظر دقت در بازیابی مدارک بین پنج راهنمای موضوعی وب اختلاف معنی داری وجود دارد. بنابراین، جهت مقایسه بین راهنماهای موضوعی و تعیین گروه های همگن و ناهمگن، از آزمون L.S.D فیشر استفاده شد. نتایج این آزمون در جدول ۳ آورده شده است.

جدول ۳. گروه های همگن از نظر دقت بازیابی اطلاعات

گروه های همگن	میانگین	تعداد	راهنمای موضوعی
*	۸/۰۵	۲۰	گوگل
*	۷/۱۵	۲۰	ياهو
	۵/۴۵	۲۰	اینفوماین
*	۵/۸۵	۲۰	اینٹیوت
	۵/۴۵	۲۰	دی موز

با توجه به داده های جدول ۳، راهنمای گوگل با فاصله به نسبت زیادی از دو راهنمای اینفوماین و دی موز قرار گرفته و راهنماهای یاهو و اینٹیوت نزدیک به هم قرار دارند. به عبارتی، در میان راهنماهای موضوعی پنجگانه، گوگل بالاترین میزان دقت را در بازیابی منابع تخصصی مربوط به فنی- مهندسی دارد. با توجه به جدول ۳، امتیازاتی که هر یک از راهنماهای موضوعی مورد بررسی با توجه به معیارهای از پیش تعیین شده کسب کردند، مشخص شد که راهنماهای موضوعی گوگل، یاهو، اینٹیوت، دی موز، و اینفوماین (به طور مشترک) به ترتیب رتبه های بالاتری به دست آوردند. با توجه به نتایج جدول ۳، برای جستجوی اطلاعات منابع فنی- مهندسی و دریافت نتایج دقیق تر، استفاده از راهنماهای موضوعی به ترتیب این اولویت پیشنهاد می شود: گوگل، یاهو، اینٹیوت، دی موز، و اینفوماین.

پرسش اساسی ۳. کدام یک از راهنماهای موضوعی وب کمترین میزان خطا را در بازیابی اطلاعات دارند؟

در بررسی ۱۰ نتیجه اول از هر جستجو در راهنمای موضوعی وب، اگر مدرکی بیش از یک بار تکرار می شد یک امتیاز منفی می گرفت. اگر پیوندی در میان مدارک بازیابی شده پیغام خطا می داد به این علت که آن صفحه وب تغییر نام داده یا نشانی آن تغییر کرده یا به هر علتی حذف شده بود و یا آنکه محل قرارگیری مدرک بر روی شبکه به درستی توسط راهنمای موضوعی ثبت نشده بود و امکان دسترسی به آن صفحه ممکن نبود، امتیاز منفی دریافت می کرد

(این موارد هر کدام پیغام خاص خود را دارد). مجموع امتیازات به دست آمده از این موارد، میزان خطای راهنمای موضوعی وب را در بازیابی اطلاعات نشان می‌داد. هرچقدر امتیاز به دست آمده از این موارد بیشتر می‌شد، میزان خطای بالا را در راهنمای موضوعی مورد نظر نشان می‌داد و هرچقدر تعداد امتیازات کمتر می‌شد، نشانگر میزان خطای پایین در راهنمای موضوعی مورد نظر بود. نسبت پیوندهای کور و پیغام‌های خطا در راهنمای موضوعی، میزان روزآمدسازی اطلاعات را در پایگاه‌های اطلاعاتی هریک از راهنماهای موضوعی و میزان کنترل آنها را بر این پیوندها نشان می‌دهد. بنابراین، هرچقدر میزان خطاها بیشتر باشد، روزآمدسازی و کنترل پایگاه‌های اطلاعاتی بر مدارک کمتر است و برعکس هرچقدر تعداد خطاها کمتر باشد، کنترل و روزآمدسازی در پایگاه‌های اطلاعاتی بهتر و سریع‌تر انجام می‌شود.

برای تعیین راهنمای موضوعی که کمترین خطا را در بازیابی اطلاعات داراست از آزمون L.S.D فیشر استفاده شد.

جدول ۴. گروه‌های همگن از نظر خطای بازیابی اطلاعات

گروه‌های همگن	میانگین	تعداد	راهنمای موضوعی
*	۰/۵	۲۰	گوگل
*	۰/۳	۲۰	ياهو
*	۱/۸	۲۰	اینفوماین
*	۰/۴۵	۲۰	اینتیوت
*	۰/۴	۲۰	دی‌موز

با توجه به نتایج به دست آمده در جدول ۴، راهنمای موضوعی اینفوماین با سایر گروه‌ها از نقطه نظر خطای بازیابی اطلاعات، همگن نبوده و دارای بیشترین خطاست. در پاسخ به سؤال اساسی مطرح شده می‌توان گفت از نظر آزمون L.S.D فیشر، چهار راهنمای موضوعی گوگل، یاهو، اینتیوت، و دی‌موز، گروه‌های همگن و دارای کمترین میزان خطا هستند، هرچند در بین آنها، یاهو با اختلاف ناچیزی کمترین خطا را داراست، یعنی یاهو، دی‌موز، اینتیوت، گوگل، و اینفوماین به ترتیب کمترین میزان خطا را دارند. به عبارتی، یاهو کمترین و اینفوماین بیشترین میزان خطا را دارا هستند. با توجه به جدول ۴، برای اینکه هنگام بازیابی اطلاعات با پیغام‌های خطا و پیوندهای کور و غیرفعال کمتری مواجه شد، به کارگیری راهنماها به ترتیب، یاهو، دی‌موز، اینتیوت، گوگل، و اینفوماین پیشنهاد گردید.

پرسش اساسی ۴: آیا اقلام بازیابی شده از کلیدواژه‌ها، در همه رده‌های موضوعی راهنماهای مورد نظر، به‌طور کامل منطبق با اقلام بازیابی شده در رده فنی- مهندسی است؟

جدول ۵. نتایج آزمون کروسکال والیس

راهنمای موضوعی	آماره توزیع خی	درجه آزادی	احتمال معنی‌داری (P)
گوگل	۸/۹۸۴	۱	۰/۰۰۳
ياهو	۱۹/۷۵۸	۱	۰
اینفوماین	۴/۳۲۰	۱	۰/۰۳۸
اینتیوت	۲/۰۴۸	۱	۰/۱۵۲
دی‌موز	۶/۵۲۹	۱	۰/۰۱۱

با توجه به نتایج به‌دست آمده و مقادیر احتمال، فقط در اقلام بازیابی شده توسط راهنمای موضوعی اینتیوت، اختلاف معنی‌داری در اقلام بازیابی شده به دو طریق اشاره شده وجود ندارد. بنابراین، در پاسخ به پرسش مطرح شده می‌توان گفت، اقلام بازیابی شده توسط تمامی راهنماهای موضوعی به‌غیر از راهنمای اینتیوت، در همه رده‌های موضوعی و در رده فنی- مهندسی، نامنطبق هستند.

به‌منظور مشخص شدن رتبه راهنماهای موضوعی پنجگانه از نقطه نظر نسبت اقلام بازیابی شده فنی به اقلام کل، پس از به‌دست آوردن نسبت این اقلام در راهنماهای موضوعی برای کلیدواژه‌های بیست‌گانه با استفاده از آزمون کروسکال والیس، وجود اختلاف معنی‌دار در اطلاعات به‌دست آمده مورد سنجش قرار گرفت. نتایج در جدول‌های ۶ و ۷ آمده است.

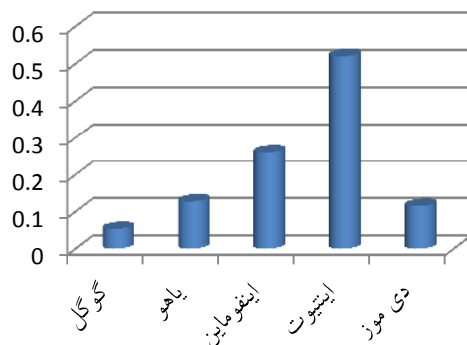
جدول ۶. میانگین رتبه‌های به‌دست آمده داده‌ها در آزمون کروسکال والیس

رتبه میانگین	اندازه نمونه	راهنمای موضوعی
۳۸/۳۷	۱۷	گوگل
۴۵/۴۳	۲۰	ياهو
۹۷/۴۶	۱۷	اینفوماین
۸۳/۶۱	۱۵	اینتیوت
۵۳/۲۶	۱۶	دی‌موز

جدول ۷. جدول نتایج کروسکال والیس

آماره توزیع خی	۱۷/۳۶۰
درجه آزادی	۴
احتمال معنی داری (P)	۰/۰۱۲

با توجه به مقدار P، اختلاف معنی داری بین راهنماهای موضوعی وجود دارد. جهت مقایسه بین راهنماهای موضوعی از حیث بازیابی کل اقسام فنی- مهندسی، آزمون کروسکال والیس انجام گرفت تا وجود اختلاف معنی دار بین اقسام بازیابی شده کل و اقسام بازیابی شده در رده فنی-مهندسی، در هر یک از راهنماهای موضوعی به طور جداگانه بررسی گردد. در نمودار ۱، نتایج آزمون که از طریق نرم افزار Stat Graph استخراج شده، نشان داده شده است.



نمودار ۱. گروه بندی راهنماهای موضوعی از نظر نسبت بازیابی شده فنی به اقسام کل

با توجه به نمودار ۱، راهنمای اینترنت با تمامی راهنماها اختلاف دارد. از طرفی، راهنمای گوگل علاوه بر راهنمای اینترنت با راهنمای اینفو ماین نیز اختلاف دارد. با توجه به این نتایج، راهنمای گوگل کمترین میزان نسبت اقسام فنی به کل و راهنمای اینترنت بیشترین میزان این نسبت را دارا هستند.

با استناد به جدول ۷ و پاسخ به پرسش اساسی ۴، می توان نتیجه گرفت که فرضیه ۲ تأیید می شود، به عبارتی اختلاف معنی داری بین اقسام بازیابی شده در رده فنی- مهندسی با اقسام بازیابی شده در تمام رده های راهنماهای موضوعی مورد نظر وجود دارد. بنابراین، هر چه اختلاف

معنی‌داری کمتر باشد، نمایه‌سازی بهتر صورت گرفته است. پس راهنماهای موضوعی اینتیوت، اینفوماین، یاهو، دی‌موز، و گوگل به‌ترتیب دارای نمایه‌سازی بهتری در حوزه فنی-مهندسی هستند. راهنماهای موضوعی عمومی گوگل و دی‌موز و یاهو به‌علت اینکه به‌طور مداوم اطلاعات زیادتری را دریافت می‌کنند، نمایه‌سازی آنها کندتر صورت می‌گیرد (چون روند نمایه‌سازی توسط انسان فرایندی وقت‌گیر است)، در حالی که راهنماهای موضوعی دانشگاهی اینتیوت و اینفوماین نسبت به راهنماهای موضوعی عمومی اطلاعات کمتری را دریافت می‌کنند، زمان بیشتری برای نمایه‌سازی دارند و سریع‌تر این نمایه‌سازی را انجام می‌دهند.

۱۴. تحلیل نتایج و مقایسه آن با پیشنهادها

بیشتر پژوهش‌های انجام‌شده در رابطه با راهنماهای موضوعی بر روی ساختار آنها و نحوه نمایه‌سازی آنهاست. راهنماهای موضوعی گوگل، یاهو، دی‌موز، اینفوماین، و اینتیوت از بهترین راهنماها هستند و پژوهش‌های زیادی نیز بر روی ساختار آنها انجام گرفته است. اما، در رابطه با مقایسه و بازیابی اطلاعات آنها کارهای کمتری به‌خصوص در داخل کشور صورت گرفته است. نمونه‌ای از کار انجام‌شده در داخل کشور مقایسه‌ای است که جوکار و جوکار (۱۳۸۶) بین راهنماهای موضوعی یاهو و اینفوماین و لی انجام داده و به این نتیجه رسیده‌اند که یاهو تعداد اقلام بیشتری را بازیابی می‌کند ولی اقلام بازیابی‌شده توسط دو راهنمای اینفوماین و لی، به‌ترتیب از نظر ربط موضوعی در جایگاه بهتری قرار دارند. همان‌طور که در این مقاله، راهنمای موضوعی گوگل از میزان دقت بیشتری نسبت به سایر راهنماها برخوردار است، ولی از نظر نسبت اقلام فنی به کل، اینتیوت و اینفوماین که راهنماهای دانشگاهی هستند، نسبت به راهنماهای موضوعی عمومی یاهو و دی‌موز و گوگل از وضعیت بهتری برخوردارند؛ به‌عبارتی نمایه‌سازی آنها بهتر انجام گرفته است. با وجود اینکه راهنمای موضوعی گوگل یک راهنمای عمومی است، در بازیابی اطلاعات فنی-مهندسی نسبت به سایر راهنماها در رابطه با میزان دقت امتیاز بیشتری را دریافت کرده است. راهنمای موضوعی گوگل از راهنماهای فعال در سایر زمینه‌های موضوعی نیز است. همان‌طور که کمیجانی (۱۳۸۱) در پایان‌نامه خود مقایسه‌ای بین هفت موتور کاوش عمومی (آلتاویستا، اکسایت، گوگل، هات‌بات، اینفوسیک، لایکاس، و یاهو) و دو موتور کاوش تخصصی در زمینه کشاورزی (اگری‌سرف و وب‌اگری) انجام داده و به این نتیجه رسیده است که موتورهای کاوش عمومی بهترین نتایج را دارند و گوگل رتبه اول را داشت. در رابطه با راهنمای گوگل، کارهای دیگری نیز صورت گرفته است از جمله پژوهشی که کریکاس، استامو، و ککسیس بر روی راهنمای گوگل در زمینه نظم و ترتیب صفحات در راهنماهای وب انجام داده‌اند و به ارائه

چارچوب منظمی از مقدار آموزندگی صفحات میان موضوعات داده شده در راهنمای گوگل پرداخته‌اند و به این نتیجه رسیده‌اند که استفاده کنندگان زیادی رتبه راهنما را بر رتبه صفحه برای نظم دادن صفحات وب داخل موضوعات راهنما ترجیح می‌دهند (Krikos, Stamou, and Kokosis) (2005).

در این پژوهش، راهنمای موضوعی یاهو نیز دارای پایین ترین خطا در بازیابی اطلاعات فنی-مهندسی است. یاهو نیز مانند گوگل از بهترین راهنماهای موضوعی عمومی است و پژوهش‌های زیادی بر روی آن انجام شده است مانند مطالعه تطبیقی که اسدی و بوویزاد بر روی شش راهنمای وب زبان فرانسه انجام داده‌اند و به این نتیجه رسیده‌اند که یاهو از نظر ساختار درختی، ارجاع متقابل، و همپوشانی بین راهنماهای مختلف بیشترین امتیاز را داراست (Assadi and Bauvisage) (2002).

۱۵. پیشنهادهای اجرایی

انتخاب درست ابزار جستجو و مشخص شدن هدف اصلی کاوش به بازیابی اقلامی مناسب و تأمین نیاز اطلاعاتی می‌انجامد. در این خصوص، براساس یافته‌های پژوهش، پیشنهادهای زیر ارائه می‌شود:

۱. برای جستجوی اطلاعات منابع فنی-مهندسی و دریافت نتایج دقیق‌تر، استفاده از راهنماهای موضوعی به ترتیب این اولویت پیشنهاد می‌شود: گوگل، یاهو، اینتیوت، دی‌موز، و اینفوماین.
۲. برای اینکه هنگام بازیابی اطلاعات با پیغام‌های خطا و پیوندهای کور و غیرفعال کمتری مواجه شویم، به کارگیری راهنماها به ترتیب، یاهو، دی‌موز، اینتیوت، گوگل، و اینفوماین پیشنهاد می‌گردد.
۳. راهنمای موضوعی اینتیوت در حوزه موضوعی فنی-مهندسی نسبت به راهنماهای دیگر بهتر نمایه‌سازی شده است، چون اقلام کلی بازیابی شده در این راهنما با اقلام بازیابی شده در رده خاص فنی-مهندسی آن اختلاف معنی‌داری ندارد. در نتیجه، استفاده از اینتیوت در قسمت راهنماهای موضوعی دانشگاهی مناسب است.

۱۶. پیشنهاد برای پژوهش‌های آینده

۱. پیشنهاد می‌شود میزان دقت راهنماهای موضوعی وب در بازیابی اطلاعات در حوزه‌ها و رشته‌های دیگر بررسی شود.
۲. پیشنهاد می‌شود در پژوهشی به مقایسه بین راهنماهای موضوعی وب و موتورهای کاوش در حوزه‌های مختلف علوم و بالاخص رشته فنی-مهندسی پرداخته شود.

۳. با توجه به راه‌اندازی موتور جستجوگر ملی در برنامه پنجم توسعه اقتصادی، اجتماعی و فرهنگی کشور، و با در نظر گرفتن موارد مطرح‌شده در پژوهش حاضر مانند نیازهای اطلاعاتی حوزه فنی-مهندسی و نقاط قوت و ضعف جستجوگرهای حاضر، پیشنهاد می‌شود پژوهشی در این زمینه در جهت نیل به این هدف با تکیه بر پژوهش حاضر انجام پذیرد.

۱۷. منابع

- آزادی، قاسم. ۱۳۸۲. میزان دقت موتورهای کاوش وب در بازیابی اطلاعات کتابداری و اطلاع‌رسانی. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران.
- ابراهیمی، مهدی. ۱۳۷۹. اینترنت. تهران: نشر کتابدار.
- جوکار، عبدالرسول، و طاهره جوکار. ۱۳۸۶. مقایسه راهنماهای یاهو، اینفوماین و لی از نظر ویژگی‌های اقلام بازیابی‌شده در
- حوزه علوم تربیتی و روان‌شناسی. فصلنامه کتاب،
www.noormags.com/View/Magazine/ViewPages.aspx? .v
- حسینی بهشتی، ملوک‌السادات. ۱۳۷۷. اصطلاحنامه نما: نظام مبادله اطلاعات علمی-فنی. تهران: مرکز اطلاعات و مدارک علمی ایران.
- حسینی بهشتی، ملوک‌السادات، مریم نوروزی اقبالی، و مهرداد نوروزی اقبالی. ۱۳۸۲. اصطلاحنامه فنی-مهندسی. تهران: مرکز اطلاعات و مدارک علمی ایران.
- خسروانی، میلاد. ۱۳۸۷. بهینه‌سازی صفحات وب بر پایه موتورهای جستجو. www.noormags.com (دسترسی در ۱۳۸۹/۲/۲).
- خسروی، فریبرز. ۱۳۸۰. اصطلاحنامه فرهنگی فارسی "اصفا". تهران: کتابخانه جمهوری اسلامی ایران.
- رضوانی گیل کلایی، شهلا. ۱۳۸۳. مقایسه هشت ابرموتور جست‌وجوی واقعی اینترنت در پاسخ‌گویی به سؤالات اختصاصی رشته کتابداری و اطلاع‌رسانی. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه شیراز.
- سلطانی، پوری، و فروردین راستین. ۱۳۶۵. اصطلاحنامه کتابداری. تهران: کتابخانه ملی جمهوری اسلامی ایران.
- شمس‌اژهای، محسن، و سیروس امیدفر. ۱۳۸۶. بررسی مؤلفه‌های مؤثر بر میزان بازیابی اطلاعات (جامعیت) و دقت بازیابی اطلاعات (مانعیت) در نظام‌های بازیابی اطلاعات وب-مدار. در همایش چالش‌های علم اطلاعات. www.lisseminar.blogfa.com (دسترسی در ۱۳۸۸/۸/۲۰).
- کمیجانی، احمد. ۱۳۸۱. مقایسه کارآیی موتورهای کاوش عمومی و تخصصی وب در بازیابی اطلاعات کشاورزی. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران.
- کومار، کریشان. ۱۳۷۴. روش‌های پژوهش در کتابداری و اطلاع‌رسانی. مترجمان فاطمه رهادوست و فریبرز خسروی. تهران: کتابخانه ملی جمهوری اسلامی ایران.
- کوشا، کیوان. ۱۳۸۱. پایگاه‌های اطلاعاتی. در *دایره‌المعارف کتابداری*. ج ۱. تهران: سازمان اسناد و کتابخانه ملی

جمهوری اسلامی ایران، ۵۵۹-۵۶۰.

کوشا، کیوان. ۱۳۸۱. *ابزارهای کاوش اینترنت: اصول، مهارت‌ها و امکانات جستجو در وب*. تهران: نشر کتابدار.
کوشا، کیوان. ۱۳۸۲. معیارهای ارزیابی موتورهای کاوش اینترنت: رویکردی متن‌پژوهی برای ارائه سیاهه واریسی. *اطلاع‌شناسی* ۱. www.noormags.com/View/Magazine/ViewPages.aspx?numberId=1305 (دسترسی در ۱۳۸۸/۸/۱۲).

کوشا، کیوان. ۱۳۸۳. طراحی راهنماهای موضوعی وب در محیط وب گامی در جهت تقویت شبکه اطلاع‌رسانی علمی، فرهنگی و ایران‌شناسی کشور Iranhoo تجربه طراحی راهنمای موضوعی سایت‌های ایران. در *مجموعه مقالات نخستین همایش ملی ایران‌شناسی: نسخه‌شناسی و کتابشناسی*. ۲۷۳۰: (۵۴) ۳۲۹-۳۴۹.
مهرآبادی، میترا. ۱۳۸۵. مقایسه موتورهای کاوش عمومی وب از لحاظ نحوه نمایه‌سازی اطلاعات (با تأکید بر حوزه کتابداری). پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی. دانشکده روان‌شناسی و علوم تربیتی. دانشگاه تهران.

مهراد، جعفر، و مژگان بینش. ۱۳۸۴. ارزیابی سایت‌های وب بازبایی‌شده توسط دو راهنمای شبکه وب (ODP و YAHOO) در حوزه علوم کتابداری و اطلاع‌رسانی براساس معیارهای محتوا و کارایی. کتابداری و اطلاع‌رسانی ۱ (۱۰): ۴۰-۶۰. www.SID.ir (دسترسی در ۱۳۸۸/۱۰/۱۶).

میرزائی، خلیل. ۱۳۸۸. پژوهش، پژوهشگری و پژوهش‌نامه‌نویسی. تهران: جامعه‌شناسان.

- Ackermann, E. 2008. Directories and virtual libraries. searching and researching on the Internet and World Wide Web. Department of Computer Science, University of Mary Washington. <http://webliminal.com/search-web04.html> (accessed 24 Nov. 2009).
- Assadi, H., and T. Bauvisage. 2002. A comparative study of six french-language web directories. Franc Telecom R&D-DIH/UCE. <http://Citeseerx.ist.psu.edu> (accessed 15 June 2010).
- Bierrakos, D., and G. Paliouras. 2005. Exploiting probabilistic latent information for the construction of community web directories. Institute of Informatics and Telecommunications, NCSR "Demokritos". <http://www.springerlink.com> (Accessed 10 Dec. 2009).
- Cacheda, F., and R. Baeza-Yates. 2004. An optimistic Model for Searching Web directories. LNCS 2997,364-377. <http://www.springerlink.com> (Accessed 10 Feb 2010).
- Chen, S. 2005. A flexible interface design for web directories to accommodate different cognitive styles. *American Society for Information Science and Technology* (1) 56: 70-83. <http://citeseerx.ist.psu.edu> (Accessed 20 Dec. 2009).
- Kaptien, R., and J. Kamps. 2009. Web directories as topical context. *Archives and Information Studies, University of Amsterdam*. <http://remote.science.uva.nl/~kamps/publications/2009/kaptweb09.pdf> (Accessed 2 Dec. 2009).
- Krikos, V., S. Stamou, and P. Kokosis. 2005. Directory Rank: Ordering Pages in Web Directories. Germany: ACM. <http://portal.acm.org/citation> (Accessed 1 Nov. 2009).
- Massmann, S., and E. Rahm. 2008. Evaluating instance-based matching of web directories. Canada: International Workshop on web and database. <http://citeseerx.ist.psu.edu> (Accessed 19 Dec. 2009).
- Mundt, K. E. 2008. Internet search: subject directories, search engines, & more. *Library & Learning Commons, Dakota State University, USA*. <http://www.departments.dsu.edu/library/engines.html> (Accessed 13 Jan. 2010).
- Nicholson, S. 2000. A proposal for categorization and nomenclature for web search tools. In *Internet searching and Indexing: the subject approach*. England: The Howorth Press.
- Perugini, S. 2007. Symbolic links in the open directory project. *Information Processing and Management*. www.sciencedirect.com (Accessed 20 Nov. 2009).

Comparing the Scale of Web Subject Directories Precision in Technical-Engineering Information Retrieval

Mehrdokht Wazirpour Keshmiri
Ph.D in library and Information Science

Bahareh Fozooni*
Postgraduate in Library and Information Science

Iranian Journal of
**Information
Processing &
Management**

Abstract: The main purpose of this research was to compare the scale of web subject directories precision in information retrieval of technical-engineering science. Information gathering was documentary and webometric. Keywords of technical-engineering science were chosen at twenty different subjects from IEEE (Institute of Electrical and Electronics Engineers) and engineering magazines that situated in sciencedirect site. These keywords are used at five subject directories Yahoo, Google, Infomine, Intute, Dmoz, that were web directories high-utilization. Usually first results in searching tools are connected to searching keywords. Because, first ten results was evaluated in every search. These assessments to consist of scale of precision, scale of error, scale retrieval items in technical-engineering categories to retrieval items entirely. The used criteria for determining the scale of precision that was according to high-utilization standards in different documents, to consist of presence of the keywords in title, appearance of keywords at the part of web retrieved pages, keywords adjacency, URL of page, page description and subject categories. Information analysis was according to Kruskal-Wallis Test and L.S.D fisher. Results revealed that there was meaningful difference about precision of web subject directories in information retrieval of technical-engineering science, Therefore this theory was confirmed. web subject directories ranked from point of precision as follows. Google, Yahoo, Intute, Dmoz, and Infomine. The scale of observed error at the first results was another criterion that was used for comparing web subject directories. In this research, Yahoo had minimum scale of error and Infomine had most of error. This research also compared the scale of retrieval items in all of categories web subject directories entirely to retrieval items in technical-engineering categories, results revealed that there was meaningful difference between them. And second theory was confirmed. That Intute with 52 percent had most, and Google with 5.3 percent had least scale of ratio technical-engineering retrieval items to retrieval items in all of categories

Keywords Web subject directories, comparing, precision, information retrieval, technical-engineering

Iranian Research Institute
For Science and Technology
ISSN 2251-8223
eISSN 2251-8231
Indexed in LISA, SCOPUS & ISC
Vol.27 | No.3 | pp: 597-616
spring 2012

* Corresponding Author: b.fozooni@gmail.com