

بررسی تطبیقی جست‌وجو از طریق عنوان و نشانی اینترنتی موتورهای کاوش در میزان ربط نتایج بازیابی اطلاعات توسط کارشناسان استانداری همدان

محمد حسن زاده^۱

سعید غفاری^۲

عاطفه زارعی^۳

حسین کمندی^{۴*}

تاریخ پذیرش: ۱۳۹۲/۰۳/۰۵

تاریخ دریافت: ۱۳۹۱/۱۱/۲۶

چکیده

هدف: پژوهش حاضر به مقایسه میزان ربط نتایج حاصل از محدودسازی عبارت جست‌وجو به روش عنوان و به طریق نشانی اینترنتی توسط کاربران در استانداری همدان و هم‌چنین میزان حضور کلیدواژه‌های مورد جست‌وجو در وب‌سایت‌های گروه‌های مورد پژوهش می‌باشد.

روش پژوهش: جامعه آماری این پژوهش ۹۰ نفر از کارشناسان استانداری همدان هستند. آنان با استفاده از قسمت جست‌وجوی پیشرفته موتور کاوش گوگل، کلیدواژه‌های (یک و چند واژه‌ای) مورد نظر خود را در قسمت نشانی اینترنتی و عنوان صفحات وب جست‌وجو کردند و میزان ربط نتایج بازیابی شده را در چک لیست‌هایی که به همین منظور تهیه شده بود، تعیین نمودند. برای دسته بندی و تحلیل داده‌ها از آمار توصیفی و استنباطی استفاده شد.

یافته‌ها: یافته‌های پژوهش نشان داد که در صورت افزایش تعداد کلیدواژه‌های مورد جست‌وجو میزان هم‌خوانی عنوان و نشانی اینترنتی کاهش می‌یابد و هر چه هم‌خوانی بین عنوان و نشانی اینترنتی بیشتر باشد میزان ربط بالاتر می‌رود.

نتیجه‌گیری: موتورهای کاوش با افزایش وزن‌دهی به کلیدواژه‌های موجود در نشانی اینترنتی می‌توانند نتایج مرتبط‌تری را رتبه‌بندی کرده و در اختیار کاربران قرار دهند و با افزایش هم‌خوانی بین عنوان و نشانی اینترنتی میزان ربط بیشتر می‌شود.

واژه‌های کلیدی: بازیابی اطلاعات، موتورهای کاوش، میزان ربط، نشانی اینترنتی.

۱- هیأت علمی گروه علوم اطلاعات و دانش‌شناسی دانشگاه تربیت مدرس تهران hasanzadeh@modares.ac.ir

۲- عضو هیأت علمی علوم اطلاعات و دانش‌شناسی دانشگاه آزاد اسلامی واحد همدان ghaffari130@yahoo.com

۳- عضو هیأت علمی علوم اطلاعات و دانش‌شناسی دانشگاه آزاد اسلامی واحد همدان Atefehzaree@gmail.com

۴- کارشناس ارشد علوم اطلاعات و دانش‌شناسی دانشگاه آزاد اسلامی واحد همدان Kamandihosseini@yahoo.com

مقدمه

بازیابی اطلاعات از شبکه جهان‌گستر وب و جزئیات فرآیند اطلاع‌یابی در این محیط به دلایل متعددی مورد توجه محققان حوزه بازیابی اطلاعات قرار گرفته است. ویژگی‌های محیط وب از جمله یک‌پارچگی آن و امکان برقراری پیوند میان اطلاعات مختلف گرچه از امتیازات برجسته برای این محیط به شمار می‌آید، اما در برخی مواقع کاربران در هنگام جست‌وجو در وب گم می‌شوند و به خاطر نمی‌آورند که مبدأ و مقصد جست‌وجوی آن‌ها کجا بوده است. این نشانی اینترنتی و عنوان در صفحات وب است که کاربران را یاری می‌کنند کجا هستند و انتظار دارند که چه چیزی را پیدا کنند. عنوان در هنگامی که کاربری سایتی را بازیابی می‌کند، استفاده می‌شود. بنابراین، عنوان مهم‌ترین جزء اطلاعات برای ذخیره و بازیابی در موتورهای کاوش است (Debolt, 2002). کاترل و گان^۱ در تحقیق خود که در نوامبر سال ۲۰۰۷ انجام دادند، بیان کردند که استفاده کنندگان ۲۴٪ از زمان جست‌وجوی خود را صرف نگاه کردن و تحلیل نشانی اینترنتی می‌کنند (Gardne, 2007) و به همین دلیل، با توجه به اهمیت عنوان و نشانی اینترنتی منابع وبی، بیشتر موتورهای کاوش امکان محدود کردن دامنه جست‌وجو به "عنوان" و "نشانی اینترنتی" را فراهم کرده‌اند. در حوزه بازیابی اطلاعات گرایش محسوسی نسبت به مطالعات مرتبط با وب مشاهده می‌شود. اغلب پژوهشگران این حوزه می‌کوشند تأثیر عوامل گوناگون را بر فرایند جست‌وجو در وب را ارزیابی کنند و نتایج این سنجش‌ها را در اختیار طراحان نظام‌های بازیابی اطلاعات مبتنی بر وب قرار دهند. در جست‌وجو و بازیابی اطلاعات در موتورهای جست‌وجو تعداد بسیار زیادی نشانی اینترنتی ارائه می‌شوند. اما همه این نتایج مرتبط نیستند این نتایج معمولاً چنان فراوان هستند که کاربران جز مرور چند صفحه اول از مرور صفحات دیگر منصرف شده و مجبور می‌شود به رتبه بندی موتور کاوش اعتماد کند و همین‌طور موتورهای کاوش از راه‌های مختلفی برای کاهش زمان جست‌وجو و افزایش ربط نتایج استفاده می‌کنند و این موتورها امکانات گوناگونی را برای محدود سازی دامنه جست‌وجو و در نتیجه بالا بردن احتمال ربط نتایج بازیابی در نظر گرفته‌اند. یکی از این امکانات، محدود کردن کلیدواژه‌ها و عبارات جست‌وجو به عنوان و یا نشانی اینترنتی است که تعداد نتایج جست‌وجو را به طور معناداری کاهش می‌دهد و در عین حال احتمال ربط آنان را افزایش می‌دهد (fattahi, 2008). موتورهای کاوش مدعی هستند که آنچه از طریق محدود کردن عنوان و نشانی بازیابی می‌شود، از ربط بیشتری برخوردارند (Greg, 2004). اما آنچه جای بحث در این زمینه دارد، این است که آیا این نتایج از ربط مناسبی از دیدگاه کاربر نیز برخوردار هستند؟ کاربران گروه‌های موضوعی مختلف با جست‌وجو در کدام قسمت آیا جست‌وجو به روش عنوان و یا از طریق نشانی اینترنتی به نتایج بهتری دست می‌یابند؟ میزان اشتراک نتایج بازیابی شده جست‌وجو از طریق عنوان و نشانی اینترنتی چقدر است؟ می‌توان گفت در مورد ربط و بازیابی اطلاعات در موتورهای کاوش پژوهش‌های زیادی انجام گرفته است، اما در مورد جست‌وجو از طریق عنوان و نشانی اینترنتی موتورهای کاوش در میزان ربط نتایج بازیابی اطلاعات که بحث اصلی این پژوهش است کمتر مورد توجه قرار گرفته است.

در ایران توسط برخی پژوهشگران علوم اطلاع‌رسانی، تحقیقاتی مربوط به موتورهای کاوش انجام گرفته است (مانند علیجانی (allijani, 2004) و طیبی عراقی (taiebi aragi, 2005)). برخی از آنان نیز معیارهای دقت و بازیافت را برای مقایسه موتورهای کاوش در نظر گرفته‌اند و در عین حال در مورد عنوان و نشانی اینترنتی پژوهش‌های بسیار اندکی صورت گرفته است. فتاحی (Fattahi, 2008) در تحقیق خود درباره عنوان و نشانی اینترنتی که با هدف بسط جست‌وجو انجام گرفته بود،

¹ Guan & Cutrell

تأثیر کلیدواژه‌های غیر موضوعی (عمومی) در عنوان و نشانی اینترنتی را در بازیابی اطلاعات مرتبط در یک موتور کاوش (گوگل) نشان داد. نیکزاد و جمالی (nikzad, jamali, 2011) در زمینه تأثیر عنوان در میزان مشاهده‌پذیری و استناد به مقالات علمی انجام داده‌اند و هم‌چنین علامی (allami, 2012) در مقاله خود با عنوان "تأثیر عنوان و نشانی اینترنتی در میزان ربط نتایج حاصل از بازیابی اطلاعات در موتورهای جست‌وجو در دو حوزه علوم کشاورزی و علوم انسانی (در کلید واژه‌های زبان انگلیسی)" به این نتیجه رسید که جست‌وجو بر اساس نشانی اینترنتی در حوزه‌های موضوعی علوم کشاورزی نتایج مرتبط‌تری را به دنبال دارد و تفاوت معنی‌داری بین تعداد کلیدواژه‌های موجود در عنوان و نشانی اینترنتی دو حوزه موضوعی وجود ندارد.

کمیته سئو (SEO committee, 2004) در سال ۲۰۰۴ تحقیقی بر روی تأثیر گذاری کلیدواژه‌های موجود در نشانی اینترنتی در دو موتور کاوش یاهو و گوگل انجام داد و مشخص شد که هر دو موتور، حضور کلیدواژه‌ها در نشانی اینترنتی را عاملی در رتبه‌بندی در نظر می‌گیرند. کیم و کنگ (Kang, kim, 2004) نیز در همان سال تحقیقی بر روی بازیابی اطلاعات در رابطه با پرسش‌های کاربر انجام دادند. آن‌ها یک روش رده‌بندی برای پرسش‌های کاربر ارائه دادند. ویدمن و استرومفر (2004)، Weideman & Strümpfer) تأثیر تعداد کلیدواژه‌های به کار رفته توسط کاربران در جست‌وجو و میزان موفقیت بازیابی را بررسی کردند. تحلیل آنان نشان می‌دهد که تفاوت زیادی در میزان موفقیت بین کاربرانی که فقط از یک کلیدواژه استفاده می‌کنند (میزان موفقیت کم) و کسانی که دو یا بیشتر کلید واژه به کار می‌برند، وجود دارد. به طور تخصصی‌تر، به دست آوردن نتیجه قابل قبول در جست‌وجو، زمانی که از دو یا چند کلید واژه استفاده می‌شود، ۵/۵ برابر زمانی است که یک کلیدواژه مورد جست‌وجو قرار می‌گیرد. شین و کارگر (Shin and Karger, 2004) با هدف تعیین الگوریتمی برای رده‌بندی وب مشخص کردند که اگر به جای محتوای متن، به نشانی اینترنتی پیوندها و مکان بصری پیوندها توجه شود، رده‌بندی وب امکان‌پذیر می‌شود. آن‌ها توانستند الگوریتمی برای رده‌بندی وب با توجه به نشانی اینترنتی و محل قرارگیری پیوندها در منابع وبی، پیشنهاد دهند. مشابه پژوهشی که شین و کارگر انجام دادند؛ یوشیکیو، کروهاشی و کنتارو (Yoshikiyo & Kurohashi and kentaro, 2008) در سال ۲۰۰۸ نیز تحقیقی درباره رده‌بندی منابع وب با هدف توسعه یک روش رده‌بندی جدید انجام دادند در این تحقیق برای آزمایش روشی در رده‌بندی کردن مدارک وب از عنوان و نشانی اینترنتی استفاده کردند و نتایج حاصله نشان داد که روش پیشنهادی با استفاده از عوامل یاد شده از بیش از 50٪ دقت برخوردار است. فتاحی، کول فلیچروویلسون (fatahi Concepción S Wilson&Fletcher Cole, 2008) در مقاله‌ای در سال ۲۰۰۸ درباره عنوان و نشانی اینترنتی با هدف بسط جست‌وجو انجام دادند و تأثیر کلیدواژه‌های عمومی در عنوان و نشانی اینترنتی را در بازیابی اطلاعات مرتبط در یک موتور کاوش (گوگل) نشان دادند. آن‌ها در این مقاله خاطر نشان کردند که بسط جست‌وجو در عنوان و در نشانی اینترنتی، با استفاده از واژه‌های عمومی که همراه کلیدواژه‌های موضوعی در عنوان و نشانی اینترنتی منابع و متون موجود در وب ظاهر می‌شوند، می‌تواند موجب افزایش میزان دقت و ربط در نتایج بازیابی در موتورهای کاوش شود (fatahi,2008).

با وجود کثرت تحقیقات در زمینه ذخیره و بازیابی اطلاعات در موتورهای کاوش، تاکنون به مقایسه عنوان و نشانی اینترنتی در بازیابی اطلاعات توجهی نشده است. در این پژوهش سعی شده است که به این مطلب پرداخته شود و هم‌چنین یافته‌های پژوهش حاضر می‌تواند در زمینه جست‌وجوی منابع مرتبط و مفید وبی به کاربران و هم‌چنین در زمینه طراحی بهتر و

مناسب‌تر صفحات وب به طراحان وب سایت یاری رساند، و متخصصان رایانه و طراحان موتورهای کاوش را در تدوین الگوریتم‌های جست‌وجو برای موتورهای کاوش و نیز صفحه رابط جست‌وجو که به شکل پیش فرض در اختیار کاربر قرار داده می‌شود، و در شیوه‌های بهتر نمایه‌سازی موتورهای کاوش به منظور بازیابی مؤثر اطلاعات و نیز هدایت کاربر در بازیابی اطلاعات مرتبط‌تر، کمک نماید.

هدف کلی این پژوهش، بررسی میزان ربط نتایج بازیابی شده حاصل از جست‌وجو در عنوان و نشانی اینترنتی و هم‌چنین عملکرد تعداد کلیدواژه مورد جست‌وجو بر میزان هم‌خوانی واژگانی بین "عنوان" و "نشانی اینترنتی" در کمک به ربط نتایج بازیابی شده اطلاعات وبی توسط کارشناسان گروه‌های مختلف موضوعی مورد پژوهش می‌باشد.

با توجه به اهداف پژوهش، سؤالات زیر پاسخ داده خواهد شد:

- ۱- میزان ربط در نتایج جست‌وجوی مبتنی بر "عنوان" و مبتنی بر "نشانی اینترنتی" منابع وبی، در گروه‌های موضوعی مختلف چه تفاوتی باهم دارد؟
- ۲- میزان ربط در نتایج جست‌وجو بر اساس «عنوان» بین کاربران گروه‌های موضوعی مورد پژوهش چگونه است؟
- ۳- میزان ربط در نتایج جست‌وجو بر اساس «نشانی» بین کاربران گروه‌های موضوعی به چه صورت است؟
- ۴- میزان هم‌خوانی کلیدواژه‌های نشانی اینترنتی و عنوان با افزایش تعداد کلیدواژه‌های مورد جست‌وجو چه تفاوتی دارد؟

روش پژوهش

این پژوهش از نوع کاربردی است و از روش پیمایشی برای بررسی میزان هم‌خوانی و به بررسی وب سایت‌ها از دیدگاه کاربران می‌پردازد. جامعه آماری این پژوهش تعداد ۹۰ نفر از کارشناسان برتر گروه‌های فنی مهندسی، علوم پایه و اقتصاد در استانداری همدان می‌باشند، و با توجه به هدف پژوهش انتخاب این گروه‌ها به دلیل تفاوت در ماهیت آنان می‌باشد. گروه‌های فنی مهندسی، علوم پایه و اقتصاد هر کدام حوزه‌ای تخصصی که کاربران آن دارای نیازهای اطلاعاتی تخصصی می‌باشند، و کارشناسان گروه‌های علوم پایه و اقتصاد بیشتر در معاونت برنامه‌ریزی با کاربرانی با نیازهای اطلاعاتی متفاوت‌تر می‌باشد، و با توجه به ماهیت پژوهش، برای تعمق بیشتر، گروه کوچکی از کاربران به نمایندگی از کل جامعه پژوهش انتخاب شدند. بنابراین تعداد ۹۰ نفر که علاقمند به همکاری با پژوهشگر در این زمینه در گروه‌های موضوعی مذکور بودند انتخاب شدند. دلیل انتخاب این گونه کارشناسان گروه‌ها این است که این گروه‌ها با توجه به ماهیت کار و فعالیت در کارهای پژوهشی و سازمانی و تهیه مطالعات توسعه، آمایش سرزمین و گزارشات آماری و ... در استان همواره به اطلاعات و منابع مرتبط‌تر نیاز دارند. جامعه دیگر این پژوهش وب سایت‌های حاصل از جست‌وجو در موتور کاوش توسط کارشناسان گروه‌های مورد پژوهش می‌باشد که توسط کاربران مورد بررسی و تحلیل قرار گرفتند. برای گردآوری داده‌های مورد نیاز این پژوهش از چک لیست‌های واریسی استفاده شد. هسی - ئی (hesi-ye, 1998) برای تحلیل داده‌ها در این پژوهش از روش آمار توصیفی (فراوانی، میانگین و انحراف استاندارد) و آمار استنباطی (آزمون تک متغیری کالموگورف اسمیرنف، آزمون تعقیبی دانکن، آزمون استقلال خی دو، آزمون t زوجی، آزمون ضرایب همبستگی پیرسون و اسپیرمن، آزمون تحلیل واریانس یک راهه و آزمون کروکسال- وایس) استفاده شد. در تحلیل داده‌ها نمره‌های ۰ تا ۴ در نظر گرفته شد. ابتدا از هر یک از آزمودنی‌ها خواسته شد تا هر کدام پنج کلیدواژه موضوعی (دارای یک تا چند واژه) را بر اساس آخرین نیاز اطلاعاتی خود در نظر

بگیرند و ابتدا در گزینه مربوط به "عنوان" در منوی (صفحه) جست‌وجوی پیشرفته، جست‌وجو را انجام دادند. از میان ارقام بازیابی شده تعداد ده سایت اول را مورد بررسی قرار داده و تعداد موارد بازیابی شده و تعداد موارد مرتبط و میزان ارتباط ده مورد اول توسط آزمودنی‌ها در چک لیست ثبت شد؛ سپس میانگین رتبه‌های تعیین شده برای هر کلیدواژه، توسط پژوهشگر محاسبه و از کاربران خواسته شد جست‌وجوی خود را عیناً با همان کلیدواژه‌های موضوعی در گزینه مربوط به "نشانی اینترنتی" تکرار کنند. تعداد موارد بازیابی شده و تعداد موارد مرتبط و میزان ارتباط مدارک مانند مرحله قبل ثبت شد. برای هر موضوع ده نتیجه برای جست‌وجو در "عنوان" و ده نتیجه برای جست‌وجوی در "نشانی اینترنتی" مورد بررسی قرار گرفت که در مجموع بیست صفحه وب برای هر موضوع مورد جست‌وجو واقع شد و هر کاربر در حدود صد صفحه وب را مورد بررسی قرار داد و هم‌چنین، این پژوهش به بررسی میزان هم‌خوانی کلیدواژه‌های نشانی اینترنتی منابع وبی با کلیدواژه‌های عنوان آنها بوده است. برای این منظور ابتدا کلیدواژه‌های مورد جست‌وجو در گروه‌های مورد پژوهش به دو قسمت یک واژه‌ای و چند واژه‌ای تقسیم شد، و سپس از هر کدام از گروه‌ها تعداد سی کلیدواژه انتخاب شدند. بنابراین در دو قسمت یک واژه‌ای و چند واژه‌ای ۶۰ کلیدواژه مورد بررسی قرار گرفت و میزان هم‌خوانی کلیدواژه‌های موجود در نشانی اینترنتی منابع وبی با کلیدواژه‌های عنوان آنها به طور جداگانه ثبت شد.

یافته‌های پژوهش

پرسش اول: میزان ربط نتایج جست‌وجوی مبتنی بر "عنوان" و مبتنی بر "نشانی اینترنتی" منابع وبی، در گروه‌های موضوعی مختلف چه تفاوتی باهم دارد؟

پیش از تعیین میزان ربط، ابتدا با استفاده از آزمون تک متغیری کالموگورف اسمیرنوف برای تعیین میزان نرمال بودن داده‌های عنوان و نشانی اینترنتی در گروه‌های موضوعی استفاده شد. با توجه به مقدار p به دست آمده از آزمون فوق در تمام گروه‌ها که از مقدار $0/05$ بزرگ‌تر است نرمال بودن توزیع داده‌ها را نشان می‌دهد؛ به همین دلیل، برای تعیین میزان ربط جست‌وجو در عنوان و نشانی اینترنتی از آزمون t استفاده شد.

جدول ۱. میزان ربط نتایج حاصل از جست‌وجو به طریق عنوان و نشانی اینترنتی با استفاده از آزمون t زوجی

گروه‌ها	میانگین اختلاف	انحراف معیار اختلاف	مقدار آماره	p-value
اقتصاد	۰/۴۵۵	۰/۱۳۸	۱۰/۸۰	۰/۰۰۰
علوم پایه	۰/۲۴۲	۰/۲۸۰	۲/۸۶۵	۰/۰۱۷
فنی مهندسی	۰/۹۰۷	۰/۲۱۴	۱۴/۰۶	۰/۰۰۰

در جدول (۱) نتایج به دست آمده از آزمون t زوجی در تمام گروه‌ها از مقدار $0/05$ کوچک‌تر است، نشان دهنده آن است که در گروه‌های فوق بین عنوان و نشانی اینترنتی در نتایج بازیابی اطلاعات مرتبط تفاوت معنی‌داری وجود دارد. پرسش دوم: آیا میزان ربط نتایج جست‌وجو بر اساس «عنوان» بین کاربران گروه‌های موضوعی مورد پژوهش به صورت معنی‌داری متفاوت است؟

میزان ربط نتایج حاصل از جست‌وجو بر اساس عنوان، با استفاده از داده‌های به دست آمده از سیاهه واریسی شماره یک تعیین شد. سپس برای بررسی معنی تفاوت ربط در بین گروه‌های مختلف از آزمون تحلیل واریانس استفاده گردید.

جدول ۲. جدول آنوا برای تفاوت میزان ربط نتایج حاصل از جست‌وجو در گروه‌ها به روش "عنوان"

منبع تغییر	مجموع مربعات خطا	درجه آزادی	میانگین مربعات خطا	مقدار آماره F	مقدار P-value
بین گروه‌ها	۱/۵۸	۲	۰/۹۲۴	۳۹/۴۳۶	۰/۰۰۰
خطا	۰/۷۰۴	۳۰	۰/۰۲۳		
کل	۲/۵۵۴	۳۲			

در جدول (۲) مقدار P در آزمون تحلیل واریانس یک راهه برابر با ۰/۰۰۰ می‌باشد و با توجه به اینکه این مقدار از ۰/۰۵ کوچک‌تر است، نشان دهنده آن است که در میان گروه‌های مورد پژوهش از دید کاربران تفاوت معناداری به لحاظ نتایج حاصل از جست‌وجو در عنوان مشاهده می‌شود.

پرسش سوّم: آیا میزان ربط نتایج جست‌وجو بر اساس «نشانی» بین کاربران گروه‌های موضوعی مورد پژوهش به صورت معنی‌داری متفاوت است؟

میزان ربط نتایج حاصل از جست‌وجو بر اساس نشانی اینترنتی، با استفاده از داده‌های به دست آمده از سیاهه واریسی شماره یک تعیین شد. سپس برای بررسی معنی تفاوت ربط در بین گروه‌های مختلف از آزمون تحلیل واریانس استفاده گردید.

جدول ۳. جدول آنوا برای تفاوت میزان ربط جست‌وجو در گروه‌ها به روش "نشانی اینترنتی"

منبع تغییر	مجموع مربعات خطا	درجه آزادی	میانگین مربعات خطا	مقدار آماره F	P-value
بین گروه‌ها	۴/۶۴۴	۲	۲/۳۲۲	۶۶/۰۴۹	۰/۰۰۰
خطا	۱/۰۵۵	۳۰	۰/۰۳۵		
کل	۵/۶۹۹	۳۲			

در جدول (۳) مقدار P در آزمون تحلیل واریانس یک راهه برابر با ۰/۰۰۰ می‌باشد و با توجه به اینکه این مقدار از ۰/۰۵ کوچک‌تر است، نشان دهنده آن است که در میان گروه‌های مورد پژوهش از دید کاربران تفاوت معناداری به لحاظ نتایج حاصل از جست‌وجو در نشانی اینترنتی مشاهده می‌شود و سپس با استفاده از آزمون تعقیبی دانکن الگوی تفاوت مشخص شد (جدول ۵).

جدول ۴. آزمون تعقیبی دانکن برای نتایج حاصل از جست‌وجو در "نشانی اینترنتی"

گروه‌ها	N	مقدار آلفا برای زیر مجموعه‌ها = ۰/۰۵	
		۱	۲
اقتصاد	۱۱	۲/۹۱۳۶	
علوم پایه	۱۱		۳/۶۱۷۳
فنی مهندسی	۱۱		۳/۷۷۳۷
سطح معنی‌داری		۰/۰۰۰	۰/۰۵۴

در جدول (۴) با توجه به داده‌های به دست آمده برای نتایج حاصل از جست‌وجو در "نشانی اینترنتی" در آزمون تعقیبی دانکن که سطح معنی داری در دو گروه فنی و مهندسی و علوم پایه برابر با ۰/۰۵۴ می‌باشد و از مقدار زیر مجموعه آلفا ۰/۰۵ بزرگ‌تر است، نشان دهنده آن است که تفاوت معنی‌داری بین گروه "اقتصاد" با سایر گروه‌ها در نتایج حاصل از جست‌وجو در "نشانی اینترنتی" وجود دارد و بین دو گروه فنی مهندسی و علوم پایه تفاوت از لحاظ آماری معنی‌دار نیست. پرسش چهارم: میزان هم‌خوانی کلیدواژه‌های نشانی اینترنتی و عنوان با افزایش تعداد کلیدواژه‌های مورد جست‌وجو چه تفاوتی دارد؟

ابتدا ۳۰ کلیدواژه یک واژه‌های و سی کلیدواژه چند واژه‌های از گروه‌های مورد پژوهش انتخاب و در مورد هر کلیدواژه ده صفحه وب مورد بررسی قرار گرفت. سپس برای بررسی تفاوت میزان هم‌خوانی کلید واژه‌ها با استفاده از آزمون خی دو در گروه‌های مختلف استفاده گردید.

جدول ۵. آزمون استقلال خی دو میزان هم‌خوانی و عدم هم‌خوانی کلید واژه‌های یک و چند واژه‌ای

سطح معنی داری دقیق یک طرفه	سطح معنی داری دقیق دو طرفه	سطح معنی داری متقارن	درجه آزادی	مقدار	
		0/000	۱	۱۴/۵۳۸	کای دو دقیق
		0/000	۱	۱۳/۸۰۲	تصحیح پیوستگی
		0/000	۱	۱۴/۷۰۴	نسبت درست نمایی
			۱		آزمون دقیق فشر
۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۱	۱۴/۵۱۳	ارتباط خطی
			۱	۶۰۰	تعداد

در جدول (۵) در آزمون استقلال خی دو با توجه به نتایج به دست آمده از در کلیه آماره‌ها سطح معنی‌داری برابر ۰/۰۰۰ است که از سطح خطای ما یعنی ۰/۰۵ کمتر است؛ یعنی هر کدام از آماره‌های کای دو دقیق، تصحیح پیوستگی و ... را که در نظر بگیریم، سطح معنی‌داری برابر با ۰/۰۰۰ است و، نشان دهنده آن است که فرض مستقل بودن نتایج حاصل از "یک واژه‌ای یا چند واژه‌ای بودن کلید واژه‌ها" از "هم‌خوانی یا عدم هم‌خوانی" رد می‌شود.

بحث و نتیجه‌گیری

نتایج حاصل از این پژوهش حاکی از این است که جست‌وجو بر اساس نشانی اینترنتی در گروه‌های موضوعی فنی مهندسی و جست‌وجو در عنوان در گروه‌های موضوعی اقتصاد نتایج مرتبط‌تری را به دنبال دارد و تفاوتی بین جست‌وجوی مبتنی بر عنوان و نشانی اینترنتی در گروه علوم پایه مشاهده نمی‌شود پژوهش حاضر در مورد افزایش میزان دقت و ربط در نتایج بازیابی در موتورهای کاوش در عنوان و نشانی اینترنتی با تحقیق فتاحی (fatahi,2008) که با هدف بسط جست‌وجو در عنوان و نشانی اینترنتی انجام داده است هم‌خوانی دارد. او در مقاله خود خاطر نشان کرد که بسط جست‌وجو در عنوان و در نشانی اینترنتی، با استفاده از واژه‌های عمومی که همراه کلیدواژه‌های موضوعی در عنوان و نشانی اینترنتی منابع و متون موجود

در وب ظاهر می‌شوند، می‌تواند موجب افزایش میزان دقت و ربط در نتایج بازیابی در موتورهای کاوش شود فتاحی (fatahi,2008) و همچنین این تحقیق از یک طرف با پژوهش علامی (allami,2012) که در مورد تأثیر عنوان و نشانی اینترنتی در میزان ربط نتایج حاصل از بازیابی اطلاعات در موتورهای جست‌وجو در دو حوزه علوم کشاورزی و علوم انسانی (در کلید واژه‌های زبان انگلیسی) انجام داده است هم‌سو است. او به این نتیجه رسید که، جست‌وجو بر اساس نشانی اینترنتی در حوزه‌های موضوعی علوم کشاورزی نتایج مرتبط‌تری را به دنبال دارد. ولیکن از طرف دیگر در مورد معنی‌داری بین تعداد کلیدواژه‌های موجود در عنوان و نشانی اینترنتی دو حوزه موضوعی که نتیجه‌گیری می‌کند که بین تعداد کلیدواژه‌های موجود در عنوان و نشانی اینترنتی دو حوزه موضوعی تفاوت وجود ندارد، با پژوهش حاضر متفاوت است. در موارد هم‌خوانی نیز شاهد میزان ربط بالایی هستیم بنابراین می‌توان نتیجه گرفت که، هم‌خوانی کلیدواژه‌های موضوعی در عنوان و نشانی اینترنتی یک صفحه وب بیانگر ارتباط بیشتر آن صفحه با موضوع مورد نظر است، و باعث بهبود ربط می‌شود. تفاوت معنی‌داری در بین تعداد کلیدواژه‌های موجود در عنوان و نشانی اینترنتی وب سایت‌های بین گروه‌ها وجود دارد. تفاوت میزان ربط نتایج حاصل از جست‌وجو در عنوان و نشانی اینترنتی بین گروه‌ها موضوعی ریشه در ماهیت واژگان بین گروه‌ها دارد و همچنین در صورت استفاده از کلید واژه‌های چند واژه‌های مورد جست‌وجو میزان هم‌خوانی عنوان و نشانی اینترنتی کاهش می‌یابد، و کنترل تعداد کلیدواژه‌های مورد جست‌وجو می‌تواند در بهبود ربط اثرگذار باشد.

در مجموع یافته‌های این پژوهش نشان داد که میان نتایج جست‌وجو‌هایی که کلیدواژه مورد جست‌وجو در عنوان و نشانی اینترنتی هم‌خوانی بیشتری وجود دارد، میزان ربط بالاتر است. هنگامی که واژه‌های مرتبط در هر دو منطقه عنوان و نشانی حضور داشته باشند، سایت مورد نظر از رتبه بالایی در بازیابی برخوردار می‌شود، و نیز مشخص شد که وب سایت‌هایی که دارای کلیدواژه‌های تک واژه‌ای هستند از هم‌خوانی بالایی برخوردارند. بنابراین، طراحان صفحات وب را می‌توان تشویق نمود تا برای نام‌گذاری عنوان صفحات و نیز برای تدوین نشانی اینترنتی از کلیدواژه‌های کوتاه و ساده و معنادار و به هم‌خوانی واژگانی بین عنوان و نشانی اینترنتی سایت، که باعث افزایش رتبه سایت در موتور کاوش گوگل و افزایش ربط را مشمول شود استفاده نمایند.

منابع

- Allami, Parisa And Fattahi, R. (2012). Comparing the Influence of Title and URL in Information Retrieval Relevance in Search Engines Results between Human Science and Agriculture Science. *Iranian Research Institute For Science and Technology* , 28 (1), 203-224 autumn 2012. [in persian]
- Concepción S. Wilson& Fletcher Cole *Information Processing and Management* 44 (2008) 1503–1516 www.elsevier.com/locate/infoproman. [in persian]
- Fattahi, R. (2008). an alternative approach to query expansion in search engines through non-topical. *Studies in Education and Psychology* , (1), 31-52. [in persian]
- Fattahi, R. (2008). *an alternative approach to natural language query expansion in search engines: text analysis of non-topical terms in web documents*.
- Greg, R. Review of Hot Bot(Inktomi). (2004). <http://www.searchengineshowdown.com/features/hotbot/> (accessed 9 Feb 2008).
- Jamali , H.R. and Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations ,*Scientometrics*, 88 (2), 653-661 [in persian]
- committee.(2008) Importance of Keywords in URLs and Search Engine
- Kang, In-Ho , and Gil Chang K. (2004).Integration of multiple evidences based on a query type for web search. *Information Processing & Management* , 40(3), (may 2004),459-478.

- Kan, min-yen; Shin, L- K, and D- R Karger.(2004). Using urls and table layout for web classification tasks . Proceedings of the 13th international conference on World Wide Web 2004 .
- Lawrence, Steve ...[et al](2001). Persistence of Web References in Scientific Research. *Computer magazine*, 34(2),26-31.
- Ranking. *search engine journal*, Retrieved march 21,2008, from <http://www.searchenginejournal.com/importance-of-keywords-in-urls-and-search-engine-ranking/1019/>.
- Web position . (2008). Retrieved January 2, 2008 from <http://en.kioskea.net/web/referencement.php3>.
- Yoshikiyo, Kato, and Kurohashi Sadao. (2008). Classifying information sender of web documents. *Internet Research* ,18 (2), 191-203.

Archive of SID