

مقایسه تأثیر عنوان و نشانی اینترنتی در میزان ربط نتایج حاصل از بازیابی اطلاعات در موتورهای جستجو در دو حوزه علوم کشاورزی و علوم انسانی

پریسا اعلامی* | کارشناس ارشد کتابداری و اطلاع‌رسانی

رحمت‌الله فتاحی^۱ | استاد،
گروه کتابداری و اطلاع‌رسانی، دانشگاه فردوسی مشهد

دریافت: ۱۳۸۹/۰۳/۰۲ | پذیرش: ۱۳۹۰/۱۱/۱۸

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳-۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS، LISA و ISC
http://jipm.irandoc.ac.ir
دوره ۲۸ | شماره ۱ | صص ۲۰۳-۲۲۴
پاییز ۱۳۹۱

نوع مقاله: پژوهشی

*parisaallami@gmail.com
1. fattahirahmat@gmail.com

چکیده: هدف پژوهش حاضر مقایسه میزان ربط نتایج حاصل از محدودسازی عبارت جستجو به عنوان و نشانی اینترنتی در دو حوزه علوم انسانی و علوم کشاورزی از دید کاربر بوده است. همچنین، مقایسه میزان حضور کلیدواژه‌های مورد جستجو در عنوان و نشانی اینترنتی وبسایت‌های دو حوزه و نیز رابطه حاکم بر کلیدواژه‌های مورد جستجو و همخوانی عنوان و نشانی اینترنتی وبسایت‌های حاصل از جستجو در موتور کاوش نیز از اهداف این پژوهش بوده است. نتایج پژوهش میزان ربط بیشتر نتایج حاصل از نشانی اینترنتی را نسبت به عنوان در حوزه علوم کشاورزی نشان داد. اما، در حوزه علوم انسانی تفاوت معنی‌داری بین میزان ربط نتایج جستجوی مبتنی بر عنوان و نشانی اینترنتی وجود نداشت. همچنین، نتایج حاکی از آن بود که هر چه تعداد کلیدواژه مورد جستجو کمتر باشد، میزان همخوانی عنوان و نشانی اینترنتی بیشتر است؛ هرچه میزان همخوانی عنوان و نشانی اینترنتی بیشتر باشد، میزان ربط نتایج نیز بالاتر است. نتیجه نهایی پژوهش این بود که کاربران با کنترل تعداد کلیدواژه‌های مورد جستجو نیز می‌توانند به موارد مرتبط‌تر دست یابند. همچنین، موتورهای کاوش تخصصی می‌توانند با افزایش وزن‌دهی به کلیدواژه‌های موجود در نشانی، نتایج مرتبط‌تری را رتبه‌بندی کنند و در اختیار کاربران خود قرار دهند.

کلیدواژه‌ها: عنوان، نشانی اینترنتی، ربط، بازیابی اطلاعات، موتورهای جستجو

۱. مقدمه

عنوان هر اثر به طور معمول، نخستین بخش از آن است که نظر خواننده را به خود جلب می‌کند. عنوان هر اثر جلوه‌ای از هویت آن اثر است. عنوان کتاب‌ها، مجله‌ها، پایان‌نامه‌ها، مقالات و... به طور معمول در خلاصه‌ترین شکل ممکن و با حروفی بزرگتر از متن نمایش داده می‌شود. در صفحات وب نیز عنوان برای جلب مخاطب و نمایه شدن در موتورهای کاوش از اهمیت زیادی برخوردار است. عنوان صفحه یکی از مهم‌ترین بخش‌ها از دید جستجوگرهاست. به طور معمول، قضاوت افراد برای انتخاب یک صفحه وب براساس عنوان آن است. به طور معمول، موتورهای کاوش، بیش از کلیدواژه‌های موجود در متن، به عنوان صفحات وب وزن می‌دهند. از این رو، یکی از معیارهای مهم و اساسی در رتبه‌بندی صفحه‌های وب، عنوان صفحه وب است. هنگام جستجو در موتورهای کاوش، صفحه‌های وب که واژه مورد جستجو در عنوان آنها به کار رفته باشد، به طور معمول به منزله مرتبط‌ترین مدارک نسبت به موضوع جستجو در نظر گرفته می‌شوند. به همین دلیل، در بیشتر موتورهای کاوش، گزینه‌ای برای محدود کردن دامنه جستجو به عنوان صفحات در نظر گرفته می‌شود. در صفحه منبع صفحه وب، عنوان صفحه وب دارای یک تگ^۱ (برچسب نمایه‌پذیر) در زبان فرامتن^۲ است که به طور معمول در بین برچسب‌های `<TITLE></TITLE>` قرار می‌گیرد.

از سوی دیگر، نشانی‌های اینترنتی^۳ نیز به منظور مکان‌یابی و دسترسی کاربران به اطلاعات در نظر گرفته شده‌اند. در اینترنت نیز برای دسترسی به واحدهای اطلاعات چه این اطلاعات یک عکس باشد، چه یک نوشته، نیاز به یک نشانی است که ما را به آن واحد اطلاعاتی راهنمایی کند. به همین دلیل، برخی موتورهای کاوش به کلیدواژه‌های موجود در نشانی اینترنتی توجه نشان می‌دهند. در همین راستا، به سازمان‌ها و نهادها توصیه می‌شود که اسم‌های پوشه‌های موجود در سایت (یعنی اسامی بامسمی و بامفهوم فایل‌های تشکیل‌دهنده سایت) را به نشانی اینترنتی صفحات بیفزایند (Web position 2008).

در هنگام جستجو و بازیابی در موتورهای کاوش تعداد نشانی‌های اینترنتی یا پیوندهایی که به منزله نتایج جستجو ارائه می‌شوند، به طور معمول بسیار زیاد هستند. اما، همه این نتایج مرتبط نیستند و حتی ممکن است عواملی مثل ابهام زبان باعث شود نتایج مناسبی به کاربر ارائه نشود. برای فراهم کردن دسترسی سریع به صفحات مناسب و اینکه صفحات با ارتباط موضوعی بیشتر، در الویت بالاتری قرار بگیرند، الگوریتم‌های جستجو در موتورهای کاوش، راهبردهای رتبه‌بندی مختلفی را به کار می‌برند.

1. tag 2. HTML 3. URL (Uniform Recourse Location)

موتورهای کاوش امکانات گوناگونی را برای محدودسازی دامنه جستجو و در نتیجه بالا بردن احتمال ربط نتایج بازیابی در نظر گرفته‌اند. با توجه به اهمیت عنوان و نشانی اینترنتی که پیشتر به آن اشاره شد، یکی از این امکانات، محدود کردن کلیدواژه و عبارت جستجو به عنوان و یا نشانی اینترنتی است که تعداد نتایج جستجو را به طور معنی داری کاهش می‌دهد و در عین حال، احتمال ربط آنان را افزایش می‌دهد (فتاحی ۱۳۸۵). موتورهای کاوش مدعی هستند آنچه که از طریق محدود کردن عنوان و نشانی بازیابی می‌شود، از بیشترین ربط برخوردار است (Greg 2004)، اما آنچه که جای بحث در این زمینه دارد این است که آیا این نتایج از ربط مناسبی از دیدگاه کاربر برخوردار هستند؟ به عبارت دیگر، در ذخیره و بازیابی مدارک وبی میزان کارآمدی عنوان و نشانی اینترنتی صفحات چگونه است؟ برای مثال، آیا کلیدواژه‌های موجود در نشانی منابع وبی از ارزش کافی برای نمایه‌سازی و نیز جستجو برخوردار هستند؟ در اصل، آیا تفاوتی در میزان ربط نتایج جستجوی مبتنی بر عنوان و مبتنی بر نشانی اینترنتی منابع وبی، در حوزه‌های موضوعی مختلف وجود دارد؟ در نهایت، کاربران حوزه‌های موضوعی مختلف با جستجو در کدام قسمت به نتایج مرتبط‌تر دست می‌یابند؟

۲. هدف‌های پژوهش

هدف‌های این پژوهش عبارت‌اند از:

(۱) مقایسه تفاوت میزان حضور کلیدواژه‌های موضوعی در عنوان و نشانی اینترنتی منابع وبی با این منظور که مشخص کند حضور کلیدواژه‌های موضوعی در کدام قسمت (عنوان یا نشانی اینترنتی منابع وبی) بیشتر است و به طور بالقوه کدام قسمت قابلیت نمایه‌سازی و جستجوی کارآمدتری دارد.

(۲) مقایسه میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی در حوزه‌های مختلف موضوعی با این منظور که مشخص کند جستجو در عنوان یا در نشانی اینترنتی در کدام حوزه‌های موضوعی (علوم انسانی و کشاورزی) به نتایج بهتری خواهد انجامید.

(۳) بررسی میزان ربط نتایج جستجوی مبتنی بر نشانی اینترنتی در مقایسه با ربط نتایج جستجوی مبتنی بر عنوان با این منظور که مشخص کند کاربران در کدام قسمت می‌توانند جستجوی موفق‌تری داشته باشند.

۳. فرضیه‌های پژوهش

۱. میزان همخوانی کلیدواژه‌های نشانی اینترنتی منابع وبی و کلیدواژه‌های عنوان آنها با افزایش تعداد کلیدواژه‌های مورد جستجو کاهش می‌یابد.

۲. میان حوزه‌های موضوعی مورد بررسی به لحاظ تعداد کلیدواژه‌های موضوعی موجود در عنوان و در نشانی اینترنتی آنها تفاوت معنی‌دار وجود دارد.
۳. میان حوزه‌های موضوعی مورد بررسی به لحاظ میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی تفاوت معنی‌دار وجود دارد.
۴. نتایج جستجوی موضوعی مبتنی بر نشانی اینترنتی منابع وبی در مقایسه با نتایج مبتنی بر عنوان از ربط بالاتری برخوردار است.

۴. سؤال‌های پژوهش

۱. میزان همپوشانی نتایج بازیابی شده حاصل از جستجو در عنوان و در نشانی چقدر است؟
۲. رتبه صفحات مشترک (همپوشان) بازیابی شده حاصل از جستجوی عنوان و نشانی چگونه است؟

۵. پیشینه پژوهش

در زمینه ربط و نمایه‌سازی در موتورهای کاوش، پژوهش‌های زیادی با استفاده از دو معیار جامعیت و مانعیت انجام گرفته است. اما، عنوان و نشانی اینترنتی وب‌سایت‌ها از جمله مواردی بوده که با وجود توجه برخی پژوهشگران در حوزه بازیابی اطلاعات کمتر بدان توجه شده است.

۱-۵. پیشینه پژوهش در خارج

لارنس و همکارانش در پژوهش خود که تحلیل استنادی بر روی ۲۷۰۹۷۷ منبع وبی استناد شده در مجله علوم رایانه^۱ بود، اعلام کردند که استناد به نشانی اینترنتی منابع وبی رو به افزایش است، اما بسیاری از آنها هنوز نامعتبرند. آنها همچنین، به تحلیل نشانی‌های نامناسب پرداختند و دلایل نامناسب بودن آنها را ارائه دادند (Lawrence et al. 2001). اگرول، الجروی، و یو در پژوهش خود در جهت معرفی خزنده‌های هوشمند، محتوای صفحه، نشانی‌های URL، ساختار پیوند صفحه وب را از عوامل مهم در طی مرور صفحات وب توسط خزنده‌های موتورهای جستجو معرفی کردند. آنها در مقاله خود به عنوان صفحات توجهی نکردند (Aggarwal, AL-Garawi, and Yu 2001).

از سوی دیگر، ویترز، کاسون، و شریپلین در پژوهش خود که در جهت ایجاد نمایه‌ای از مجلات الکترونیکی بودند، چالش‌های ایجاد چنین نمایه‌ای را تغییرات زیاد در عنوان و نشانی

اینترنتی مجله بیان نمودند. آنها در زمینه توسعه نظام نمایه‌سازی که به‌طور خودکار اطلاعات را از فهرست‌های پیوسته استخراج می‌کنند و در وب‌سایت‌های پویا^۱ قرار می‌دهند، آزمایش‌هایی انجام دادند و متوجه شدند که چون استخراج اطلاعات از فهرست‌ها خودکار است، همپوشانی عنوان و URL اجتناب‌ناپذیر است (Withers, Casson, and Shrimplin 2002).

شاید مرتبط‌ترین پژوهش با پژوهش حاضر مربوط به کان و نگوین تای باشد که سودمندی نشانی اینترنتی را با تجزیه نشانی اینترنتی و وزن‌دهی به کلیدواژه‌های آن ثابت کردند (Kan and Nguyen Thi 2004).

شین و کارگر با هدف تعیین الگوریتمی برای رده‌بندی وب مشخص کردند که اگر به‌جای محتوای متن، به نشانی اینترنتی پیوندها و مکان بصری پیوندها توجه شود، رده‌بندی سایت‌های وب امری امکان‌پذیر است. آنها توانستند الگوریتمی برای رده‌بندی وب با توجه به نشانی اینترنتی و محل قرارگیری پیوندها در منابع وبی، پیشنهاد دهند. آنان برای تعیین این الگوریتم به واژه‌های موجود در نشانی اینترنتی و محل پیوندهای اینترنتی (به‌طور مثال، پیوندهای بالای صفحه) بدون در نظر گرفتن عنوان صفحات، توجه کردند (Shin and Karger 2004).

کمیته SEO^۲ پژوهشی بر روی تأثیرگذاری کلیدواژه‌های موجود در نشانی اینترنتی در دو موتور کاوش یاهو و گوگل انجام داد که مشخص کرد هر دو موتور، حضور کلیدواژه‌ها در نشانی صفحات را عامل مهمی در رتبه‌بندی در نظر می‌گیرند (SEO Committee 2004).

کیم و کنگ در پژوهش خود در زمینه کارآمدی نمایه‌سازی صفحات وب، بر سه عنصر اطلاعات متن، اطلاعات پیوندها، و اطلاعات نشانی‌های اینترنتی تأکید کردند. یکی از نتایج اصلی پژوهش آنها این بود که اطلاعات موجود نشانی‌های اینترنتی و پیوندها برای یافتن صفحات خانگی مرتبط‌تر مفید هستند (Kim and Kang 2004).

تأثیر عناصر موجود در صفحات وب بر رؤیت‌پذیری آن صفحات، یکی دیگر از زمینه‌های پژوهشی است که نظر برخی پژوهشگران را به‌خود جلب کرده است. ژانگ و دیمیتروف در پژوهش خود که تأثیر عناصر فراداده را در رؤیت‌پذیری صفحات وب در موتورهای کاوش بررسی کردند، نشان دادند که کلیدواژه‌های استخراج‌شده از عنوان و متن صفحه وب در رؤیت‌پذیری آن در موتورهای جستجو تأثیر بسزایی دارند. در پژوهش آنان، نشانی‌های اینترنتی مورد نظر نبودند (Zhang and Dimitroff 2005).

در زمینه تأثیر عنوان و نشانی صفحات در رده‌بندی وب‌سایت‌ها نیز پژوهش‌هایی انجام

1. dynamically-generated

2. search engine optimization

شده است. فوشفکفو، کروهاشف، و کنتارو پژوهشف درباره ردهبندف منابع وب با هدف توسعه فک روش ردهبندف ففد انجام دادند. در افن پژوهش، برای آزمافش روشف در ردهبندف کردن مدارک وب از عنوان و نشانی افترنتف استفاده کردند و نتافچ نشان داد که روش ففنهادهف با استفاده از عوامل فادشده از ففش از ۵۰٪ دقت برخوردار است (Yoshikiyo and Kurohashi 2008).

۲-۵. ففشفه پژوهش در افان

در افان فف پژوهشهای مربوط به موتورهای کاوش و تأففر عناصر اففاعاتف بر کارآمدف آنها مورد توجه قرار گرفته است. در عفن حال، پژوهش در زمفنه کارآمدف عنوان و نشانی صفحات وب ففبار اندک بوده است.

درباره عنوان و نشانی افترنتف، فتاحف (۱۳۸۵) در پژوهش خود که با هدف بسط جستجو انجام گرفته بود، تأففر کلفدواژههای ففر موضوعف (عمومف) در عنوان و نشانی افترنتف را به طور فداگانه در بازفبف اففاعات مرتبط در فک موتور کاوش (گوگل) نشان داد. وف در پژوهش خود خاطر نشان کرد که بسط جستجو در عنوان و در نشانی افترنتف، با استفاده از واژههای عمومف که همراه کلفدواژههای موضوعف در عنوان و نشانی افترنتف منابع و متون موجود در وب ظاهر مف شوند، مف تواند موجب افزافش میزان دقت و ربط در نتافچ بازفبف در موتورهای کاوش شود.

در پژوهشف ففگر، فتاحف و رحفمف (۱۳۸۶) محدودفب اصطلاخانماهها را در رابطه با وجود یا عدم وجود واژههای ففر موضوعف یا عام که مف تواند بر جستجو و بازفبف اففاعات در نظامهای اففاعاتف مؤثر باشد، مورد بررسی قرار دادند. آنها مشخص کردند که هرچه ففشر توصففگرها با واژههای عمومف همراه باشند، مف توانند به عمقف ترشدن نمافهسازف با هدف فراهم آوردن شرایط برای افزافش ربط و دقت در بازفبف منتهف شوند. آنان ابفدا با استفاده از روش تحلیل محتواف مقالات حوزه علوم انسانی، واژههای عمومف را که به طور معمول، همراه واژههای تخصصف در عنوان و چکفده و متن منابع افن حوزه ظاهر مف شوند، شناسافف کردند و سفاهه آنها را با محتواف اصطلاخانماههای همان حوزه مقافسه کردند. سپس، واژههای عمومف دارای بسامد بالا را با برخی کلفدواژههای موضوعف ترکیب کردند و در ففلفدهای عنوان و نشانی صفحات در موتور کاوش گوگل مورد جستجو قرار دادند. فافتهها نشان داد میزان موفقفب جستجو در کسب نتافچ مرتبط و ففقق با استفاده از واژگان عمومف در ففلفد عنوان و نشانی صفحات ففبار

بالاست. آنان نتیجه‌گیری کردند چنانچه قرار باشد زبان اصطلاحنامه‌ها به زبان جستجو نزدیک‌تر شود و بازیابی با نتایج دقیق‌تر و مرتبط‌تر همراه باشد، لازم است واژه‌ها و عبارات‌های عمومی برگرفته از عنوان و متن و نشانی صفحات به اصطلاح‌های تخصصی افزوده گردند.

روند کلی پژوهش‌ها نشان می‌دهد که نظم‌دهی شبکه جهانی وب از نابسامانی موجود و تسهیل فرایند اطلاع‌یابی کاربران، بیش از آنکه نیازمند طراحی نظام‌های بازیابی پیچیده‌تر باشد، در گروه ارتقای سطح دانش طراحان این نظام‌ها درباره چگونگی تعامل کاربران با این محیط، درک نیازهای کاربران، و ساده‌سازی مراحل جستجو است (منصوریان ۱۳۸۶). هنوز پرسش‌های فراوانی در ارتباط با چگونگی تعامل کاربران با محیط وب در حین فرایند جستجو وجود دارد. پاسخگویی به این پرسش‌ها نیازمند انجام پژوهش‌های تازه است. درباره راهکارهای جستجو در یک موتور کاوش و بررسی ربط از دید کاربران، پژوهش زیادی انجام نگرفته است. بسیاری از موتورهای کاوش گزینه‌هایی برای محدود کردن نتایج جستجو به عنوان و نشانی اینترنتی منابع وبی دارند. با کاوش در این قسمت‌ها نتایج بسیار متفاوتی به دست می‌آید.

با توجه به این مطلب، مقایسه میزان کارآمدی عنوان و نشانی اینترنتی در بسط جستجو و بررسی ربط نتایج به دست آمده از دید کاربر ضروری به نظر می‌رسد. با وجود کثرت پژوهش‌ها در موضوع ذخیره و بازیابی اطلاعات در موتورهای کاوش، تاکنون به عنوان و نشانی اینترنتی در بازیابی اطلاعات توجه کمی شده است. بیشتر توجه پژوهشگران به کلیدواژه‌های نشانی اینترنتی در پی ارائه الگوریتمی برای رده‌بندی وب بوده است (از جمله Shin and Karger 2004 و Yoshikiyo and Kurohashi 2008) و توجه به قابلیت‌های عنوان و نشانی اینترنتی (به خصوص نشانی اینترنتی) در نمایه‌سازی کمتر مورد توجه بوده است. در این پژوهش، سعی شده است که تاحدودی به این مطلب پرداخته شود و نتایج قابل توجهی، هم برای کاربران وب و هم برای طراحان وب‌سایت و طراحان موتورهای کاوش فراهم شود.

۶. روش پژوهش، جامعه آماری و ابزار گردآوری اطلاعات

در این پژوهش که پژوهشی کاربردی است و در مورد نتایج بازیابی حاصل از جستجوی کاربران در موتور جستجوی گوگل صورت می‌گیرد، از دو روش تحلیل محتوی و روش پیمایشی استفاده می‌گردد. روش تحلیل محتوی در مورد محتوای وب‌سایت‌ها از دیدگاه کاربر و روش پیمایشی در مورد میزان همخوانی و نیز تأثیر تعداد کلیدواژه‌ها در میزان همخوانی انجام می‌گیرد.

فامعه آماری این پژوهش (آزمودنی های پژوهش) دانشجویان کارشناسی ارشد حوزه های کشاورزی و علوم انسانی در دانشگاه فردوسی مشهد هستند. این دو حوزه با توجه به ماهیت متفاوت و نیز تنوع موضوعی آنان انتخاب شدند. با توجه به هدف و ماهیت پژوهش، برای تعمق بیشتر در این پژوهش، گروه کوچکی از کاربران به نمایندگی از کل فامعه پژوهش انتخاب شدند. تعداد ۶۰ نفر دانشجوی کارشناسی ارشد (۳۰ نفر دانشجوی کارشناسی ارشد حوزه علوم انسانی و ۳۰ نفر دانشجوی کارشناسی ارشد حوزه کشاورزی) ترم سه به بعد (دانشجویانی که در مرحله نوشتن طرح پیشنهادی پایان نامه و یا نوشتن پایان نامه و علاقه مند به همکاری با پژوهشگر در این زمینه در حوزه های موضوعی علوم انسانی و کشاورزی بودند) برای بررسی میزان ربط نتایج در وبسایت های بازبایی شده انتخاب شدند. دلیل محدودسازی فامعه پژوهش به این دانشجویان آن بود که آنها با توجه به انجام پایان نامه خود به طور جدی به اطلاعات و منابع مرتبط بیشتر نیاز داشتند و می توانستند براساس فضای واقعی به پرسشنامه پاسخ دهند.

این دانشجویان با توجه به نیاز اطلاعاتی خود به جستجو در موتور کاوش گوگل پرداختند. علت انتخاب موتور جستجوی گوگل آن بود که موتور جستجوی گوگل پر استفاده ترین موتور جستجو در بین دانشجویان تحصیلات تکمیلی است. هر یک از آنان بنا بر نیاز اطلاعاتی خود، تعداد پنج کلیدواژه را مورد کاوش قرار دادند و هر یک از کلیدواژه ها را یک بار در عنوان^۱ و بار دیگر در نشانی اینترنتی^۲ وارد کردند. آزمودنی های پژوهش پس از جستجو در موتور کاوش، موارد بازبایی شده توسط موتور کاوش (۱۰ نتیجه اول) را به صورت یکی از موارد خیلی زیاد، زیاد، متوسط، کم، و خیلی کم تعیین می کردند.

از هر یک از آزمودنی ها خواسته شد تا کلیدواژه های مورد جستجوی خود را به طور دقیق و بدون هیچ گونه تغییری روی سیاهه واریسی یادداشت کند و به پژوهشگر بدهد که برای بررسی تعداد و میزان همخوانی کلیدواژه ها در عنوان و نشانی (این بار توسط خود پژوهشگر) مورد استفاده قرار گیرد. به منظور بررسی همخوانی بین عنوان و نشانی اینترنتی در حوزه های کشاورزی و علوم انسانی، ابتدا کلیدواژه های مورد جستجو در دو حوزه کشاورزی و علوم انسانی به سه گروه یک واژه ای، دو واژه ای، و سه واژه ای تقسیم شدند، سپس تعداد ۴۵ کلیدواژه از حوزه کشاورزی (۱۵ کلیدواژه یک واژه ای، ۱۵ کلیدواژه دو واژه ای، و ۱۵ کلیدواژه سه واژه ای) و تعداد ۴۵ کلیدواژه از حوزه علوم انسانی (۱۵ کلیدواژه یک واژه ای، ۱۵ کلیدواژه دو واژه ای، و ۱۵ کلیدواژه سه واژه ای) به طور تصادفی ساده انتخاب شدند (با استفاده از سیاهه واریسی شماره ۱ که بیشتر توسط آزمودنی ها تکمیل شده بود).

1. in title

2. in URL

بنابراین، در هر گروه (یک‌واژه‌ای و دوواژه‌ای و سه‌واژه‌ای) ۳۰ کلیدواژه (۱۵ کلیدواژه در حوزه کشاورزی و ۱۵ کلیدواژه در حوزه علوم انسانی) مورد بررسی قرار گرفت و میزان همخوانی کلیدواژه‌های موجود در نشانی اینترنتی منابع وبی با کلیدواژه‌های عنوان آنها به‌طور جداگانه در هر قسمت در سیاهه‌های مربوط ثبت شد. برای بررسی تعداد کلیدواژه‌ها در عنوان و نشانی اینترنتی هر حوزه نیز تعداد ۶۰ کلیدواژه از هر حوزه انتخاب شد (۳۰ کلیدواژه در حوزه کشاورزی و ۳۰ کلیدواژه در حوزه علوم انسانی) و تعداد کلیدواژه‌های موجود در عنوان و نشانی اینترنتی هر حوزه محاسبه و در یک سیاهه ثبت گردید. برای هر کلیدواژه، ۳۰ نتیجه اول مورد بررسی قرار گرفت. علت انتخاب ۳۰ نتیجه اول این بود که بیشتر کاربران اینترنت سه صفحه اول نتایج را مورد بررسی قرار می‌دهند.

همان‌گونه که اشاره شد، به‌منظور گردآوری اطلاعات مورد نیاز این پژوهش از سیاهه‌های واری استفاده شد. هسی-ئی در بخشی از پژوهش خود سیاهه‌ای تهیه کرد که در آن کاربران به میزان ربط نتایج بازیابی شده در ۸ موتور جستجو به ترتیب مرتبط بودن، از یک تا ۱۰ امتیاز می‌دادند (Hesi-Yee 1998 در هسی-ئی ۱۹۹۸). با استفاده از این سیاهه و براساس مطالعه متون و دیدگاه‌های صاحب‌نظران، سیاهه‌ای مشابه برای ۱۰ گزینه اول بازیابی شده تهیه شد و در معرض پیش‌آزمون قرار گرفت. این سیاهه در اختیار ۱۰ دانشجوی کارشناسی ارشد (۵ دانشجو در حوزه کشاورزی و ۵ دانشجو در حوزه علوم انسانی) قرار گرفت. در این مرحله، تغییراتی در سیاهه اولیه به پیشنهاد آزمودنی‌های مرحله پیش‌آزمون، جهت تسهیل در ارزیابی ربط مدارک انجام گرفت. بنابراین، پس از چندبار بازنگری سیاهه‌ها و پیش‌آزمون آنان، سیاهه‌های واری جهت ارزیابی ربط به جامعه پژوهش (آزمودنی‌ها) ارائه شد.

۷. تجزیه و تحلیل داده‌ها و ارائه نتایج

فرضیه ۱: میزان همخوانی کلیدواژه‌های نشانی اینترنتی منابع وبی و کلیدواژه‌های عنوان آنها با افزایش تعداد کلیدواژه‌های مورد جستجو کاهش می‌یابد.

برای پاسخ به این فرضیه، تعداد ۴۵ کلیدواژه از هر حوزه موضوعی انتخاب شد و به قسمت‌های یک‌واژه‌ای و دوواژه‌ای و سه‌واژه‌ای تقسیم شد. میزان همخوانی در ۳۰ وب‌سایت اول نتایج موتور جستجو سنجیده شد. ابتدا، با استفاده از آزمون تک‌متغیری کلموگروف اسمیرنوف به تعیین میزان نرمال بودن داده‌ها پرداخته شد (جدول ۱).

جدول ۱. آزمون تک‌متغیری کلموگروف اسمیرنف برای میزان همخوانی عنوان و نشانی اینترنتی

تعداد داده	میانگین	انحراف معیار	p-value
۹۰	۱۰/۸۷	۶/۹۳	۰/۲۹۰

با توجه به مقدار $p (p\text{-value} = 0.290 > 0.05)$ که نشان می‌دهد توزیع داده‌ها در متغیر میزان همخوانی عنوان و نشانی اینترنتی براساس تعداد کلیدواژه‌ها نرمال است، از ضریب همبستگی پیرسون استفاده شد (جدول ۲).

جدول ۲. ضریب همبستگی پیرسون برای تعداد کلیدواژه‌ها و میزان همخوانی

تعداد	ضریب همبستگی پیرسون	- مقدار p
۹۰	-۰/۸۱۷	۰/۰۰

با توجه به ضریب همبستگی $r = -0.817$ همبستگی به نسبت کامل، اما منفی است. یعنی، با افزایش تعداد کلیدواژه مورد جستجو میزان همخوانی کاهش می‌یابد. بنابراین، فرضیه یک تأیید می‌شود.

فرضیه ۲: میان حوزه‌های موضوعی مختلف (مورد بررسی) به لحاظ تعداد کلیدواژه‌های موضوعی موجود در عنوان و در نشانی اینترنتی آنها تفاوت وجود دارد. برای پاسخ به این فرضیه، تعداد وب‌سایت‌هایی که کلیدواژه موضوعی مورد نظر در عنوان و نشانی اینترنتی آنها وجود داشت، ثبت شد. با استفاده از آزمون تک‌متغیری کلموگروف اسمیرنف می‌توان وضعیت جامعه را مشخص کرد (جدول ۳).

جدول ۳. آزمون تک‌متغیری کلموگروف اسمیرنف برای تعداد کلیدواژه‌های عنوان و نشانی اینترنتی حوزه‌های علوم انسانی و علوم کشاورزی

تعداد	میانگین	انحراف معیار	- مقدار p
۳۰	۲۵/۱۳	۵/۰۵	۰/۳۶۵
۳۰	۱۵/۴۳	۶/۷۸	۰/۵۰۶
۳۰	۲۵/۱۰	۳/۵۷	۰/۳۸۰
۳۰	۱۴/۴۳	۴/۴۳	۰/۵۹۲

با توجه به مقادیر $p > 0.05$ ، جامعه آماری نرمال است و از آزمون‌های پارامتری مقایسه میانگین دو جامعه استفاده می‌شود.

جدول ۴ میانگین و انحراف استاندارد تعداد کلیدواژه‌های موضوعی موجود در عنوان حوزه‌های موضوعی مورد بررسی را با استفاده از آزمون t نشان می‌دهد.

جدول ۴. تعداد کلیدواژه‌های موضوعی در عنوان دو حوزه علوم انسانی و کشاورزی

حوزه موضوعی	تعداد کلیدواژه	میانگین	انحراف استاندارد	p-value
علوم انسانی	۳۰	۲۵/۱۳	۵/۰۵	۰/۷۷۹
علوم کشاورزی	۳۰	۲۵/۱۰	۳/۵۷	۰/۷۷۹

با توجه به $P\text{-value} = 0.779 > 0.05$ فرض صفر تأیید می‌شود و میان حوزه‌های موضوعی مختلف (مورد بررسی) به لحاظ تعداد کلیدواژه‌های موضوعی موجود در عنوان تفاوت معنی‌دار وجود ندارد.

جدول ۵ میانگین و انحراف استاندارد تعداد کلیدواژه‌های موضوعی موجود در نشانی حوزه‌های موضوعی مورد بررسی را با استفاده از آزمون t نشان می‌دهد:

جدول ۵. تعداد کلیدواژه‌های موضوعی در نشانی اینترنتی دو حوزه کشاورزی و علوم انسانی

حوزه موضوعی	تعداد کلیدواژه	میانگین	انحراف استاندارد	p-value
علوم انسانی	۳۰	۱۵/۴۳	۶/۷۸	۰/۵۱۱
علوم کشاورزی	۳۰	۱۴/۴۳	۴/۷۳	۰/۵۱۱

با توجه به $P\text{-value} = 0.511 > 0.05$ فرض صفر تأیید می‌شود و میان حوزه‌های موضوعی مختلف (مورد بررسی) به لحاظ تعداد کلیدواژه‌های موضوعی موجود در نشانی اینترنتی تفاوت معنی‌دار وجود ندارد.

فرضیه ۳: میان حوزه‌های موضوعی مورد بررسی به لحاظ میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی تفاوت وجود دارد.

برای پاسخ‌گویی به این فرضیه، از میانگین‌های به دست آمده برای جستجوی مبتنی بر عنوان و نشانی اینترنتی برای هر کلیدواژه (که با توجه به سیاهه‌های شماره یک، توسط آزمودنی‌ها تکمیل شده‌اند) استفاده شد.

ابتدا با استفاده از آزمون تک‌متغیری کلموگروف اسمیرنوف به تعیین میزان نرمال بودن داده‌های عنوان و نشانی در هر دو حوزه پرداخته شد (جدول ۶).

جدول ۶. آزمون تک‌متغیری کلموگروف اسمیرنوف برای تعیین میزان نرمال بودن داده‌های عنوان و نشانی در هر دو حوزه

حوزه	میانگین	انحراف معیار	p-value
کشاورزی	۲/۸	۰/۵۶	۰/۰۰۶
علوم انسانی	۲/۷	۲/۱	۰/۰۰۰

با توجه به مقدار p ($p\text{-value} < 0.05$)، توزیع داده‌ها در متغیرهای میزان ربط عنوان و نشانی اینترنتی در هر دو حوزه نرمال نیست، از این رو از آزمون غیرپارامتری نمونه‌های جفت و وابسته استفاده می‌شود. آزمون ویلکاکسون از آزمون‌های آماری غیرپارامتری مناسب برای نمونه‌های جفت و وابسته است (هویدا ۱۳۸۷).

جدول ۷. میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی با استفاده از آزمون ویلکاکسون

حوزه	میانگین	انحراف معیار	p-value
کشاورزی	۰/۳۴	۰/۶۱	۰/۰۰۰
علوم انسانی	۰/۳۶	۲/۹	۰/۴۳۵

با توجه به مقادیر P ($P\text{-value} = 0.00 < 0.05$)، بین میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی در حوزه کشاورزی تفاوت معنی‌دار وجود دارد، اما بین میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی در حوزه علوم انسانی تفاوت معنی‌دار وجود ندارد ($P\text{-value} = 0.435 > 0.05$).

با توجه به نتایج فرضیه ۳، تفاوت میزان ربط بین عنوان و نشانی اینترنتی دو حوزه کشاورزی و علوم انسانی ناشی از تفاوت تعداد کلیدواژه در عنوان و نشانی اینترنتی آنها نیست، بلکه می‌توان ادعا نمود که تفاوت‌های موجود در میزان ربط حوزه‌های مورد بررسی ناشی از تفاوت در ماهیت نیازهای اطلاعاتی دانشجویان است. دانشجویان حوزه کشاورزی بیشتر به دنبال اسامی و ابزارهای خاص هستند، اما دانشجویان حوزه علوم انسانی بیشتر به دنبال موضوعات مفهومی هستند.

فرضیه ۴: نتایج جستجوی موضوعی مبتنی بر نشانی اینترنتی منابع وبی در مقایسه با نتایج مبتنی بر عنوان از ربط بالاتری برخوردار است.

از آنجا که بین میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی در حوزه علوم انسانی تفاوت معنی‌دار وجود ندارد، این امر در حوزه موضوعی کشاورزی سنجدیده می‌شود.

با توجه به فرض صفر و یک:

فرض صفر: جستجوی موضوعی مبتنی بر نشانی اینترنتی منابع وبی در مقایسه با نتایج مبتنی بر عنوان از ربط بالاتری برخوردار نیست.

فرض یک: جستجوی موضوعی مبتنی بر نشانی اینترنتی منابع وبی در مقایسه با نتایج مبتنی بر عنوان از ربط بالاتری برخوردار است.

برای پاسخ به این فرضیه از آزمون فرض یک طرفه استفاده می شود (جدول ۸).

جدول ۸. آزمون یک طرفه برای میانگین های حاصل از جستجو در عنوان و نشانی اینترنتی حوزه کشاورزی

P-value	درجه آزادی	حوزه کشاورزی
۰/۰۰۰	۱۴۹	

با توجه به مقدار $P (P\text{-value}=0.00 < 0.05)$ ، فرض یک تأیید می شود و جستجوی موضوعی مبتنی بر نشانی اینترنتی منابع وبی در مقایسه با نتایج مبتنی بر عنوان در حوزه کشاورزی از ربط بالاتری برخوردار است. همان گونه که پیشتر اشاره شد، نشانی اینترنتی یک وبسایت از فایل ها و پوشه های مختلفی تشکیل شده است؛ این فایل ها و پوشه ها هر یک دارای نام خاصی هستند. با توجه به ماهیت رشته کشاورزی انتظار می رود اسامی تشکیل دهنده این فایل ها کلمات تخصصی حوزه کشاورزی باشند که در مرحله ایجاد صفحات وب به آنها اختصاص داده شده اند و در هنگام جستجو در نشانی اینترنتی، بازیابی می شوند و میزان ربط را به طور قابل توجهی بالا می برند.

۸. پاسخ به سؤالات پژوهش

سؤال ۱: میزان همپوشانی نتایج بازیابی شده حاصل از جستجو در عنوان و در نشانی اینترنتی چقدر است؟

در این قسمت به بررسی نتایج بازیابی شده حاصل از جستجو در عنوان و در نشانی اینترنتی و نیز به مقایسه رتبه های موارد همپوشان پرداخته شده است. ابتدا، رتبه وبسایت مورد نظر در جستجوی مبتنی بر عنوان، سپس رتبه همان سایت در جستجوی مبتنی بر نشانی اینترنتی ثبت شد. در این قسمت، اگر وبسایت مورد نظر در جستجوی مبتنی بر عنوان دارای رتبه پایین تر از جستجوی مبتنی بر نشانی اینترنتی بوده است، آن سایت دارای افزایش رتبه در جستجوی مبتنی بر نشانی اینترنتی بوده است (به عنوان مثال، در جستجوی مبتنی بر عنوان دارای رتبه ۸ و در

جستجوی مبتنی بر نشانی اینترنتی دارای رتبه ۲ است). به همین ترتیب، اگر وبسایت مورد نظر در جستجوی مبتنی بر عنوان دارای رتبه بیشتر از جستجوی مبتنی بر نشانی اینترنتی بوده است، آن سایت دارای کاهش رتبه در جستجوی مبتنی بر نشانی اینترنتی بوده است (به عنوان مثال، در جستجوی مبتنی بر عنوان دارای رتبه ۱ و در جستجوی مبتنی بر نشانی اینترنتی دارای رتبه ۲ است). همین طور برخی از وبسایتها در هر دو مورد جستجو دارای یک رتبه هستند که رتبه آن وبسایتها ثابت است (به عنوان مثال، در جستجوی مبتنی بر عنوان دارای رتبه ۱ و در جستجوی مبتنی بر نشانی اینترنتی هم دارای رتبه ۱ است).

جدولهای ۹ و ۱۰ نمایانگر میزان همپوشانی و نیز تغییر رتبه نتایج بازیابی شده حاصل از جستجو در عنوان و در نشانی اینترنتی دو حوزه علوم انسانی و کشاورزی است.

جدول ۹. میزان همپوشانی و تغییر رتبه نتایج بازیابی شده حاصل از جستجو در عنوان و در نشانی اینترنتی حوزه کشاورزی

میزان ربط	افزایش رتبه		کاهش رتبه		رتبه ثابت		جمع
	تعداد سایت	درصد	تعداد سایت	درصد	تعداد سایت	درصد	
بسیار مرتبط	۷۲	۳۸/۷۰	۳	۱/۶۱	۲۹	۱۵/۵۹	۵۵/۹۱
مرتبط	۳۳	۱۷/۷۴			۷	۳/۷۶	۲۱/۵۰
کم ربط	۱۶	۸/۶۰	۱	۰/۵۳	۶	۳/۲۲	۱/۲۳
نامرتبط	۱۳	۶/۹۸			۶	۳/۲۲	۱/۰۲
جمع	۱۳۴	۷۲/۰۴	۴	۲/۱۵	۴۸	۲۵/۸۰	۱۰۰

جدول ۱۰. میزان همپوشانی و تغییر رتبه نتایج حاصل از جستجو در عنوان و نشانی حوزه علوم انسانی

میزان ربط	افزایش رتبه		کاهش رتبه		رتبه ثابت		جمع
	تعداد سایت	درصد	تعداد سایت	درصد	تعداد سایت	درصد	
بسیار مرتبط	۶۲	۳۳/۸۷	۳	۱/۶۳	۲۴	۱۳/۱۱	۴۸/۶۳
مرتبط	۳۶	۱۹/۶۷	۱	۰/۵۴	۳	۱/۶۳	۲۱/۸۵
کم ربط	۲۵	۱۳/۶۶	-	-	۲	۱/۰۹	۱۴/۷۵
نامرتبط	۲۵	۱۳/۶۶	-	-	۲	۱/۰۹	۱۴/۷۵
جمع	۱۴۸	۸۰/۸۷	۴	۲/۱۸	۳۱	۱۶/۹۳	۱۰۰

با توجه به جدول‌های ۹ و ۱۰، در حوزه کشاورزی از تعداد ۲۸۱۸ وبسایت مورد بررسی، ۱۸۶ وبسایت (۶/۶٪) و در حوزه علوم انسانی، از تعداد ۲۸۰۹ وبسایت مورد بررسی، ۱۸۳ وبسایت (۶/۵٪) همپوشانی داشته‌اند. با توجه به این یافته‌ها، میزان همپوشانی در جستجوهای مبتنی بر عنوان و نشانی اینترنتی دارای نسبتی کم است که نشان‌دهنده تفاوت نتایج جستجو حاصل از جستجو در عنوان و نشانی اینترنتی و نیز تفاوت در الگوریتم رتبه‌بندی نتایج جستجوهای عنوان و نشانی اینترنتی است. کاربران با جستجو در هر دو مورد به نتایج متفاوتی دست می‌یابند.

سؤال ۲: رتبه صفحات مشترک (همپوشان) بازیابی شده حاصل از جستجوی عنوان و نشانی چگونه است؟

با توجه به جدول ۹، در حوزه کشاورزی ۷۲/۰۴٪ از نتایج همپوشان از افزایش رتبه، ۲/۱۵٪ از نتایج از کاهش رتبه برخوردارند و ۲۵/۸۰٪ از نتایج بدون تغییر رتبه و دارای رتبه یکسان و ثابتی در نتایج بازیابی هستند. همچنین، ۵۵/۹۱٪ از نتایج دارای میزان ربط بسیار بالا، ۲۱/۵۰٪ از نتایج دارای میزان ربط متوسط، ۱/۲۳٪ از نتایج کم‌ربط و ۱/۰۲٪ از نتایج نامربط هستند.

همچنین با توجه به جدول ۱۰، در حوزه علوم انسانی ۸۰/۸۷٪ از نتایج همپوشان از افزایش رتبه، ۲/۱۸٪ از نتایج از کاهش رتبه برخوردارند و ۱۶/۹۳٪ از نتایج بدون تغییر رتبه و دارای رتبه ثابت هستند. همچنین، ۴۸/۶۳٪ از نتایج دارای میزان ربط بسیار مرتبط، ۲۱/۸۵٪ از نتایج دارای میزان ربط مرتبط، ۱۴/۷۵٪ از نتایج کم‌ربط، و ۱۴/۷۵٪ از نتایج دارای میزان ربط نامربط هستند.

با توجه به اعداد بالا، موارد همپوشان در هر دو حوزه مورد بررسی بیشتر اوقات از میزان ربط بالا و افزایش رتبه در جستجوی مبتنی بر نشانی اینترنتی برخوردارند. موارد همپوشان مواردی بودند که کلیدواژه مورد جستجو در عنوان و نشانی اینترنتی آنها تکرار شده‌اند، یعنی میزان همخوانی بین عنوان و نشانی اینترنتی در این موارد بالاست.

۹. نتیجه‌گیری

با توجه به اینکه در این پژوهش از کاربران دو حوزه علوم انسانی و علوم کشاورزی استفاده شده است و این دو حوزه در درون خود دارای تنوع‌هایی هستند و نیز جامعه این پژوهش محدود به دانشجویان دانشگاه فردوسی شده و از یک موتور کاوش (موتور کاوش گوگل) استفاده شده است، برای تعمیم دادن نتایج به کل کاربران و نیز محیط وب احتیاط بیشتری صورت می‌گیرد.

گرچه ربط امری نسبی است، در نظر گرفتن عوامل مختلف تأثیرگذار بر میزان ربط می‌تواند تا حدودی مسأله ربط یا تطابق میان واژگان عبارت پرشش و مدارک موجود در مجموعه را حل کند. تعداد کلیدواژه مورد جستجو و محل قرارگیری آنها از عوامل تأثیرگذار در بالا بردن میزان ربط نتایج جستجو است. بر این اساس، بسیاری از موتورهای جستجو، گزینه‌هایی برای محدود کردن واژه‌ها و عبارات مورد جستجوی کاربران در مکان‌های خاص (مانند عنوان، متن، نشانی اینترنتی، و پیوندها) قرار داده‌اند. در این پژوهش تلاش شد تا به بررسی میزان ربط با استفاده از بسط جستجو و امکانات موتور جستجوی گوگل پرداخته شود. هدف این پژوهش، بررسی میزان ربط نتایج بازایی شده حاصل از جستجوی مبتنی بر عنوان و جستجوی مبتنی بر نشانی اینترنتی و همچنین، تأثیر تعداد کلیدواژه مورد جستجو بر میزان همخوانی واژگانی بین عنوان و نشانی اینترنتی در دو حوزه علوم کشاورزی و علوم انسانی بوده است.

یافته‌ها نشان داد جستجو براساس نشانی اینترنتی در حوزه‌های موضوعی تخصصی نظیر حوزه علوم کشاورزی، نتایج مرتبط‌تری را به دنبال دارد. همچنین، تعداد کلیدواژه‌های عبارت جستجو نیز می‌تواند در میزان ربط نتایج تأثیرگذار باشد. در موارد همپوشان، که کلیدواژه مورد جستجو هم در عنوان و هم در نشانی اینترنتی آنان تکرار شده است نیز میزان ربط بالایی وجود دارد. این موارد بیشتر دارای رتبه بالایی در صفحه نتایج جستجو بودند. به این ترتیب، همخوانی بین عنوان و نشانی اینترنتی یک صفحه وب بیانگر ارتباط بیشتر آن صفحه با موضوع مورد نظر است. همچنین، همخوانی بین عنوان و نشانی اینترنتی یک صفحه وب می‌تواند در بالا بردن رتبه آن وبسایت در صفحه ارائه نتایج موتور کاوش تأثیر زیادی داشته باشد.

از سوی دیگر، تفاوت معنی‌داری بین تعداد کلیدواژه‌های موجود در عنوان و نشانی اینترنتی وبسایت‌های دو حوزه وجود نداشت. بنابراین می‌توان نتیجه گرفت که تفاوت میزان ربط نتایج حاصل از جستجو در عنوان و نشانی اینترنتی دو حوزه موضوعی ریشه در کلیدواژه‌های دو حوزه دارد. با بررسی کلیدواژه‌های مورد جستجو مشخص شد که کاربران حوزه کشاورزی در بیشتر موارد در جستجوی اسامی خاص (مانند نام ویروس‌ها، باکتری‌ها، بیماری‌ها و ...) و یا ابزارهای خاص حوزه (مانند نام آنتی‌سرم‌های خاص، وسایل خاص و ...) و کاربران حوزه علوم انسانی بیشتر به دنبال موارد مفهومی (مانند موضوع‌ها، رویدادها، علت‌ها و ...) بودند.

۱۰. پیشنهادها و رهنمودهای پژوهش

پیشنهادهای این پژوهش که برخاسته از یافته‌ها و نتایج آن است، در سه مقوله ارائه می‌شود:

۱۰-۱. پیشنهادهایی به طراحان وبسایت‌ها برای طراحی وبسایت‌ها به منظور بالا بردن رتبه وبسایت در موتورهای جستجو

یافته‌های این پژوهش نشان داد که میان نتایج جستجوهای که کلیدواژه مورد جستجو در عنوان و نشانی اینترنتی همخوانی بیشتری وجود دارد، میزان ربط بالاتر است. این مطلب در هر دو حوزه صادق بود. با توجه به این مطلب، به صاحبان و طراحان وبسایت‌ها توصیه می‌گردد که در انتخاب کلیدواژه‌ها برای عنوان و نشانی اینترنتی وبسایت دقت نظر داشته باشند. هنگامی که واژه‌های مرتبط در هر دو منطقه عنوان و نشانی حضور داشته باشند، سایت مورد نظر از رتبه بالایی در بازیابی برخوردار می‌شود. همچنین، طراحان صفحات وب را می‌توان تشویق نمود تا برای نام‌گذاری عنوان صفحات و نیز برای تدوین نشانی اینترنتی از کلیدواژه‌ها و عبارات تخصصی معنی‌دار استفاده کنند تا میزان بازنمونی صفحات در موتورهای جستجو افزایش یابد. همان‌طور که پیشتر نیز گفته شد وبسایت‌هایی که رهنمودهایی برای طراحان وب دارند، آنها را به استفاده از واژه‌های ساده و کوتاه تشویق می‌کنند. در این پژوهش نیز مشخص شد وبسایت‌هایی که دارای کلیدواژه‌های تک‌واژه‌ای هستند، از همخوانی بالایی برخوردارند. بنابراین، طراحان وبسایت بهتر است که از واژه‌هایی کوتاه و ساده استفاده نمایند.

۱۰-۲. پیشنهادهایی به کاربران و جستجوگران وب به منظور یافتن نتایج مرتبط‌تر با نیاز اطلاعاتی آنان

نتایج این پژوهش نشان داد وبسایت‌هایی که اطلاعات و اصطلاحات خاص و تخصصی را در نشانی اینترنتی خود دارند، دارای ربط بالایی بودند. همچنین، میزان ربط وبسایت‌هایی که بین عنوان و نشانی اینترنتی آنها همخوانی بالایی وجود دارد، بالا است و نیز هنگامی که کلیدواژه مورد جستجو تک‌واژه‌ای باشد، میزان همخوانی عنوان و نشانی اینترنتی بالا می‌رود. بنابراین، کاربران و جستجوگران وب که به اطلاعات تخصصی نیاز دارند می‌توانند با محدود کردن عبارات و کلیدواژه‌های جستجو به عنوان و نشانی (به‌خصوص در مورد اسامی خاص، به نشانی اینترنتی) به نتایج مرتبط‌تری دست یابند. همچنین، با کنترل تعداد کلیدواژه‌های مورد جستجو می‌توانند به اطلاعات مفیدتری دست پیدا نمایند.

۱۰-۳. پیشنهادهایی به طراحان موتورهای کاوش

در این پژوهش، از یک موتور کاوش عمومی برای پژوهش استفاده شد، اما نتایج آن برای موتورهای کاوش تخصصی نیز می‌تواند مناسب و مفید باشد. موتورهای کاوش به‌خصوص موتورهای کاوش تخصصی می‌توانند وزن بیشتری برای کلیدواژه‌های موجود در نشانی اینترنتی

صفحات وب در نظر گیرند. از آنجاکه آنها بیشتر، توجه زیادی به کلیدواژه‌های موجود در عنوان صفحات وب می‌کنند، اگر به کلیدواژه‌های موجود در نشانی اینترنتی نیز توجه بیشتری نمایند، می‌توانند نتایج مرتبط‌تری را رتبه‌بندی کنند و در اختیار کاربران قرار دهند.

۱۱. منابع

- رمضانی درح، عباسعلی. ۱۳۸۴. معیارهای قضاوت ربط مدارک الکترونیکی از دیدگاه دانشجویان تحصیلات تکمیلی دانشگاه فردوسی و مقایسه آن با الگوریتم‌های رتبه‌بندی موتورهای کاوش. پایان‌نامه کارشناسی ارشد، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه فردوسی مشهد.
- فتاحی، رحمت‌اله. ۱۳۸۵. شناسایی و تحلیل واژگان عمومی در منابع وب: رویکردی نو به بسط عبارت جستجو با استفاده از زبان طبیعی در موتورهای کاوش. *مطالعات تربیتی و روانشناسی* ۷(۱): ۳۱-۵۲.
- فتاحی، رحمت‌الله، و ماریه رحیمی. ۱۳۸۶. جای خالی واژگان عمومی در اصطلاحنامه‌ها: نگاهی به دشواریهای تحقق ربط و دقت در بازیابی اطلاعات با استفاده از زبان کنترل‌شده. در *همایش ملی اصطلاحنامه، قم*. <http://profsite.um.ac.ir/~fattahi/farsipub.htm> (دسترسی در ۵ خرداد ۱۳۸۷).
- منصوریان، یزدان. ۱۳۸۶. عوامل مؤثر بر جستجو و بازیابی اطلاعات در شبکه جهانگستر وب. *مجله کتابدار*. <http://www.ketabdar.org/magazine/detailarticle.asp?number=23> (دسترسی در ۵ خرداد ۱۳۸۷).
- هسی-ئی اینگرید. ۱۹۹۸. اینترنت: سازماندهی و جستجو. ترجمه قاسم آزادی. ۱۳۸۲. *فصلنامه اطلاع‌رسانی* ۱۸ (۳ و ۴): ۹۴-۱۰۳.
- Aggarwal, C., F. Al-Garawi, and P. Yu. 2001. Intelligent crawling on the World Wide Web with arbitrary predicates. In *Proceedings of the Tenth International World Wide Web Conference*, Michael RL, ed., 96-105. Hong Kong: ACM Press.
- Greg, R. *Review of HotBot (Inktomi)* (April 15, 2004). <http://www.searchengineshowdown.com/features/hotbot/> (accessed Nov. 15 2008).
- Kan, Min-Yen, Nguyen Thi, and Hoang Oanh. 2004. Fast webpage classification using URL features. In *Fast webpa Proceedings of the 14th ACM International Conference on Information and Knowledge Management*. ACM press, 325-326. <http://dl.acm.org/citation.cfm?doid=1099554.1099649> (accessed Nov. 15 2008).
- Kang, In-Ho, and Gil Chang Kim. 2004. Integration of multiple evidences based on a query type for web search. *Information Processing & Management* 40 (3): 459-478.
- Lawrence, Steve, et al. 2001 Persistence of Web references in scientific research. *Computer magazine* 34 (2): 26-31.
- SEO Committee. 2004. Importance of keywords in URLs and search engine ranking. *Search Engine Journal*. <http://www.searchenginejournal.com/importance-of-keywords-in-urls-and-search-engine-ranking/1019/> (accessed 21 Mar. 2008).
- Shin, L.-K., and D.-R. Karger. 2004 Using urls and table layout for web classification tasks. In *Proceedings of the 13th international conference on World Wide Web*. <http://dl.acm.org/citation.cfm?id=988672&picked=prox&CFID=121566981&CFTOKEN=32301290> (accessed Nov. 15 2008).
- Web position*. 2008. <http://en.kioskea.net/web/referencement.php3> (accessed 2 Jan. 2008).
- Withers, Rob, Rob Casson, and Aaron Shrimplin. 2002. Creating Web-based listings of electronic journals with Library Collections. *Acquisitions and Technical Services* 26 (2): 107-112.

Yoshikiyo, Kato, and Sadao Kurohashi. 2008 Classifying information sender of web documents. *Internet Research* 18 (2): 191-203.

Zhang, Jin, and Alexandra Dimitroff. 2005 The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Information Processing & Management* 41 (3): 665-690.

۱۲. پیوست

پیش از پرداختن به آزمون اصلی پژوهش، سیاهه‌ای از معیارهای قضاوت ربط در اختیار آزمودنی‌ها قرار داده شد تا ضمن قضاوت در مورد ربط نتایج بازیابی شده، توجهی نیز به معیارهای ارائه شده داشته باشند تا در صورت بروز ابهام و سؤال، این سیاهه بتواند جوابگوی آنان باشد. این معیارها برگرفته از پایان‌نامه کارشناسی ارشد رضانی‌درج (۱۳۸۴) است که به بررسی قضاوت ربط دانشجویان دانشگاه فردوسی از منابع الکترونیکی پرداخته بود. با توجه به کثرت معیارهای قضاوت ربط، دو مقوله محتوی و پدیدآورنده که کاربران بیشترین توجه را به آنها دارند، انتخاب شدند. معیارهای فرعی هر کدام به‌قرار زیر هستند:

معیارهای ربط دانشجویان

- محتوای مدرک که با استفاده از شاخص‌های زیر بررسی می‌شود:
 - روزآمدی سایت؛
 - عمق و دامنه بررسی موضوع (ارزشمندی محتوی)؛
 - وجود اطلاعاتی که برای شما تازگی دارد؛
 - وجود دستورالعمل‌های کاربردی در مدرک؛
 - وجود تصاویر، جدول‌ها، و یا موارد دیگر که مورد نیاز کاربر بوده است؛
 - سطح محتوای مدرک (اینکه آیا با سطح نیازهای کاربر همخوانی دارد؟)؛ و
 - دسترس‌پذیر بودن محتوای سایت.
- پدیدآورنده که با استفاده از شاخص‌های زیر بررسی می‌شود:
 - شهرت و موقعیت علمی پدیدآورنده؛
 - رشته تخصصی و اعتماد به صلاحیت تخصصی پدیدآورنده؛
 - اعتبار سازمانی که پدیدآورنده بدان وابسته است (یا سایت مربوط به آن است)؛ و
 - معرفی کتاب‌ها و مقالات دیگر از پدیدآورنده.

Comparing the Influence of Title and URL in Information Retrieval Relevance in Search Engines Results between Human Science and Agriculture Science

Parisa Allami*

Master of Library and Information Science

Rahmatollah Fattahi¹

Professor of Library and Information Science,
Ferdowsi University of Mashhad

Iranian Journal of
**Information
Processing &
Management**

Iranian Research Institute
For Science and Technology

ISSN 2251-8223

eISSN 2251-8231

Indexed in LISA, SCOPUS & ISC

Vol.28 | No.1 | pp: 203-224

autumn 2012

Abstract: When the World Wide Web provides suitable methods for producing and publishing information to scientists, the Web has become a mediator to publishing information. This environment has been formed billions of web pages that each of them has a special title, special content, special address and special purpose. Search engines provide a variety of facilities limit search results to raise the possibility of relevance in the retrieval results. One of these facilities is the limitation of the keywords and search terms to the title or URL. It can increase the possibility of results relevance significantly. Search engines claim what are limited to title and URL is most relevant. This research tried to compare the results relevant between results limited in title and URL in agricultural and Humanities areas from their users sights also it notice to Comparison of the presence of keywords in the title and URL between two areas and the relationship between search query numbers and matching keywords in title and their URLs. For this purpose, the number of 30 students in each area whom were in MA process and in doing their thesis was chosen. There was a significant relevant of the results that they limited their information needs to title and URL. There was significantly relevance in URL results in agricultural area, but there was not any significant difference between title and URL results in the humanities. For comparing the number of keywords in title and URL in two areas, 30 keywords in each area were chosen. There was not any significantly difference between the number of keywords in the title and URL of websites in two areas. To show relationship between number of search keyword and the matching of title and URL 45 keywords in each area were chosen. They were divided to three parts (one keyword, two keywords and three keywords). It was determined that if search keyword was less, the amount of matching between title and URL was more and if the matching between title and URL was more, the rate of relevance of website was more. The research concluded that if users with specific names limit their search in the URL, they can get more related results. Also they can get better results by controlling of their search keywords. Specific search engines can also increase the weighting of keywords in the URLs and rank more relevant results to their users.

Keywords: Keywords: title, URL, relevance, information retrieval, search engines

*Corresponding author: parisaallami@gmail.com

1. fattahirahmat@gmail.com