اطلاع‌شناسی
INFOLOGY

مقایسهٔ کارآیی موتورهای کاوش ...

احمد کمیجانی

چکیده

هدف اصلی این پژوهش مقایسهٔ کارآیی موتورهای کاوش عمومی و موتورهای کاوش تخصصی کشاورزی وب در بازیابی اطلاعات حوزهٔ کشاورزی است. بدین منظور، عناوین پنج طرح تحقیقاتی وزارت جهاد کشاورزی به صورت تصادفی انتخاب شده و در ۷ موتورکاوش عمومی و ۲ موتور کاوش تخصصی وب جست‌وجو شدند. ۲۰ نتیجهٔ نخست در هر بار جست‌وجو با استفاده از معیارهای ارتباط، اعتبار، روزآمد بودن، و تعداد پیوندهای غیرفعّال مورد بررسی قرار گرفتند. برای پاسخ به این سوّال که آیا میان موتورهای کاوش در بازیابی اطلاعات در هر یک از معیارها اختلاف معنی‌داری وجود دارد، از آزمون آماری بلوک تصادفی فریدمن استفاده گردید. یافته‌ها نشان می‌دهد که اختلاف معنی‌داری میان موتورهای کاوش در بازیابی مدارک مرتبط و معتبر وجود داشته ولی اختلاف معنی‌داری میان آنها در بازیابی اطلاعات روزآمد شده و نیز تعداد صفحات غیر فعال دیده نمی‌شود. در همهٔ معیارها موتورهای کاوش عمومی بهترین وموتورهای کاوش تخصصی ضعیف‌ترین نتایج را بازگرداندند. در این پژوهش، موتورهای کاوش گوگل و یاهو در همهٔ معیارها درمیان سه موتور کاوش نخست دیده می شوند. بهترین موتور کاوش در بازگرداندن نتایج مرتبط و معتبر موتورکاوش گوگل بود و موتور کاوش لیکس در دومعیار ارتباط واعتبار درجایگاه دوم قرار گرفت.

کلیدواژه‌ها: موتور کاوش عمومی. موتور کاوش تخصّصی. کشاورزی. بازیابی اطلاعات. ارزیابی. وب

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مقایسهٔ کارآیی موتورهای کاوش عمومی و تخصصی وب در بازیابی اطلاعات کشاورزی

احمد کمیجانی[1]

مقدمه

شبکهٔ جهانی اینترنت، به‌ویژه وب، در مراحل مختلف هر پژوهشی می تواند منبع اطلاعاتی محسوب شود. نتایج تحقیقات نشان می‌دهد که اگرچه محقّقان و اعضای هیأت علمی دانشگاه‌ها به‌تدریج تمایل بیشتری به استفاده از اینترنت به‌عنوان نوعی ابزار و منبع اطلاعاتی در پژوهش‌های خود نشان می‌دهند، در مقابل، پرسش‌هایی مبنی بر صحت و قابل اطمینان بودن اطلاعات برگرفته از منابع اینترنتی را نیز مطرح می‌سازند (هرینگ[2]، ۲۰۰۱).

بر روی شبکهٔ اینترنت منابع معتبر و قابل اطمینانی برای پژوهشگران و مشاوران و سایر کاربران کشاورزی وجود دارد؛ همچون سایت وب سازمان خواربارو کشاورزی سازمان ملل متحد[3] که خود با ایجاد پیوندهایی به مرکز اطلاعات کشاورزی جهانی[4] مرتبط است، و یا سایت‌های وب تحت نظارت دولت ایالات متحده نظیر سایت وب وزارت کشاورزی امریکا[5] و یا کتابخانهٔ ملی کشاورزی[6] که خود زیر نظر وزارت کشاورزی امریکاست، و یا سایت‌های تحت نظارت مؤسسات پژوهشی و یا دانشگاه‌ها

۱. کارشناس ارشد کتابداری و اطلاع‌رسانی

2. Herring

3. http://www.fao.org

4. World Agricultural Information Centre (WAICENT) ...http://www.fao.org/waicent/search/default.htm

5. http://www.usda.gov

6. National Agricultural Library (NAL).. http://www. nal.usda.gov

که از اعتبار بالایی برخوردارند. اما کاربرانی که از وجود چنین سایت‌هایی مطلع نیستند یا می‌خواهند چکیده‌ای از اطلاعات خاص موضوعی را سریعاً به‌دست آورند غالباً از یکی از موتورهای کاوش وب[1] استفاده می‌کنند.

اگرچه موتورهای کاوش وب به سبب عدم استفادهٔ گسترده از نظریه‌های رده‌بندی موضوعی و نیز غالباً به‌دلیل عدم جست‌وجوی کامل بر روی وب و یا ناکامی در بازگرداندن اطلاعات قابل اعتماد مورد تحلیل انتقادی قرار می‌گیرند، در عین حال به عنوان نوعی روش جست‌وجوی گسترده توسط ۸۰ درصد از کاربران اینترنت به‌کار می‌روند (کراسن[2] ۱۹۹۹).

بیان مسئله

مقایسهٔ میان نظام‌های بازیابی اطلاعات از دیر باز مورد توجه متخصصان اطلاع‌رسانی بوده است. استفاده‌کنندگان این نظام‌ها، آگاهانه یا غیرآگاهانه، نظام را برحسب اینکه تا چه حد نیازهای اطلاعاتی آنان را تأمین می‌کند ارزیابی می‌کنند. استفاده‌کنندگان معمولاً به چگونگی کارکرد نظام علاقه‌مند نیستند، بلکه فقط به نتایج نهایی توجه دارند. لذا بررسی بیشتر نیازهای استفاده‌کنندگان از نظام‌های بازیابی اطلاعات از اهمیت برخوردار است (لنکستر، ۱۳۷۹).

در پژوهش حاضر، با استفاده از مرور منابع مختلف، ۹ موتورکاوش وب شامل ۷ موتورکاوش عمومی که از پراستفاده‌ترین آنها بوده و ۲ مورد آنها نیز از موتورهای کاوش تخصصی کشاورزی هستند انتخاب گردید و مقایسه‌ای میان کارآیی این موتورها در بازیابی اطلاعات حوزهٔ کشاورزی صورت گرفت. واژه‌های کلیدی استفاده شده در موتورهای کاوش از عناوین پروژه‌های پژوهشی وزارت جهاد کشاورزی و به شکل نمونه‌گیری تصادفی انتخاب گردیده است.

پیشینه پژوهش

گانگ وو و جی لی[3] (۱۹۹۹) از دانشگاه ایالتی وین، در مقایسه‌ای میان ۷ موتور کاوش آلتاویستا، هات‌بات، اکسایت، اینفوسیک، نورثرن لایت[4]، یاهو، و مدیکال ورلد سرچ برای بررسی کارآیی آنها در پاسخگویی به پرسش‌های کاربران علوم بهداشتی

1. Web search engine
2. Krasne
3. Gang wu, Jie li
4. Northern light (http://www.(northernlight.com)

دریافتند که سه موتور کاوش اینفوسیک ، آلتاویستا، و اکسایت به‌عنوان موتورهای کاوش ردهٔ بالا، با بیشترین درصد نتایج بازیابی شده، دارای عملکرد خوبی از نظر روزآمد بودن منابع اطلاعات و ویژگی‌های پیشرفتهٔ جست‌وجو هستند. یاهو و مجموعهٔ منحصر به‌فرد "نورثرن لایت" برای جست‌وجوی منابع دارویی غیرمصنوعی و محصولات طبیعی موتورهای خوبی هستند. موتور کاوش هات بات، پوشش جامع و ویژگی‌های جست‌وجوی پیشرفته داشته و منبع خوبی برای جست‌وجوی تصاویر و فایل‌های چند رسانه‌ای است ولی مشکل پیوندهای تکراری در این موتور کاوش به چشم می‌خورد.

مارتینز و سانچز[1] در سال ۱۹۹۹، در مقاله‌ای با عنوان "مقایسه‌ای میان ابزارهای جست‌وجوی اینترنت" که در بیست و سومین نشست بین‌المللی اطلاعات پیوسته ارائه گردید ۱۰ موتور کاوش را با به‌کارگیری واژه‌های گردآوری اطلاعات، تحلیل استنادی، و بازیابی اطلاعات مورد ارزیابی قراردادند. آنان دریافتند که رابطه‌ای میان مورد استقبال واقع شدن یک موتور کاوش از سوی کاربران و توانمندی‌ها و ابزارهای جست‌وجوی اطلاعات موجود در آنان وجود ندارد. به‌طور مثال، موتورکاوش نورثرن لایت، اگرچه کمترین امتیاز را از لحاظ استقبال مردمی دارا بود ولی میزان اطلاعات مرتبط بازگردانده شده با به‌کارگیری واژهٔ بازیابی اطلاعات نسبت به موتور یاهو، که یکی از پرطرفدارترین موتورهای کاوش است، به مراتب بالاتر بوده و رتبهٔ نخست را کسب کرده و یاهو از موفقیت چندان مناسبی برخوردار نبوده است.

لیزا ویشارد[2] (۱۹۹۸) در مقاله‌ای تحت عنوان "مانعیت[3] درمیان موتورهای کاوش وب" دست به مطالعهٔ موردی در علوم زمین شناسی زده و با به‌کارگیری سه پرسش، مقایسه‌ای میان مانعیت ۳.۷ موتور کاوش وب صورت داده است. موتورهای کاوش در سه دسته راهنماهای وب، موتورهای کاوشی که به‌طور مستقل و با استفاده از روبات‌های خود اطلاعات را نمایه می کنند و ابرموتورهای کاوش[4] قرار گرفتند. ۱۵ نتیجهٔ نخست در هر موتور کاوش بررسی شده و مرتبط بودن آنان با پرسش مطرح شده مشخص گردید و نتایج ذیل به‌دست آمد:

در دستهٔ موتورهای کاوش راهنمای GO2, Infomine و Argus و در دستهٔ موتورهای کاوش به اصطلاح کلید واژه‌ای یا روباتی، موتورهای اکسایت، اینفوسیک، و

1. Martinez A. M., Sanchez E. F.
2. Wishard, Lisa
3. Precision
4. Meta search engines

نورثرن لایت به‌عنوان برترین‌ها با مانعیت بالا معرفی شدند و هیچ ابرموتور کاوشی نسبت به دیگری به عنوان برتر انتخاب نگردید.

در سال ۱۹۹۷، لیتون و سری‌واستاوا[1] از گروه کامپیوتر دانشگاه ایالتی ویرجینیا و دانشگاه مینه‌سوتا با ارائهٔ مقاله‌ای به بررسی میزان مانعیت در بازیابی اطلاعات در پنج موتور کاوش آلتاویستا، هات بات، لیکس[2]، اکسایت، واینفوسیک پرداختند. آنان جست‌وجوهای خود را در فاصلهٔ زمانی ۴۴ روزه در اوایل سال ۱۹۹۷ و با استفاده از ۱۵ موضوع متفاوت انجام دادند و پس از بررسی ۲۰ نتیجهٔ نخست در هر موتور کاوش و ارتباط آنها با موضوع‌های جست‌وجو شده و نهایتاً تجزیه و تحلیل آماری اعلام کردند موتورهای کاوش آلتاویستا، اکسایت، و اینفوسیک به ترتیب دارای بالاترین رتبه‌بندی در ارائهٔ نتایج مرتبط هستند.

لبدوف[3] (۱۹۹۷) در مقاله‌ای تحت عنوان "بهترین موتورهای کاوش در یافتن اطلاعات علمی در شبکهٔ اینترنت" در فاصلهٔ زمانی سوم اوت ۱۹۹۶ تا ۱۰ فوریه ۱۹۹۷ با استفاده از کلیدواژه‌های شیمی و فیزیک دست به مقایسهٔ کارآیی ۸ موتور کاوش وب زد. وی با مقایسهٔ تعداد نتایج بازیابی شده در موتورهای کاوش، سرانجام با قاطعیت موتور کاوش آلتاویستا را بهترین موتور در بازیابی اطلاعات در این دو زمینهٔ موضوعی معرفی کرده و ابرموتور *متاکرولر*[4] را در میان ابرموتورها انتخاب می‌کند. البته وی بررسی جالبی در نتایج بازیابی شده از موتورهای کاوش وب و حضور یا عدم حضور آنها در پایگاه اطلاعاتی تخصصی INSPEC کرده و به این نتیجه رسیده است که حدود ۱۰ تا ۲۰ درصد نتایج مرتبط در موتورهای وب در این بانک تخصصی یافت می‌شود.

چو و روزنتال[5] از دانشکدهٔ علوم کتابداری و اطلاع‌رسانی دانشگاه لانگ آیلند در نیویورک در سال ۱۹۹۶ با انتشار مقاله‌ای با عنوان "موتورهای کاوشی برای تار جهان‌گستر: مطالعه‌ای مقایسه‌ای و ارزیابانه" اقدام به مقایسه و ارزیابی توانمندی‌های سه موتور کاوش آلتاویستا، اکسایت، و لیکس در زمینه‌های ارائهٔ امکانات جست‌وجو همچون به‌کارگیری عملگرهای بولی، کوتاه‌سازی در جست‌وجو، نوع فیلدهای جست‌وجو، جست‌وجوی واژه و عبارت[6]، و نیز عملکرد آنان در بازیابی نظیر میزان مانعیت و زمان پاسخگویی به کاوش درخواستی، با استفاده از پرسش‌های واقعی مطرح شده در بخش مرجع کتابخانه، کردند. پژوهشگران دریافتند که موتور کاوش آلتاویستا

1. Leighton, Vernon, H. and Srivastava, J.
2. Lycos(http://www.lycos.com)
3. Lebedev, Alexander
4. Metacrawler (http://www.metacrawler.com)
5. Chu Hetting, Rosenthal Marilyn
6. Exact phrase

به‌نحو چشمگیری نسبت به دو موتور دیگر در ارائهٔ امکانات جست‌وجو و عملکرد بازیابی برتری داشت، در صورتی که میزان پوشش صفحات وب در موتور لیکس نسبت به دو موتور دیگر بالاتر بود. در این مقاله روشی برای ارزیابی موتورهای کاوش وب ارائه گردیده است.

اهداف پژوهش

مقایسهٔ کارآیی موتورهای کاوش عمومی و تخصصی کشاورزی مورد بررسی در بازیابی اطلاعات مرتبط و معتبر حوزهٔ کشاورزی و نیز مقایسهٔ این موتورهای کاوش در روزآمدسازی پیوندهای وب تحت پوشش خود و جلوگیری از بروز خطا در دسترسی به مدارک هدف اصلی این پژوهش است. از اهداف فرعی این پژوهش می‌توان به معرفی توانایی‌ها و ویژگی‌های هر یک از موتورهای کاوش مورد بررسی در پوشش صفحات وب، ارائهٔ امکانات جست‌وجو در سطح شبکهٔ اینترنت، رتبه‌بندی موتورهای کاوش بر اساس عملکرد آنها در جست‌وجوی مباحث خاص حوزهٔ کشاورزی، و ارائهٔ راهنمایی برای پژوهشگران حوزهٔ کشاورزی در استفادهٔ بهینه از موتورهای کاوش وب اشاره کرد.

پرسش‌های اساسی

۱. آیا اختلاف معناداری میان موتورهای کاوش منتخب در بازیابی مدارک مرتبط وجود دارد؟

۲. آیا در تعداد مدارک مرتبط معتبر بازیابی شده در هر موتورکاوش با دیگر موتورها اختلاف معنی‌داری وجود دارد؟

۳. آیا تفاوت معنی‌داری میان موتورهای کاوش در مورد تعداد پیوندهای کور[1] یا غیرفعّال بازیابی شده دیده می شود؟

۴. آیا اختلاف معنی‌داری میان موتورهای کاوش منتخب در این پژوهش در ارائهٔ مدارک روزآمد شده در یک سال اخیر وجود دارد؟

روش پژوهش

در این پژوهش از سه روش در مراحل مختلف طرح بهره گرفته شده است.

1. Dead links

روش پژوهش کتابخانه‌ای یا سندی[1] به‌منظور انتخاب موتورهای کاوش، تعیین معیارهای ارزیابی صفحات وب، و تعیین تعداد مدارک بازیابی شده‌ای که می‌بایست در هر موتور کاوش ارزیابی شوند؛ روش پیمایشی به‌منظور بررسی و اعمال معیارها بر صفحات وب برگزیده و دستیابی به داده‌های لازم جهت تجزیه و تحلیل داده‌ها؛ و سرانجام به منظور مقایسهٔ کارآیی و عملکرد موتورهای کاوش روش مقایسه‌ای به‌کار رفته است.

جامعهٔ مورد مطالعه

جوامع نمونهٔ متعددی در قسمت های مختلف پژوهش مورد توجه قرار گرفته است. در مورد موتورهای کاوش وب ، ۷ موتورکاوش عمومی وب شامل:

Altavista(http://www.av.com)

Excite(http://www.excite.com)

Google(http://www.google.com)

Hotbot(http://www.hotbot.com)

Infoseek(http://infoseek.go.com)

Lycos(http://www.lycos.com)

Yahoo(http://www.yahoo.com)

و ۲ موتورکاوش تخصصی کشاورزی وب شامل:

Agrisurf(http://www.agrisurf.com)

WebAgri(http://www.web-agri.com)

جامعهٔ پژوهش را تشکیل می‌دهند.

حوزهٔ مورد بررسی کشاورزی بوده و به‌منظور واقعی بودن واژه‌های کلیدی به‌کار رفته در موتورهای کاوش، جامعهٔ مورد مطالعه در این حوزه، طرح‌های پژوهشی مصوب وزارت جهاد کشاورزی بوده که از آن میان پنج عنوان منتخب برای تبدیل به واژه‌های کلیدی عبارتند از:

۱. بررسی اثر پلیمر BT53 در کنترل فرسایش و حفاظت خاک

(Effects of polymer BT53 to soil erosion control and soil protection))

1. Library research method

۲. شناسایی مکان‌های زیست آرتمیا

(Identification of Artemia·SPP. habitats)

۳. مطالعات آت اکولوژی گیاهان مهم مرتعی

(Survay of rangeplants autecology)

۴. بررسی روش‌های شکست خواب بذور

(survay of seed dormancy breaking)

۵. بررسی واکنش تعدادی از ژنوتیپ‌های زیتون به کلرید سدیم

(Effects of sodium) chloride on some of olive genotypes)

و پس از انجام کاوش در موتورهای منتخب ،از میان جامعه رکوردهای اطلاعاتی منتج از کاوش، ۲۰ نتیجهٔ نخست در هر عمل کاوش جهت ارزیابی انتخاب گردیدند.

شیوهٔ نمونه‌گیری و اعمال معیارها

شیوهٔ نمونه‌گیری در مراحل مختلف پژوهش متفاوت بوده است. در انتخاب طرح‌های پژوهشی وزارت جهاد کشاورزی از هریک از مؤسسات مادر تحت پوشش معاونت آموزش، تحقیقات و ترویج کشاورزی این وزارتخانه به‌طور تصادفی یک طرح جهت تبدیل عنوان آن به واژه‌های کلیدی انتخاب گردید. در مورد موتورهای کاوش عمومی وب، پس از بررسی بیش از ۱۵ مقاله برای مقایسهٔ این موتورها که در سال‌های اخیر منتشر شده‌اند، ۷ موتور کاوش که در همهٔ این مقالات به‌عنوان پراستفاده‌ترین موتورهای کاوش[1] شناخته شده‌اند و بیشترین میزان حضور را داشتند انتخاب گردیدند. در مورد موتورهای تخصصی کشاورزی، با بررسی موتورهای کاوش موجود در این زمینه، موتورهای کاوشی برگزیده شدند که امکان استفادهٔ کامل از عملگرهای بولی و کوتاه‌سازی را به کاربر می‌دهند و جالب اینکه این دو موتور کاوش، یکی نخستین موتور کاوش کشاورزی و دیگری مدّعی‌ترین آنها در نمایهٔ تعداد بالای صفحات وب کشاورزی بودند.

برای دستیابی به معیارهای لازم برای ارزیابی صفحات وب ، بیش از ۱۶ لیست ارائه شده از معیارهای ارزیابی توسط کتابخانه های دانشگاهی نظیر:

ـ کتابخانه‌های دانشگاه ایالتی کلورادو[2]

1. Popular search engine
2. Colorado State University Libraries

ـ کتابخانه یادبود ولف گنگ[1]

ـ بخش خدمات مرجع کتابخانه دانشگاه کرنل[2]

ـ بخش خدمات علوم اجتماعی کتابخانه دانشگاه فلوریدا[3]

و معیارهای ارائه شده در پروژهٔ اینترنتی اسکات[4]، معیارهای ارائه شده در مقاله‌های بسو[5] (۱۹۹۹)، کاپون[6] (۱۹۹۸)، و هریس[7] (۱۹۹۷) و معیارهای ارائه شده توسط سوزان بک[8] (۱۹۹۷) در مقاله‌ای تحت عنوان "خوب، بد، زشت: چرا این اندیشه‌ای مناسب برای ارزیابی صفحات وب است؟" و نیز سایر مقالات مرتبط مورد بررسی قرار گرفت و معیارهای مشترک در همهٔ این منابع برگزیده شد. انتخاب ۲۰ نتیجهٔ نخست هر کاوش جهت ارزیابی با استفاده از معیارها نیز براساس مرور منابع صورت گرفته است. نتایج حاصل از مطالعهٔ رفتار کاربران در حین استفاده از موتورهای کاوش بیانگر این است که، بیش از ۷۵ درصد از آنان، فقط ۲۰ نتیجهٔ نخست بازگردانده شده توسط موتورهای کاوش در هر بار مراجعه به آن را مورد بررسی قرار می‌دهند (ساراسویک[9]، ۲۰۰۰).

برای همه موتورهای کاوش تنظیمات پیش فرض[10] یعنی تنظیمات اولیه صورت گرفته توسط موتورهای کاوش نظیر زبان‌های مدارک بازیابی شده، نوع فایل‌های مورد جست‌وجو، یا تعداد مدارکی که پس از هر جست‌وجو در یک صفحهٔ نمایش داده می‌شوند به‌کار گرفته شد. برای دستیابی به بهترین نتایج کاوش، استراتژی جست‌وجو، یا به‌عبارت دیگر، نحوهٔ به‌کارگیری واژه‌های کلیدی به گونه‌ای تنظیم گردید که از حداکثر قابلیت‌های ارائه شده توسط هر موتور کاوش در ترکیب کلید واژه‌ها بهره گرفته شود. کلیدواژه‌ها با استفاده از عملگرهای بولی و عملگرهای غیربولی (همچون + و یا "") و نیز نشانگرهای کوتاه‌سازی (مثل *) با یکدیگر ترکیب شدند. برای ارزیابی نتایج حاصل از کاوش، از معیارهای مرتبط بودن[11] مدرک، اعتبار[12] منبع و مطلب، روزآمد بودن[13] مدرک یا همان آخرین تاریخ بازبینی صفحه وب، و تعداد پیوندهای غیرفعال[14]، یا به اصطلاح پیوندهای کور، در نتایج حاصل از جست‌وجو استفاده گردید.

مدارک مرتبط مدارکی هستند که بسیاری از جنبه‌های موضوع مورد نظر را در بر می‌گیرند و توسط متخصصان و کارشناسان مربوط به‌عنوان منبع مفید و قابل استفاده ارزیابی می‌شوند. اعتبار منبع براساس پدیدآور، منابع به‌کار گرفته شده در مدرک، و

1. Wolfgang Memorial Library
2. Reference Division, Cornell University
3. University of Florida Library, Humanities and Social Sciences Services
4. Internet Scout Project Report selection criteria
5. Basu
6. Kapoun
7. Marris
8. Beck
9. Saracevic
10. Default setting
11. Relevancy
12. Authority
13. Currency
14. Inactive links

اعتبار ارگان‌ها یا مراکز ارائه‌دهنده یا حمایت‌کننده در ارائهٔ این مطالب تعیین می‌گردد.

دانشگاه‌ها، مؤسسات آموزشی و پژوهشی، نمایندگی‌های دولتی و انجمن‌های علمی غالباً مطالب مرتبط و معتبر در زمینه‌های تخصصی خود ارائه می‌دهند. برای نمونه در بررسی اعتبار مطالب مندرج در صفحات وب به آدرس منحصر به‌فرد صفحه وب یا ارل[1] می‌توان توجه کرد. مؤسسات دانشگاهی و پژوهشی معمولاً دارای نام حوزهٔ اصلی edu در آدرس خود هستند.

مدارک روزآمد، مدارکی تلقی می‌شود که در یک سال اخیر منتشر گردیده یا صفحات آنها به روز شده باشند و به‌طور معمول تاریخ روزآمد شدن صفحات در انتهای آنها می‌آید.

پیوندهای وبی که، پس از درخواست، شامل خطای ۴۰۴ یا پیامی مبنی بر یافت نشدن سرویس‌دهنده یا خطا در خدمات DNS سایت موردنظر باشند پیوندهای غیرفعّال یا کور تلقی شده‌اند. خطای ۴۰۴ هنگامی به وقوع می پیوندد که:

۱. مدرک موردنظر به نقطه‌ای دیگر منتقل گردیده، تغییرنام داده، یا حذف شده است؛

۲. آدرس محل قرارگیری مدرک یا صفحهٔ موردنظر بر روی شبکهٔ اینترنت به‌درستی در میلهٔ آدرس مرورگر اینترنت وارد نشده باشد.

نسبت پیوندهای کور در موتورهای کاوش مختلف بیانگر مقیاسی برای درک چگونگی روزآمدسازی اطلاعات در پایگاه‌های اطلاعاتی هر یک از موتورهای کاوش و میزان کنترل آنان بر این پیوندها در فواصل زمانی مطلوب است.

کلیدواژه‌ها در محدودهٔ زمانی ۲۰ روزه، از روز بیست و پنجم خرداد ماه تا روز سیزدهم تیر ماه ۱۳۸۱، جست‌وجو شد. ۲۰ نتیجهٔ نخست در هر کاوش ذخیره گردید و در اختیار کارشناسان مربوط قرار گرفت تا از میزان و تعداد مدارک مرتبط آگاهی حاصل گردد.

شیوهٔ تجزیه وتحلیل یافته‌ها

در این پژوهش از طرح آماری بلوک تصادفی فریدمن[2] برای تجزیه و تحلیل آماری نتایج و تعیین معنی‌دار بودن اختلاف موجود در میان موتورهای کاوش با توجه به معیارهای زیر استفاده شده است:

1. URL (Universal Resource Locator)

2. Friedmanns Randomized Block Design

معیار مرتبط بودن

جدول ۱. تعداد نتایج مرتبط در موتورهای کاوش

	آلتاویستا	اکسایت	گوگل	هات بات	اینفوسیک	لیکس	یاهو	اگری سرف	وب اگری
سؤال ۱	۱۲	۲	۱۴	۱۱	۱۶	۸	۱۴	۴	۷
سؤال ۲	۱۳	۱۲	۱۱	۶	۹	۱۱	۹	۰	۴
سؤال ۳	۶	۴	۴	۲	۱	۶	۵	۱	۰
سؤال ۴	۶	۹	۱۳	۶	۱	۱۱	۱۰	۰	۰
سؤال ۵	۲	۵	۴	۲	۲	۳	۳	۰	۰

نتایج جدول ۱ پس از رتبه‌بندی در هر بلوک (که در اینجا هر سؤال در حکم بلوک است) به صورت جدول ۲ قابل ارئه است:

جدول ۲. رتبه‌بندی نتایج مرتبط در موتورهای کاوش

	آلتاویستا	اکسایت	گوگل	هات بات	اینفوسیک	لیکس	یاهو	اگری سرف	وب اگری
سؤال ۱	۶	۱	۷/۵	۵	۹	۴	۷/۵	۲	۳
سؤال ۲	۹	۸	۶/۵	۳	۴۰۵	۶/۵	۴/۵	۱	۲
سؤال ۳	۸/۵	۵/۵	۵/۵	۴	۲/۵	۸/۵	۷	۲/۵	۱
سؤال ۴	۴/۵	۶	۹	۴/۵	۳	۸	۷	۱/۵	۱/۵
سؤال ۵	۴	۹	۸	۴	۴	۶/۵	۶/۵	۱/۵	۱/۵
جمع	۳۲	۲۹/۵	۳۶/۵	۲۰/۵	۲۳	۳۳/۵	۳۲/۵	۸/۵	۹

در اینجا برای رسیدن به پاسخ این سؤال که میان این موتورهای کاوش اختلاف معنی‌داری در بازیابی مدارک مرتبط وجود دارد یا خیر، بایستی مقدار ثابت x_r^2 برای آزمون فریدمن محاسبه گردد.

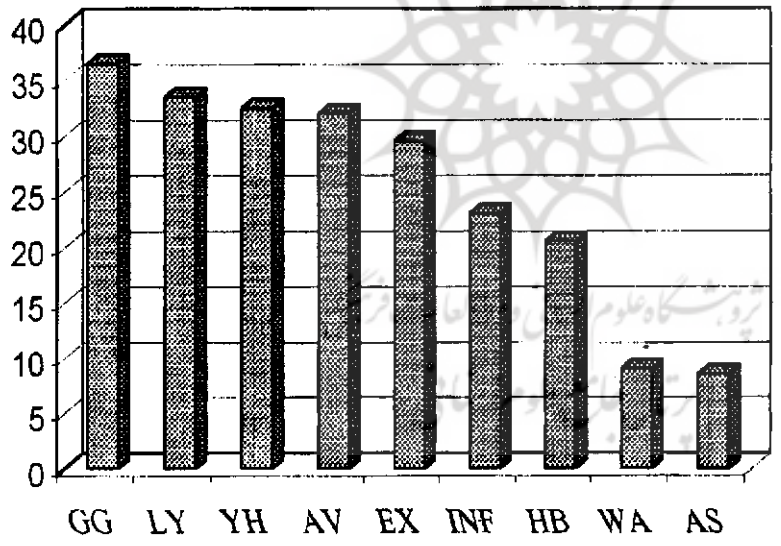
$$x_r^2 = \frac{12}{Nk(k+l)} \sum_{j=l}^{k} (R_j)^2 - 3N(k+l)$$

که در آن، N تعداد ردیف‌ها، K تعداد ستون‌ها، R_j برابر حاصل جمع رتبه‌ها در ستون j ام، و $\sum_{}^{k}$ جمع مربع حاصل جمع رتبه‌ها برای همهٔ K موقعیت درنظر گرفته شده است؛ که نتیجهٔ محاسبه به شرح زیر است:

$$x_r^2 = \frac{12}{5\times9(9+1)} \left[(32)^2 + (29.5)^2 + (20.5)^2 + (23)^2 + (33.5)^2 + (32.5)^2 + (8.5)^2 + (9)^2\right]$$

$$-3\times5(9+1) = 23.53$$

اکنون با مراجعه به جدول x^2 با درجه آزادی df=k-1=9-1=8 درمی‌یابیم که در سطح، α=0.01، میزان x_r^2 به‌دست آمده بزرگ‌تر از میزان موجود در جدول است و بدین معنی است که اختلاف بسیار معنی‌داری میان موتورهای کاوش با میزان ۹۹ درصد اطمینان وجود دارد.



نمودار ۱. مجموع رتبه‌های مرتب شده براساس معیار ارتباط در موتورها

اکسایت = EX آلتاویستا = AV یاهو = YH لیکس = LY گوگل = GG

اگری سرف = AS پوب اگری = WA هات بات = HB اینفوسیک = INF

معیار اعتبار

جدول ۳. تعداد نتایج معتبر در موتورهای کاوش

	آلتاویستا	اکسایت	گوگل	هات بات	اینفوسیک	لیکس	یاهو	اگری سرف	وب اگری
سؤال ۱	۶	۲	۱۲	۱۰	۱۳	۸	۱۲	۴	۷
سؤال ۲	۷	۱۱	۱۱	۶	۹	۱۰	۹	۰	۴
سؤال ۳	۶	۴	۴	۲	۱	۶	۵	۱	۰
سؤال ۴	۶	۹	۱۰	۶	۱	۹	۹	۰	۰
سؤال ۵	۲	۴	۴	۲	۲	۳	۳	۰	۰

حال نتایج جدول بالا را در هر گروه یعنی سؤالات ۱ تا ۵ رتبه‌بندی کرده تا آزمون آماری صورت گیرد.

جدول ۴. رتبه‌بندی نتایج معتبر در موتورهای کاوش

	آلتاویستا	اکسایت	گوگل	هات بات	اینفوسیک	لیکس	یاهو	اگری سرف	وب اگری
سؤال ۱	۳	۱	۷/۵	۶	۸	۵	۷/۵	۲	۴
سؤال ۲	۴	۸/۵	۸/۵	۳	۵/۵	۷	۵/۵	۱	۲
سؤال ۳	۸/۵	۵/۵	۵/۵	۴	۲/۵	۸/۵	۷	۲/۵	۱
سؤال ۴	۴/۵	۷	۹	۴/۵	۳	۷	۷	۱/۵	۱/۵
سؤال ۵	۴	۸/۵	۸/۵	۴	۴	۶/۵	۶/۵	۱/۵	۱/۵
جمع	۲۴	۳۰/۵	۳۹	۲۱/۵	۲۳	۳۴	۳۳/۵	۸/۵	۱۰

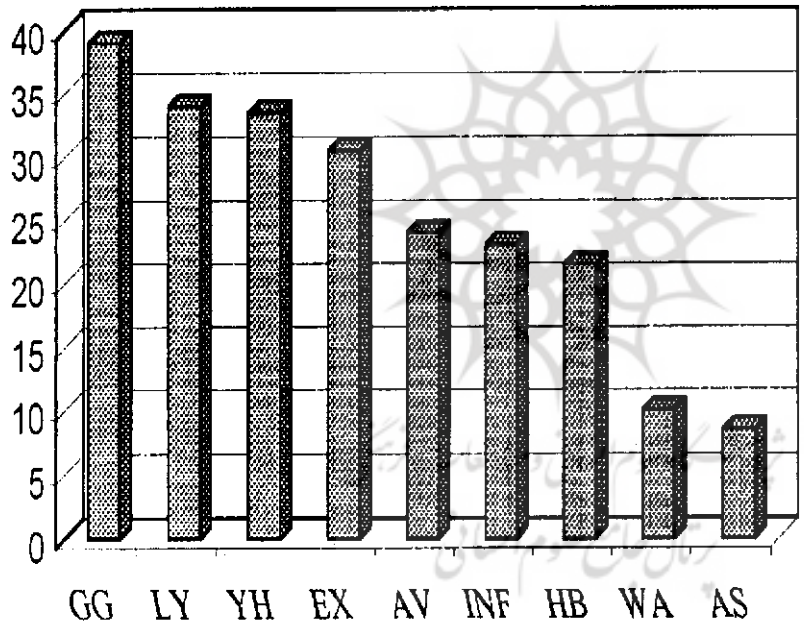
برای پاسخ به این پرسش که آیا موتورهای کاوش مورد بررسی در بازیابی مدارک معتبر اختلاف معنی‌دار آماری دارند یا خیر بایستی مقدار χ^2_r برای این صفت اندازه‌گیری شود.

$$x_r^2 = \frac{12}{5 \times 9(9+1)} \left[(24)^2 + (30.5)^2 + (39)^2 + (21.5)^2 + (23)^2 + (34)^2 + (33.5)^2 + (8.5)^2 + (10)^2 \right]$$

$$-3 \times 5(9+1) = 22.50$$

با مراجعه به جدول x^2 درمی‌یابیم که میزان x_r^2 به دست آمده در این صفت از میزان موجود در جدول در سطح p=۰/۰۱ بزرگ‌تر است و این بدان معنی است که میان موتورهای کاوش با سطح اطمینان ۹۹ درصد اختلاف معنی‌داری در بازیابی مدارک معتبر وجود دارد.

در نمودار ۲ موتورهای کاوش براساس مجموع رتبه‌ها در پنج سؤال و در صفت اعتبار مرتب شده‌اند که خود بیانگر قوی‌ترین و ضعیف‌ترین آنها در این صفت است.



نمودار ۲. مجموع رتبه‌های مرتب شده براساس معیار اعتبار در موتورها

معیار روزآمد بودن

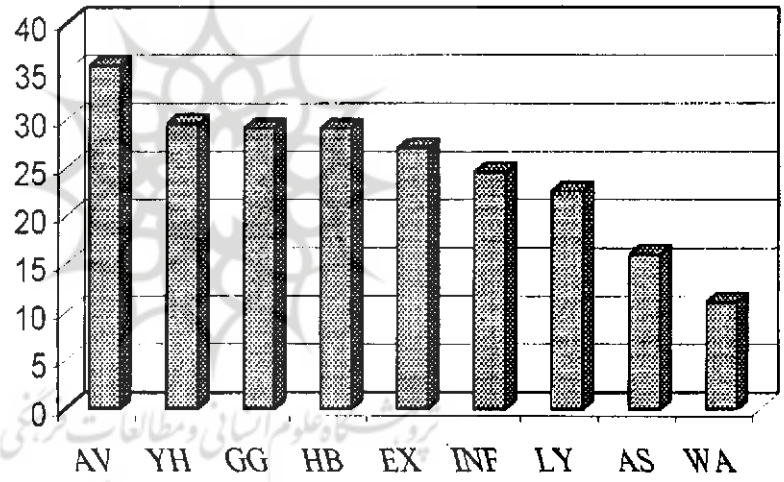
برای پاسخ به این سؤال که آیا در بازیابی صفحات روزآمد شده اختلاف معنی‌داری در میان موتورهای کاوش مورد بررسی وجود دارد یا خیر، میزان x_r^2 برای مجموع

رتبه‌های به‌دست آمده ازموتورهای کاوش محاسبه گردید:

$$x_r^2 = \frac{12}{5 \times 9(9+1)} \left[(35.5)^2 + (27)^2 + (29)^2 + (29)^2 + (24.5)^2 + (22.5)^2 + (29.5)^2 + (16)^2 + (11)^2 \right]$$
$$- 3 \times 5 = (9+1) = 10.66$$

با مقایسهٔ مقدار x_r^2 به‌دست آمده با جدول x^2 در می‌یابیم که اختلاف معنی‌داری میان موتورهای کاوش در سطوح $\alpha=0.01$ و $\alpha=0.05$ در بازیابی صفحات روزآمد شده در یک‌سال اخیر وجود ندارد.

در نمودار ۳ مجموع مرتب شدهٔ رتبه‌ها برای موتورهای کاوش آمده است.



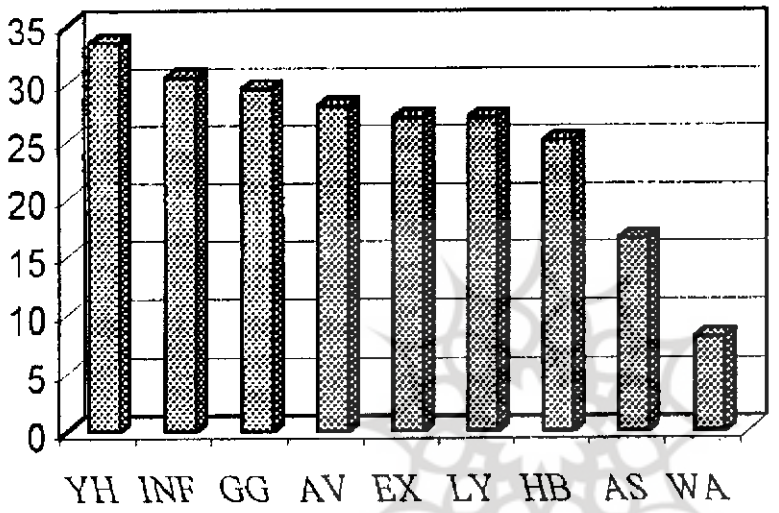
نمودار ۳. مجموع رتبه‌های مرتب شده براساس معیار روزآمدی

معیار پیوندهای غیرفعّال

همچون سایر معیارها مقدار x_r^2 برای معیار پیوندهای غیر فعّال نیز محاسبه گردید.

$$x_r^2 = \frac{12}{5 \times 9(9-1)} \left[(28)^2 + (27)^2 + (29.5)^2 + (25)^2 + (30.5)^2 + (27)^2 + (33.5)^2 + (16.5)^2 + (8)^2 \right]$$
$$- 3 \times 5\,(9+1) = 13.36$$

با مقایسه مقدار x_p^2 به‌دست آمده با جدول x^2 با درجه آزادی df =k-1=9-1=8 در می‌یابیم که احتمال معنی‌دار بودن این اختلافات برابر p =٪۱۰ است که این از لحاظ آماری میزان خطای بالایی تلقی شده و باید چنین نتیجه‌گیری کرد که در سطح معنی‌دار $\alpha=0.05$ اختلاف معنی‌داری میان موتورهای کاوش مورد مقایسه در این پژوهش، در تعداد پیوندهای غیرفعال وجود ندارد.



نمودار ۴. مجموع رتبه‌های مرتب شده براساس تعداد پیوندهای فعال

نتیجه‌گیری

با نگاهی به نتایج به‌دست آمده از آزمون فریدمن برای تجزیهٔ واریانس دوطرفه از طریق رتبه بندی مشاهده گردید که اختلاف معنی‌داری میان موتورهای کاوش مورد بررسی در بازیابی اطلاعات مرتبط با سؤالات مطروحه در حوزهٔ کشاورزی دیده می‌شود و این بیانگر آن است که با مراجعه به موتور کاوشی که نتایج بهتری را نشان می‌دهد دستیابی به مدارک مرتبط بیشتر و با اختلاف قابل توجهی میسّر می‌گردد.

در معیار ارتباط، سه موتور کاوش به ترتیب اولویت گوگل ، لیکس و یاهو از جمله موتورهای کاوش عمومی بودند وآخرین رتبه ها به موتورهای کاوش تخصصی کشاورزی تعلق داشت. در خصوص معیار اعتبار مدارک نیز براساس آزمون فریدمن ،

اختلاف بسیار معنی داری میان موتورهای کاوش دیده می شود و این بیانگر اثبات فرضیهٔ مطرح شده در این پژوهش است.

در این معیار نیز سه موتورکاوش گوگل، لیکس، و یاهو از گروه موتورهای کاوش عمومی دارای رتبه‌های نخست بودند و باز هم موتورهای کاوش کشاورزی رتبه‌های آخر را دارا بودند. در مورد معیار روزآمد بودن صفحات بازیابی شده برخلاف فرضیه مطرح شده، آزمون آماری اختلاف معنی‌داری را میان موتورهای کاوش نشان نداد.

موتورهای کاوش مورد بررسی دربازیابی پیوندهای غیرفعّال درنتایج کاوش، اختلاف معنی‌داری را نشان نمی‌دهند و این بدان معنی است که این موتورها کنترل مطلوبی بر پایگاه‌های داده‌ای خود در حذف پیوندهای کور دارند. اگرچه آزمون آماری اختلاف معنی‌داری میان موتورهای کاوش در مدارک روزآمد و پیوندهای غیرفعّال نشان نمی‌دهد ، با نگاهی به رتبه‌بندی موتورهای کاوش درمی‌یابیم که در این معیارها نیز موتورهای کاوش عمومی از رتبهٔ بالاتری نسبت به موتورهای کاوش تخصصی برخوردارند.

سرانجام می‌توان گفت که موتورهای کاوش عمومی در هر دو معیار ارتباط و اعتبار که به‌نظر می‌رسد به ترتیب مهم‌ترین معیارهای مورد توجه از سوی کاربران متخصص کشاورزی باشند برخلاف انتظار با اختلاف چشمگیری بر موتورهای کاوش تخصصی کشاورزی برتری دارند، به نحوی که موتورهای کاوش کشاورزی همیشه در انتهای جدول رتبه‌بندی جای داشتند و این بیانگر آن است که آن موتورها هنوز به بلوغ و سطح قابل قبولی برای مراجعهٔ کاربران کشاورزی در امور پژوهشی دست نیافته‌اند. شاهد این ادعا را می‌توان در بازنگرداندن حتی یک رکورد در برخی پرسش‌های مطرح شده دانست، در حالیکه موتورهای کاوش عمومی تعداد قابل توجهی رکورد را در این زمینه‌ها بازیابی کرده بودند.

نتیجهٔ دیگر آنکه در مورد موتورهای کاوش عمومی، دو موتورکاوش گوگل و یاهو همیشه در میان سه موتور کاوش برتر در کلّیهٔ معیارهای مورد بررسی در این پژوهش دیده می‌شوند. براساس آمارهای ارائه شده توسط سولیوان[1] (۲۰۰۱) و نوتس[2] (۲۰۰۲) و تخمین تعداد واقعی صفحات نمایه شده در موتورهای کاوش و اشارهٔ آنان بر رشد روزافزون صفحات تحت پوشش موتور گوگل، در مقایسه با سایر موتورهای کاوش

1. Sullivan
2. Notess

وب در سال‌های آینده، می‌توان پیش‌بینی کرد که این موتور و نیز موتور کاوش یاهو که از پایگاه داده گوگل به‌طور مشترک استفاده می‌کند رتبه و جایگاه خود را در این رتبه‌بندی حفظ کرده و از برترین موتورهای کاوش وب در سال‌های آتی باشند.

مآخذ

لنکستر، فردریک ویلفرید (۱۳۷۹). *نظامهای بازیابی اطلاعات، ویژگیها، آزمون و ارزیابی*. ترجمه جعفر مهراد. شیراز: نوید.

Basu,k. (1999) "Evaluation of internet resources. CALIBER 99. Academic libraries in the internet era". *Proceedings of the 6th national convention for automation of libraries in education and research*. nagpur, India, p. 296-301.

Beck, Susan (1997). "Evaluation criteria: The good, The bad & the ugly: or, Why It's a good idea to evaluate Web sources". [online] Available: http://lib.nmsu.edu/instruction/evalcrit.html [Jan. 2002]

Chu, Hetting; Rosenthal, Marilyn (1996). "Search engines for the world wide web: A comparitive study and evaluation methodology. *ASIS Annual Conference Proceedings*

Daniel, Wayne W. (1990). *Applied nonparametric statistics*. Boston PWS-Kent publishing company.

Gang wu, Jie li. (1999). "Comparing Web search engine performance in searching. consumer health information: Evaluation and recommendations". *Bulletin of Medical Library Association*. 87(4): 456-461

Harris, Robert (1997). "Evaluation internet research sources". [online] Available: http:// http://www.virtualsalt.com/evalu8it.htm

Herring-S-D. (2001). "Using the World Wide Web for research: are faculty satisfied?". *Journal of Academic Librarianship*, 27(3): 213-19.

"Internet Scout projectscout report selection criteria" [online] Available: http://scout.cs.wisc.edu/report/ sci-eng/criteria.html

1. Levy
2. Ross

Kapoun, Jim (1998). "Teaching undergrads Web evaluation: A guide for library instruction." *C & RL News* (July/August): 522-523.

Krasne, Alexandra (1999). "What are your search habit?". Pcworld.com. Retrieved from the World Wide Web on 20 July 2002 from http://pcworld.com/news/article/0,aid, 12826, 00,asp

Lebedev, Alexander (1997). Best search engines for finding scientific information in the web. [Online]http://www.chem.msu.su/eng/comparison.html

Leighton, Vernon, H. and Srivastava, J.(1997). Precision among World Wide Web search services (Search engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos. [Online] http:// www.winona.msus.edu/library/webind2/webind2.htm

Martinez A.M.,Sanchez E.F.(1999). "Comparing internet search tools". Online information 99. proceedings of the 23rd. international online information meeting.P 263-266

Notess, Greg R. (2002). Search engines reviews. Searchengineshowdown.com

Saracevic,T.,Jansen,B.J.,Spink, A.(2000). Real life,real users,and real needs: A study and analysis of users queries on the web. *Information processing and management*. 36(2), 207-227.

Sullivan, Danny (2001). "Search engine sizes". searchenginewactch.com. Retrieved from the World Wide Web on 25 Dec 2002 from http://searchenginewatch.com/reports/ sizes.html

Wishard, Lisa (1998). "precision among internet search engines: an earth sciences case study". *Issues in science and technology librarianship*. spring (18)