

بررسی عناصر ابر داده‌ای موجود در صفحات وب حاصل از جستجو در موتورهای جستجوی عمومی

سیما شفیعی علویجه*
کارشناس ارشد کتابداری و اطلاع‌رسانی
دانشگاه الزهرا

امیر غائبی
دکترای تکنولوژی اطلاعات
عضو هیئت علمی و استادیار دانشگاه الزهرا

سعید رضایی شریف‌آبادی
دکترای کتابداری و اطلاع‌رسانی
عضو هیئت علمی و دانشیار دانشگاه الزهرا

اطلاعات
علوم و فناوری

دریافت: ۱۳۸۸/۰۴/۲۷ پذیرش: ۱۳۸۸/۰۷/۱۴ مقاله برای اصلاح به مدت یک ماه و ۱۳ روز نزد پدیدآوران بوده است.

چکیده: در پژوهش حاضر ۹۰ صفحه وب از طریق جستجو در موتورهای جستجوی عمومی (گوگل، یاهو، و ام‌اس‌ان) به عنوان نمونه انتخاب شدند و میزان حضور عناصر ابر داده‌ای (هسته دابلین و ابرنشان‌های زبان نشانه‌گذاری فرامتن) در این صفحات و وجود رابطه معنادار بین حضور عناصر ابر داده‌ای و نوع موتور جستجو مورد بررسی قرار گرفت.

نتایج، نشان‌دهنده حضور بسیار کم عناصر ابر داده‌ای هسته دابلین و ابرنشان‌های زبان نشانه‌گذاری فرامتن در صفحات بازیابی شده می‌باشد، که بیانگر کاربرد بسیار کم عناصر ابر داده‌ای در صفحات وب است. به علاوه، یافته‌های این تحقیق نشان می‌دهد رابطه معناداری بین نوع موتور جستجو و حضور عناصر ابر داده‌ای وجود ندارد، و این بدان معنا است که موتورهای جستجو از نظر لحاظ کردن ابر داده‌ها در بازیابی منابع وب، تفاوت معناداری با یکدیگر ندارند.

کلیدواژه‌ها: عناصر ابر داده‌ای؛ هسته دابلین؛ ابرنشان‌های زبان نشانه‌گذاری فرامتن؛ موتورهای جستجوی عمومی؛ گوگل؛ یاهو؛ ام‌اس‌ان؛ صفحات وب

* پدیدآور رابط simashafii@yahoo.com

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا (چاپی) ۱۷۳۵-۵۲۰۶
شاپا (الکترونیکی) ۲۰۰۸-۵۵۸۲
نمایه در LISA & SCOPUS
<http://jst.irandoc.ac.ir>
دوره ۲۵ | شماره ۱ | صص ۷۱-۸۹
پاییز ۱۳۸۸

نوع مقاله: علمی پژوهشی

۱. مقدمه

در اوایل دهه ۱۹۹۰، شبکه وب توسط «تیم برنرز-لی»^۱ راه اندازی شد. وب مشتمل بر تعداد زیادی از منابع اطلاعاتی است و از همان ابتدا حجم و تنوع منابع موجود در آن به سرعت رشد یافت. از این رو، یکی از مسائلی که ذهن محققان را به خود مشغول کرده، سازمان‌دهی و بازیابی اطلاعات وب است.

منابع در محیط وب دارای ویژگی‌هایی هستند که آن‌ها را از منابع سنتی متفاوت می‌کند؛ از این رو کارکرد استانداردهای پیشین (مانند قواعد فهرست‌نویسی انگلوامریکن) که با اشیای فیزیکی سروکار دارند در این محیط کمرنگ شده است. به علاوه، محققان دریافتند به دلیل اندازه، وسعت و طبیعت پویای وب، روش‌های سنتی ذخیره و بازیابی اطلاعات، کارآمد نیستند. یکی از راه‌حل‌های پیشنهادشده برای این مشکل، استفاده از ابر داده برای توصیف منابع وب است. ابر داده‌ها مجموعه‌هایی از عناصر داده‌ای هستند که یا به صورت مستقیم در عنصر سرآیند^۲ یک مدرک، یا در داده‌پایگاه جداگانه‌ای جای داده می‌شوند.

عمومی‌ترین طرح‌های ابر داده‌ای که برای سازمان‌دهی و توصیف منابع وب مورد استفاده قرار می‌گیرند، طرح ابر داده‌ای هسته دابلین و ابر نشانه‌های زبان نشانه‌گذاری فرامتن می‌باشند. اگر منابع وب به صورت دقیق توسط ابر داده توصیف شوند و موتورهای جستجو نیز از فیلدهای ابر داده‌ای برای جستجو و بازیابی منابع استفاده کنند، مجموعه‌های بازیابی شده به سطوح قابل بررسی کاهش می‌یابند و دقیق‌تر رتبه‌بندی می‌شوند.

کاربران اغلب برای جستجوی منابع اطلاعاتی از موتورهای جستجو بهره می‌گیرند. با توجه به حجم و تنوع منابع وب، تعداد نتایج بازیابی شده توسط موتورهای جستجو اغلب بسیار زیاد است و کاربران تنها چند صفحه اول از نتایج بازیابی شده را مورد بررسی قرار می‌دهند (Noruzi 2007; Zhang and Dimitroff 2005a; Bar-Ilan 2005; Jansen, Spink, and Saracevic 2000)؛ بنابراین انتظار می‌رود نتایجی که دارای ربط بیشتری با موضوع مورد جستجو هستند در رتبه‌های اول سیاهه نتایج جستجو قرار گیرند.

^۱ Tim Burners-Lee

^۲ HEAD

با توجه به تحقیقات انجام شده، کاربرد عناصر ابر داده‌ای در منابع وب موجب بهبود بازیابی و رتبه آن‌ها در میان مدارک بازیابی شده می‌گردد. توجه به حضور عناصر ابر داده‌ای در الگوریتم‌های بازیابی موتورهای جستجو موجب بهبود در کشف منبع و دستیابی به منابع اطلاعاتی می‌گردد (Zhang and Dimitroff 2004; Mohamed 2006)؛ شریف (۱۳۸۵).

ضرورت انجام این پژوهش از آن جهت است که تعیین شود در مدارک بازیابی شده که دارای رتبه بهتری در سیاهه نتایج جستجو هستند؛ کدامیک از عناصر ابر داده‌ای بیش تر به کار رفته‌اند. با شناسایی این عناصر می‌توان به ارائه پیشنهاد در جهت به کارگیری این عناصر ابر داده‌ای در هنگام طراحی صفحات وب، خصوصاً صفحات وب فارسی، برای بالا بردن امکان بازیابی و بهبود رتبه آن‌ها در سیاهه نتایج جستجو پرداخت. هدف از انجام این پژوهش، بررسی میزان حضور عناصر ابر داده‌ای (عناصر ابر داده‌ای هسته دابلین و ابر نشانه‌های زبان نشانه‌گذاری فرامتن) در صفحات حاصل از جستجو در موتورهای جستجوی عمومی (گوگل، یاهو، و ام‌اس‌ان) است.

۲. بیان مسئله

همان‌طور که گفته شد، افزایش منابع اطلاعاتی منجر به تلاش‌های زیادی در جهت سازمان‌دهی و تسهیل در یافتن اطلاعات شده است. بسیاری از این تلاش‌ها بر استفاده از ابر داده برای سازماندهی تأکید دارند. یک رکورد ابر داده‌ای مشتمل بر مجموعه ویژگی‌ها یا عناصری است که برای توصیف منابع وب ضروری‌اند. این عناصر شامل اطلاعاتی درباره شکل منبع و محتوای آن (مثل «عنوان»، «نویسنده»، «تاریخ انتشار»، یا جزئیات دیگری) هستند که می‌توانند به جستجوگران و تهیه‌کنندگان اطلاعات کمک کنند. در صورتی که موتورهای جستجو در جهت کاربرد ابر داده طراحی شده باشند، جستجو و یافتن منابع اطلاعاتی از طریق فیلدهای ابر داده‌ای مثل «عنوان»، «نویسنده»، و ... تسهیل می‌شود (Vineyard 2001).

باین که طرح‌های ابر داده‌ای برای تسهیل در یافتن اطلاعات در محیط‌های مختلف ایجاد شده‌اند، اما به نظر نمی‌رسد کاربرد ابر داده برای کشف منابع اینترنت، عملی شده باشد، و تعداد کمی از موتورهای جستجو در الگوریتم‌های جستجوی خود به ابر داده توجه

کرده‌اند. هر یک از موتورهای جستجو از الگوریتم خاصی برای بازیابی منابع اطلاعاتی استفاده می‌کند (Henshaw and Valauskas 2001). ساختارهای ابر داده‌ای^۱، محتوای صفحات^۲، و وضعیت ارجاعات درون‌متنی^۳، از جمله عوامل اثرگذار بر بازیابی و رتبه‌بندی صفحات وب هستند (ژانگ و جاسترام ۲۰۰۵، نقل در شریف ۱۳۸۵). بنابراین مسئله‌ای که در این پژوهش مطرح می‌شود این است که آیا موتورهای جستجو در الگوریتم‌های بازیابی خود به ابر داده توجه می‌کنند؟ به عبارت دیگر، با توجه به این که نتایج اولیه حاصل از جستجو در موتورهای جستجو، معمولاً باربندترین نتایج هستند، آیا در نتایج اولیه حاصل از جستجو در موتورهای جستجو، عناصر ابر داده‌ای به کار رفته‌اند؟ و دیگر این که در صورت کاربرد، کدامیک از عناصر ابر داده‌ای بیشترین کاربرد را در نتایج اولیه حاصل از جستجو داشته‌اند؟

در این پژوهش منظور از عناصر ابر داده‌ای، مجموعه عناصر ابر داده‌ای هسته دابلین و مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن می‌باشد.

عناصر ابر داده‌ای هسته دابلین مشتمل بر ۱۵ عنصر ابر داده‌ای می‌باشند که برای توصیف منابع وب به کار می‌روند. این عناصر عبارت‌اند از: «عنوان»، «پدیدآور»، «موضوع»، «توصیف»، «ناشر»، «همکار»، «تاریخ»، «نوع»، «قالب»، «شناسگر»، «منبع»، «زبان»، «رابطه»، «پوشش» و «حقوق».

از مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن، ۱۲ عنصر زیر در این پژوهش مورد بررسی قرار گرفتند: «عنوان»، «پدیدآور»، «کلیدواژه»، «تاریخ»، «نوع منبع»، «زبان»، «حق مؤلف»، «تاریخ انقضا»، «توصیف»، «تولیدکننده»، «روبو»، «تجدیدنظر کردن». لازم به ذکر است که «عنوان» در این پژوهش علیرغم نشانه بودن، به منظور همسان‌سازی با عناصر ابر داده‌ای هسته دابلین، در حکم یک ابر نشانه انگاشته شده است.

منظور از موتورهای جستجو در پژوهش حاضر، سه موتور جستجوی عمومی گوگل، یاهو، و ام‌اس‌ان می‌باشد.

پرسش‌های پژوهش حاضر از این قرارند:

¹ metadata structure

² webpage content

³ hyperlink cited status

۱. آیا در صفحات باز یابی شده از موتورهای جستجوی عمومی، عناصر ابر داده‌ای به کار رفته‌اند؟

۱-۱. میزان به کارگیری عناصر ابر داده‌ای هسته دابلین در صفحات وب چقدر است؟
۲-۱. میزان به کارگیری ابر نشانه‌های زبان نشانه گذاری فرامتن در صفحات وب چقدر است؟

فرضیات این پژوهش عبارت‌اند از:

۲. بین نوع موتور جستجو و وجود عناصر ابر داده‌ای از مجموعه عناصر ابر داده‌ای هسته دابلین رابطه معناداری وجود دارد.

۳. بین نوع موتور جستجو و وجود عناصر ابر داده‌ای از مجموعه ابر نشانه‌های زبان نشانه گذاری فرامتن رابطه معناداری وجود دارد.

۳. مروری بر ادبیات و پیشینه پژوهش

پژوهش‌هایی که در حوزه ابر داده انجام شده‌اند، به دو نوع پژوهش قابل تقسیم‌اند. نوع اول پژوهش‌هایی هستند که در آن‌ها، محقق با قرار دادن ابر داده برای صفحات وب، اثربخشی کاربرد آن را بر باز یابی و رتبه صفحات مورد بررسی قرار می‌دهد. از جمله این پژوهش‌ها می‌توان به مطالعات انجام شده توسط افرادی همچون «هنشا» و «والاسکاس»، «ژانگ» و «دیمیتروف»، «صفری»، «محمد» و «شریف» اشاره کرد (Henshaw and Valauskas 2001; Zhang and Dimitroff 2004; Zhang and Dimitroff 2005a; Zhang and Dimitroff 2005b; Mohamed 2006; صفری ۱۳۸۲؛ شریف ۱۳۸۵).

نتایج به دست آمده از هر یک از این پژوهش‌ها تا حدودی با یکدیگر متفاوت‌اند: نتایج حاصل از تحقیق «صفری» نشان داد که هیچ‌گونه تفاوت معناداری میان دو گروه گواه و آزمودنی بعد از افزودن عناصر ابر داده‌ای به صفحات مورد بررسی، در بهبود باز یابی صفحات وب وجود ندارد (صفری ۱۳۸۲). این در حالی است که نتایج تحقیقاتی که توسط «هنشا» و «والاسکاس»، و «ژانگ» و «دیمیتروف»، «محمد»، و «شریف» انجام شد حاکی از آن است که رتبه صفحات مورد بررسی بعد از افزودن عناصر ابر داده‌ای، تا حدودی بهبود یافته است (Henshaw and Valauskas 2001; Zhang and Dimitroff 2004;

(Zhang and Dimitroff 2005a; Zhang and Dimitroff 2005b; Mohamed 2006) شریف
۱۳۸۵.

نوع دوم پژوهش‌های انجام‌شده در حوزه ابر داده، پژوهش‌هایی هستند که در آن‌ها میزان استفاده از عناصر ابر داده‌ای مورد سنجش قرار گرفته‌اند. پژوهش درباره ابر داده و کاربرد آن نه تنها برای ناشران و نویسندگان صفحات وب که می‌خواهند بازیابی وبسایت‌هایشان افزایش یابد دارای اهمیت است، بلکه همچنین برای طراحان موتورهای جستجو نیز اهمیت دارد؛ به طوری که آن‌ها می‌توانند الگوریتم‌های بازیابی را به گونه‌ای تدوین و اصلاح کنند که تا حد امکان، وبسایت‌های مرتبط، بازیابی و رتبه‌بندی شوند. در ادامه مروری کوتاه بر پژوهش‌هایی که ارتباط بیش‌تری با پژوهش حاضر دارند خواهیم داشت.

۳-۱. پیشینه پژوهش در ایران

«پیشوا» و «مجیدی» در پژوهشی به بررسی میزان استفاده از دو ابرنشانه - یعنی «کلیدواژه» و «توصیف» - در مجله‌های دسترسی آزاد وب پرداختند. یافته‌های پژوهش نشان دادند که میزان به‌کارگیری «کلیدواژه‌ها» و «توصیف» در نمونه آماری، به ترتیب ۳۳/۱ درصد و ۲۹/۹ درصد می‌باشد. همچنین از بین رده‌های موضوعی مختلف، «مجله‌های عمومی» با ۵۵/۶ درصد بیش‌ترین میزان استفاده، و مجله‌های رده موضوعی «ریاضیات و آمار» با ۱۱/۱ درصد کم‌ترین میزان استفاده از ابرنشانه «کلیدواژه» را داشته‌اند. همچنین مجله‌های عمومی و مجله‌های رده «شیمی» به ترتیب با ۵۵/۶ درصد و ۱۵/۴ درصد، بیش‌ترین و کم‌ترین میزان استفاده از ابرنشانه «توصیف» را داشتند. بر اساس یافته‌های این پژوهش و در مقایسه با یافته‌های دیگر پژوهش‌های مشابه، میزان استفاده از این ابرنشانه‌ها رشد چندانی نداشته است (پیشوا و مجیدی ۱۳۸۶).

«سعادت» مطالعه‌ای بر روی صفحات وب کتابخانه‌های ملی جهان از نظر استفاده از عناصر ابر داده‌ای هسته دابلین انجام داد. نتایج نشان داد در میان ۶۹ وبسایت کتابخانه‌ای متعلق به ۶۱ کشور جهان تنها ۲۰/۳ درصد از آن‌ها از عناصر ابر داده‌ای استفاده کرده‌اند (سعادت ۱۳۸۵).

«علیمحمدی» در پژوهشی به اندازه‌گیری حضور ابر نشانه‌های «توصیف» و کلیدواژه در وب‌سایت‌های ایرانی بر اساس شاخص‌های بین‌المللی پرداخت. برای انجام پژوهش ۳۴۶ وب‌سایت ایرانی از میان ۳۳۴۲ وب‌سایت از ایران‌هو^۱ انتخاب شدند. یافته‌های پژوهش نشان دادند که ۳۱/۵ درصد و ۲۴/۶ درصد از وب‌سایت‌های ایرانی به ترتیب دارای ابر نشانه‌های «کلیدواژه» و «توصیف» هستند. نتایج نشان‌دهنده استفاده کم‌تر وب‌سایت‌های ایرانی نسبت به وب‌سایت‌های غیر ایرانی از ابر نشانه‌ها می‌باشد (علیمحمدی ۱۳۸۲).

۳-۲. پیشینه پژوهش در خارج

«کریون» استفاده از ابر نشانه «توصیف» را در صفحات وب زبان‌های مختلف مورد بررسی قرار داد. وی در پژوهش خود ۲۰ زبان اصلی را با استفاده از لیستی که توسط گوگل ارائه شده انتخاب کرد. با استفاده از موتور جستجوی گوگل، برای هر زبان و کشوری که همان زبان، زبان رسمی آن است جستجویی انجام داد و صفحات دارای رتبه بالا را به عنوان جامعه مورد مطالعه انتخاب کرد. با بررسی صفحات، تفاوت‌هایی از لحاظ میزان کاربرد ابر نشانه «توصیف» با توجه به زبان مشاهده شد، به طوری که صفحاتی که به زبان‌های اروپای غربی بودند بیش‌تر از ابر نشانه «توصیف» استفاده کرده بودند؛ در حالی که صفحات چینی زبان، این عنصر را کم‌تر به کار برده بودند (Craven 2004).

«وینیارد» پنج کلیدواژه را در سه موتور جستجو - هات‌بات^۲، آلتاویستا، و گوگل - مورد جستجو قرار داد و ۲۰ نتیجه اول از هر جستجو را بررسی کرد. از ۲۹۹ صفحه وب بررسی شده، ۲۱۲ صفحه (۷۰/۹۰٪) دارای حداقل یک ابر نشانه بودند و تنها ۷ صفحه (۲/۳۴٪) از عناصر ابر داده‌ای هسته دابلین استفاده کرده بودند. محقق پیشنهاد می‌کند که طراحان صفحات وب برای توصیف محتوای صفحات خود، از عناصر ابر داده‌ای استفاده کنند. به علاوه، طراحان موتورهای جستجو نیز امکان جستجو از طریق فیلدهای ابر داده‌ای را برای کاربران فراهم نمایند (Vineyard 2001).

«کریون» در پژوهش دیگر، ۱۹۳۷ صفحه وب را از نظر میزان استفاده از ابر نشانه «توصیف» مورد بررسی قرار داد. این صفحات به سه دسته تقسیم شدند: دسته اول،

¹ Iranhoo

² Hotbot

صفحات بازیابی شده از یاهو بودند؛ دسته دوم صفحاتی بودند که پیوندهای آنها روی صفحات گروه اول قرار داشت؛ دسته سوم شامل صفحاتی بودند که پیوندهای آنها روی صفحات گروه دوم قرار داشت. این صفحات از نظر استفاده یا عدم استفاده از ابرنشانه «توصیف» در سال‌های ۲۰۰۰ و ۲۰۰۱ مورد بررسی قرار گرفتند. نتایج این پژوهش نشان داد ۳۶/۵ درصد از صفحات در سال ۲۰۰۰، و ۳۵/۵ درصد از صفحات در سال ۲۰۰۱ از این ابرنشانه استفاده کرده‌اند (Craven 2001a). این محقق در پژوهش دیگری، تعداد ۳۵۱۰ صفحه وب را در سه گروه از لحاظ کاربرد ابرنشانه «توصیف» مورد بررسی قرار داد: (۱) صفحاتی که از طریق یاهو بازیابی شده بودند؛ (۲) صفحاتی که پیوندهای آنها روی صفحات گروه اول قرار داشت؛ (۳) صفحاتی که پیوندهای آنها روی صفحات گروه دوم قرار داشت. یافته‌ها حاکی از آن بودند که ۳۳/۳ درصد از صفحات مورد بررسی، از ابرنشانه «توصیف» استفاده کرده‌اند (Craven 2001b).

«تلوال» در پژوهشی ۶۰،۰۸۷ صفحه وب را برای بررسی وضعیت طراحی صفحات وب و ثبت این صفحات در موتورهای جستجو مورد بررسی قرار داد. موتورهای جستجوی یاهو، هات‌بات، آلتاویستا، ام‌اس‌ان، و اینفوسیک به عنوان ابزار جستجو انتخاب شدند. سپس جستجو بر روی دامنه وب‌سایت‌ها با استفاده از این موتورهای جستجو انجام شد. نتایج نشان دادند که حدود ۳۵ درصد از صفحات وب از ابرنشانه «کلیدواژه» و ۳۳ درصد از ابرنشانه «توصیف» استفاده کرده‌اند (Thelwall 2000).

«کریون» در پژوهش خود با عنوان «ساختار ابرنشانه‌های «توصیف» در صفحات خانگی عمومی» گزارش داد که از ۶۲۸ صفحه وب ثبت شده در یاهو، ۳۵۷ مورد (۵۶/۸٪) دارای ابرنشانه بودند که ۱۶۳ مورد (۲۵/۹٪) از این ابرنشانه‌ها، ابرنشانه «توصیف» بودند (Craven 2000).

مرور پژوهش‌های پیشین نشان می‌دهد که با وجود اثربخشی عناصر ابر داده‌ای هسته دابلین بر بهبود بازیابی و رتبه‌بندی صفحات وب، در نمونه‌های مورد بررسی، از این عناصر به میزان بسیار کمی استفاده شده است. همچنین در میان ابرنشانه‌های زبان نشانه‌گذاری فرامتن، تنها عناصر «توصیف» و «کلیدواژه» بیشترین کاربرد را داشته‌اند و بیشترین توجه به این دو عنصر بوده و برخی از محققین در پژوهش‌های خود تنها حضور این دو عنصر را مورد بررسی قرار داده‌اند.

مرور زمینه‌های نظری پژوهش بیانگر اهمیت کاربرد عناصر ابر داده‌ای برای سهولت دستیابی به منابع رقومی است. همان‌طور که قبلاً نیز بیان شد، به دلیل حجم و تنوع منابع وب، روش‌های سنتی ذخیره و بازیابی اطلاعات کارآمد نیستند؛ به همین جهت استفاده از عناصر ابر داده‌ای برای توصیف و سازماندهی منابع وب پیشنهاد شد.

همان‌طور که مشاهده می‌شود، تحقیقات بسیار کمی به بررسی تمام عناصر ابر داده‌ای هسته دابلین و ابر نشانه‌های زبان نشانه گذاری فرامتن پرداخته‌اند. به علاوه، در تحقیقات قبلی هیچ موردی یافت نمی‌شود که به بررسی حضور عناصر ابر داده‌ای هسته دابلین و ابر نشانه‌های زبان نشانه گذاری فرامتن در صفحات حاصل از جستجو در موتورهای جستجوی گوگل، یاهو و ام‌اس‌ان، که از پرکاربردترین موتورهای جستجو هستند پرداخته باشد. این موارد می‌توانند نشان‌دهنده ضرورت و اهمیت اجرای این پژوهش باشند.

۴. روش‌شناسی پژوهش

این پژوهش از نوع کمی است. از روش تحلیل صفحات وب برای تعیین میزان حضور عناصر ابر داده‌ای به کاررفته در رکوردهای بازیابی شده و در نهایت از روش‌های آماری توصیفی و آزمون‌های همبستگی متغیرها برای تحلیل داده‌ها استفاده شده است.

با توجه به وسعت، تنوع و همچنین طبیعت پویای وب، توصیف دقیق محتوای وب مشکل به نظر می‌رسد. بررسی دقیق همه صفحات وب، یک امر پرهزینه و مشکل است و هر نتیجه‌ای به سرعت قدیمی می‌شود (Rasmevichington et al 2001). به جای بررسی وب به صورت کامل، راه‌حلی پیشنهاد شده است. یکی از راه‌حل‌های پیشنهاد شده، نمونه‌برداری از وب است (Baykan, Cstelberg, and Henzinger 2006). در این پژوهش نیز محقق با طراحی عبارت جستجو و جستجوی آن در موتورهای جستجوی مورد بررسی (گوگل، یاهو، و ام‌اس‌ان)، نمونه خود را انتخاب نمود.

بعد از وارد کردن عبارت جستجوی طراحی شده در هر موتور جستجو، تعداد نتایج به دست آمده از هر موتور جستجو به این ترتیب بود: گوگل: ۱۳۱ مورد، یاهو: ۱۴۷ مورد، و ام‌اس‌ان: ۹۵ مورد. در مجموع حدود ۴۲۵ رکورد بازیابی شده از موتورهای جستجو، جامعه آماری پژوهش حاضر را تشکیل می‌دهند.

۴-۱. مراحل اجرای پژوهش

۴-۱-۱. گزینش موتورهای جستجو. یکی از روش‌های رایج جستجو در وب، جستجو در موتورهای جستجو می‌باشد. برای اجرای پژوهش حاضر سه موتور جستجوی عمومی گوگل، یاهو، و ام‌اس‌ان انتخاب شدند. سایت‌هایی مانند «سرچ‌انجین‌واچ»^۱، «الکسا»^۲، و «سرچ‌انجین‌گاید»^۳ به رتبه‌بندی موتورهای جستجو از لحاظ میزان بازدید و استفاده می‌پردازند. به منظور گزینش موتورهای جستجو، این وب‌سایت‌ها در ژوئن ۲۰۰۷ مورد بازدید قرار گرفتند. در این بازدید مشاهده شد که موتورهای جستجوی گوگل، یاهو و ام‌اس‌ان در سال ۲۰۰۷ بیش‌ترین بازدید و استفاده را داشته‌اند.

در مرحله بعد، برای ساخت عبارت جستجو، موتورهای جستجو از لحاظ امکانات جستجو (در جستجوی پیشرفته) مورد بررسی قرار گرفتند. باید توجه داشت که عبارت جستجو باید به گونه‌ای طراحی می‌شد که با امکانات موجود در جستجوی پیشرفته هر سه موتور جستجو، پشتیبانی شود.

۴-۱-۲. ساخت عبارت جستجو. از آنجا که موتورهای جستجوی مورد بررسی، عمومی هستند و مشتمل بر حوزه‌های مختلف علوم می‌باشند، محقق برای طراحی عبارت جستجو از لحاظ موضوعی محدودیتی نداشت. در این پژوهش عبارت جستجویی انتخاب شد که تعداد نتایج تقریباً برابری در هر سه موتور جستجو به دست دهد تا شرایط برای هر سه موتور جستجو یکسان باشد. پس از جستجوی اصطلاحات مختلف در حوزه علوم کتابداری و اطلاع‌رسانی، عبارت زیر برای نمونه‌گیری از موتورهای جستجو انتخاب شد:

"Knowledge management and information technology" site:edu

عبارت جستجوی حاضر با استفاده از زبان طبیعی ساخته شد. همچنین محدود کردن عبارت جستجو از لحاظ حوزه موضوعی به حوزه آموزشی، تنها امکان‌بازایی صفحات وب آموزشی را فراهم می‌کند، که این خود موجب افزایش اعتبار نتایج به دست آمده می‌شود.

¹ www.searchenginewatch.com

² www.alexa.com/site/ds/top_500

³ www.searchengineguid.com

۴-۱-۳. **گردآوری داده‌ها.** در هفته اول تیرماه ۱۳۸۶ (برابر با هفته آخر ژوئن ۲۰۰۷)، عبارت جستجوی طراحی شده وارد موتورهای جستجوی مورد بررسی شد و ۳۰ نتیجه اول حاصل از جستجوی هر یک از سه موتور جستجو (در مجموع ۹۰ یافته) که در قالب اچ تی ام ال بودند برای بررسی بازیابی شدند. بعد از انتخاب رکوردها، عناصر ابر داده‌ای به کاررفته در سرآیند^۱ صفحات مورد بررسی قرار گرفتند. داده‌های مورد نیاز این تحقیق از طریق مشاهده و بررسی کد منبع^۲ صفحات مورد بررسی گردآوری شد و ابر داده‌های موجود، ثبت شدند.

۵. تحلیل داده‌ها

عناصر ابر داده‌ای به کاررفته در رکوردهای بازیابی شده از موتورهای جستجو، با توجه به فرضیات تحقیق مورد تحلیل قرار گرفت. با توجه به این که داده‌های به دست آمده از نوع اسمی می‌باشند، از آزمون مجذور کای برای آمار تحلیلی استفاده شد. در جدول شماره ۱ فراوانی و درصد فراوانی عناصر ابر داده‌ای به کاررفته در ۹۰ رکورد بازیابی شده به صورت کلی ارائه شده. با توجه به این جدول مشاهده می‌شود که عناصر «عنوان» و «موضوع» در میان مجموعه عناصر ابر داده‌ای هسته دابلین بیشترین کاربرد را داشته‌اند. همچنین، در میان ابر نشانه‌های به کاررفته در هر سه موتور جستجو، ابر نشانه‌های «عنوان»، «نوع منبع»، «کلیدواژه»، «توصیف»، «تولید کننده»، «نویسنده» و «زبان» به میزان بیش تری مشاهده شده‌اند؛ در حالی که ابر نشانه‌های «تجدید نظر کردن»، «روبو» و «حق مؤلف» حضور کم تری داشته‌اند. با این حال، همان طور که مشاهده می‌شود، عناصر ابر داده‌ای هسته دابلین در رکوردهای بازیابی شده نسبت به ابر نشانه‌های زبان نشانه گذاری فرامتن، حضور بسیار کمی داشته‌اند و این بیانگر کاربرد بسیار کم عناصر ابر داده‌ای هسته دابلین در طراحی صفحات وب می‌باشد. به علاوه می‌توان گفت که موتورهای جستجوی مورد بررسی، در بازیابی نتایج توجه چندانی به حضور عناصر ابر داده‌ای هسته دابلین نداشته‌اند.

¹ HEAD

² Source code

جدول ۱. فراوانی عناصر ابر داده‌ای به کار رفته در موتورهای جستجو

عناصر ابر داده‌ای هسته	فراوانی	درصد	ابرنشانه‌های زبان فرامتن	فراوانی	درصد
عنوان (E 1)	۱۰	٪۱۱/۱	عنوان (E'1)	۸۷	٪۹۶/۷
پدیدآور (E 2)	۵	٪۵/۶	پدیدآور (E'2)	۱۳	٪۱۴/۴
موضوع (E 3)	۸	٪۸/۹	کلیدواژه (E'3)	۳۳	٪۳۶/۷
توصیف (E 4)	۳	٪۳/۳	تاریخ (E'4)	۶	٪۶/۷
ناشر (E 5)	۳	٪۳/۳	نوع منبع (E'5)	۵۸	٪۶۴/۴
همکار (E 6)	۳	٪۳/۳	زبان (E'6)	۱۲	٪۱۳/۳
تاریخ (E 7)	۳	٪۳/۳	حق مؤلف (E'7)	۳	٪۳/۳
نوع (E 8)	۳	٪۳/۳	تاریخ انقضا (E'8)	۶	٪۶/۷
قالب (E 9)	۳	٪۳/۳	توصیف (E'9)	۳۳	٪۳۶/۷
شناسگر (E 10)	۴	٪۴/۴	تولیدکننده (E'10)	۲۳	٪۲۵/۶
منبع (E 11)	۳	٪۳/۳	روایت (E'11)	۵	٪۵/۶
زبان (E 12)	۳	٪۳/۳	تجدیدنظر کردن (E'12)	۱	٪۱/۱
رابطه (E 13)	۳	٪۳/۳	-----	-----	-----
پوشش (E 14)	۳	٪۳/۳	-----	-----	-----
حقوق (E 15)	۳	٪۳/۳	-----	-----	-----

۶. یافته‌های پژوهش

در رکوردهای بازیابی شده از گوگل، عناصر ابر داده‌ای هسته دابلین به میزان بسیار کم‌تری نسبت به ابرنشانه‌های زبان نشانه‌گذاری فرامتن به کار رفته بودند. بیش‌ترین میزان کاربرد عناصر ابر داده‌ای هسته دابلین در رکوردهای بازیابی شده از گوگل، مربوط به عناصر «عنوان»، «موضوع» و «پدیدآور» بود؛ درحالی‌که در میان ابرنشانه‌های زبان نشانه‌گذاری فرامتن، عناصر «عنوان»، «نوع منبع»، «توصیف» و «کلیدواژه» بیش‌ترین کاربرد را داشتند.

در رکوردهای بازیابی شده از یاهو، ابر نشانه‌های زبان نشانه‌گذاری فرامتن نسبت به عناصر ابر داده‌ای هسته دابلین حضور بیش تری داشتند. در میان عناصر ابر داده‌ای هسته دابلین به کاررفته در رکوردهای بازیابی شده، عناصر «عنوان»، «پدیدآور» و «شناسگر» بیش ترین کاربرد را داشتند؛ در حالی که عناصر «عنوان»، «نوع منبع»، «کلیدواژه» و «توصیف» از مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن، در رکوردهای بازیابی شده کاربرد بیش تری داشتند.

در رکوردهای بازیابی شده از ام‌اس‌ان، عناصر ابر داده‌ای هسته دابلین میزان حضور بسیار کمی دارند؛ در حالی که ابر نشانه‌های زبان نشانه‌گذاری فرامتن در این رکوردها به میزان بیش تری مشاهده شده‌اند. عناصر «عنوان» و «موضوع» از مجموعه عناصر ابر داده‌ای هسته دابلین و عناصر «عنوان»، «نوع منبع»، «توصیف» و «کلیدواژه» از مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن، بیش ترین کاربرد را در رکوردهای بازیابی شده از ام‌اس‌ان داشته‌اند.

ام‌اس‌ان بیش ترین میزان کاربرد عناصر ابر داده‌ای هسته دابلین را داشته است؛ در حالی که کم ترین میزان کاربرد این عناصر مربوط به یاهو می‌باشد. در میان عناصر ابر داده‌ای هسته دابلین، عناصر «عنوان» و «موضوع» بیش ترین کاربرد را در هر سه موتور جستجو داشته‌اند. در میان ابر نشانه‌های زبان نشانه‌گذاری فرامتن به کاررفته در هر سه موتور جستجو، عناصر «عنوان»، «نوع منبع»، «کلیدواژه» و «توصیف» به میزان بیش تری مشاهده شده‌اند.

در مجموع، یافته‌های این پژوهش نشان می‌دهند که در رکوردهای بازیابی شده از موتورهای جستجو، عناصر ابر داده‌ای هسته دابلین به میزان بسیار کمی به کار رفته‌اند؛ در حالی که ابر نشانه‌های زبان نشانه‌گذاری فرامتن، کاربرد بیش تری داشته‌اند. به علاوه، براساس آزمون‌های آماری انجام شده و نتایج حاصل از آنها، هیچ گونه رابطه معناداری بین نوع موتور جستجو و حضور عناصر ابر داده‌ای (مجموعه عناصر ابر داده‌ای هسته دابلین و مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن) وجود ندارد. این بدان معنا است که با تغییر موتور جستجو، میزان استفاده از عناصر ابر داده‌ای تغییر نمی‌کند و موتورهای جستجو از حیث لحاظ کردن ابر داده‌ها در بازیابی منابع وب، تفاوت معناداری با یکدیگر ندارند.

۱-۶. آزمون فرضیه‌های تحقیق

فرضیه اول. بین نوع موتور جستجو و وجود عناصر ابر داده‌ای از مجموعه عناصر ابر داده‌ای هسته دابلین رابطه معناداری وجود ندارد.

جدول ۲. نتایج آزمون مجذور کای پی‌رسون عناصر ابر داده‌ای هسته دابلین

عناصر ابر داده‌ای هسته دابلین	Value	df	Asymp. Sig. (2-sided)
عنوان	۰/۹۰۰	۲	۰/۶۳۸
پدیدآور	۰/۴۲۴	۲	۰/۸۰۹
موضوع	۱/۹۲۱	۲	۰/۳۸۳
توصیف	۲/۰۶۹	۲	۰/۳۵۵
ناشر	۲/۰۶۹	۲	۰/۳۵۵
همکار	۲/۰۶۹	۲	۰/۳۵۵
تاریخ	۲/۰۶۹	۲	۰/۳۵۵
نوع	۲/۰۶۹	۲	۰/۳۵۵
قالب	۲/۰۶۹	۲	۰/۳۵۵
شناسگر	۰/۵۲۳	۲	۰/۷۷۰
منبع	۲/۰۶۹	۲	۰/۳۵۵
زبان	۲/۰۶۹	۲	۰/۳۵۵
رابطه	۲/۰۶۹	۲	۰/۳۵۵
پوشش	۲/۰۶۹	۲	۰/۳۵۵
حقوق	۲/۰۶۹	۲	۰/۳۵۵

همان طور که جدول ۲ نشان می‌دهد، مقدار p برای همه عناصر ابر داده‌ای بیش‌تر از احتمال ۰/۰۵ است؛ بنابراین فرضیه پژوهش تأیید می‌شود و به احتمال ۹۵ درصد بین حضور مجموعه عناصر ابر داده‌ای هسته دابلین و موتور جستجو (گوگل، یاهو، و ام‌اس‌ان) رابطه معناداری وجود ندارد.

فرضیه دوم. بین نوع موتور جستجو و وجود عناصر ابر داده‌ای از مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن، رابطه معناداری وجود ندارد.

جدول ۳. نتایج آزمون مجذور کای پیرسون ابر نشانه‌های زبان نشانه‌گذاری فرامتن

Asymp. Sig. (2-sided)	df	Value	عناصر ابر نشانه‌های زبان نشانه‌گذاری فرامتن
۰/۳۵۵	۲	۲/۰۶۹	عنوان
۰/۹۱۴	۲	۰/۱۸۰	پدیدآور
۰/۶۵۰	۲	۰/۸۶۱	کلیدواژه
۰/۲۰۰	۲	۳/۲۱۴	تاریخ
۰/۷۱۲	۲	۰/۶۷۹	نوع منبع
۰/۳۱۵	۲	۲/۳۰۸	زبان
۱/۰۰۰	۲	۰/۰۰۰	حق مؤلف
۰/۵۸۵	۲	۱/۰۷۱	تاریخ انقضا
۰/۵۶۳	۲	۱/۱۴۸	توصیف
۰/۹۴۳	۲	۰/۱۱۷	تولیدکننده
۰/۰۶۴	۲	۵/۵۰۶	روبووت
۰/۳۶۴	۲	۲/۰۲۲	تجدید نظر کردن

همان‌طور که در جدول ۳ مشاهده می‌شود، با توجه به نتایج به دست آمده و احتمال وجود رابطه معنادار بیش از ۰/۰۵، فرضیه پژوهش تأیید می‌شود و با اطمینان ۹۵ درصد می‌توان گفت بین حضور عناصر مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن و نوع موتور جستجو (گوگل، یاهو، و ام‌اس‌ان) رابطه معناداری وجود ندارد.

۲-۶. پاسخ به پرسش‌های پژوهش

یافته‌های پژوهش حاضر برای پاسخگویی به یک پرسش اساسی و دو پرسش فرعی می‌باشد.

۱. آیا در صفحات باز یابی شده از موتورهای جستجوی عمومی، عناصر ابر داده‌ای به کار رفته‌اند؟

۱-۱. میزان به کارگیری عناصر ابر داده‌ای هسته دابلین در صفحات وب چقدر است؟

۲-۱. میزان به کارگیری ابر نشانه‌های زبان نشانه‌گذاری فرامتن در صفحات وب چقدر است؟

در موتور جستجوی گوگل در میان عناصر ابر داده‌ای هسته دابلین، عناصر «عنوان»، «موضوع» و «پدیدآور» بیش تر استفاده شده‌اند. در رکوردهای بازیابی شده از موتور جستجوی یاهو، تنها عناصری چون «عنوان»، «پدیدآور» و «شناسگر» قابل مشاهده‌اند. همچنین، در میان رکوردهای بازیابی شده از موتور جستجوی ام‌اس‌ان، عناصر «عنوان» و «موضوع» بیش ترین کاربرد را داشته‌اند. در مجموع با توجه به جدول شماره ۱، عناصر «عنوان» و «موضوع» بیش ترین کاربرد را داشته‌اند.

در موتور جستجوی گوگل، عناصر «عنوان»، «نوع منبع»، «توصیف» و «کلیدواژه» در میان ابر نشانه‌های زبان نشانه‌گذاری فرامتن بیش ترین کاربرد را داشته‌اند. عناصر «عنوان»، «نوع منبع»، «کلیدواژه» و «توصیف»، عناصری هستند که در صفحات بازیابی شده از موتور جستجوی یاهو بیش ترین کاربرد را داشته‌اند. همچنین در صفحات بازیابی شده از موتور جستجوی ام‌اس‌ان، عناصری مانند «عنوان»، «نوع منبع»، «توصیف» و «کلیدواژه» بیش ترین کاربرد را داشته‌اند. در مجموع جدول شماره ۱ نشان می‌دهد که عناصر «عنوان»، «نوع منبع»، «کلیدواژه» و «توصیف» نسبت به دیگر عناصر به میزان بیش تری مشاهده شده‌اند.

نتایج حاصل از پژوهش، بیانگر کاربرد عناصر ابر داده‌ای در صفحات بازیابی شده از موتورهای جستجوی عمومی می‌باشد. اما همان‌طور که مشاهده شد، عناصر ابر داده‌ای هسته دابلین به حدی کم است که می‌توان گفت این عناصر در صفحات بازیابی شده به کار نرفته‌اند. این در حالی است که در این صفحات، برخی از عناصر مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن به میزان نسبتاً بیش تری به کار رفته‌اند. با این حال، میزان کاربرد عناصر ابر داده‌ای در این صفحات کم است.

۷. بحث و نتیجه‌گیری

با وجود تحقیقات انجام شده در مورد اثربخشی حضور عناصر ابر داده‌ای در بازیابی و رتبه‌بندی صفحات وب (مانند تحقیقات انجام شده توسط افرادی همچون «هنشا» و «الاسکاس»، «ژانگ» و «دیمیتروف»، «محمد»، و «شریف»)، نتایج تحقیق حاضر و نیز تحقیقات پیشین بیانگر کاربرد اندک این عناصر در صفحات وب است (Henshaw and Valauskas 2001; Zhang and Dimitroff 2004; Zhang and Dimitroff 2005a; Zhang and Dimitroff 2005b; Mohamed 2006). شریف ۱۳۸۵.

نتایج حاصل از این پژوهش با نتایج تحقیق «وینارد» مبنی بر استفاده بسیار کم از عناصر ابر داده‌ای هسته دابلین هماهنگی دارد (Vineyard 2001).

در پژوهش انجام شده توسط «تلوال»، نتایج نشان داد که ۳۵ درصد از صفحات وب از ابر نشانه «کلیدواژه» و ۳۳ درصد از ابر نشانه «توصیف» استفاده کرده‌اند (Thelwall 2000). همچنین در پژوهش انجام شده توسط «کریون» مشاهده شد که کاربرد ابر نشانه «توصیف» در صفحات وب مورد بررسی در سال‌های ۲۰۰۰ و ۲۰۰۱، به ترتیب ۳۶/۵ و ۳۵/۵ درصد بوده است (Craven 2001a). پژوهش دیگر وی، نشان داده که از مجموع ۳۵۱۰ صفحه مورد بررسی، ۳۳/۳ درصد از ابر نشانه «توصیف» استفاده کرده‌اند (Craven 2001b). همان‌طور که مشاهده می‌شود نتایج حاصل از این پژوهش (کاربرد ۳۶/۷ درصد عنصر «توصیف» در صفحات مورد بررسی) با پژوهش‌های انجام شده توسط «کریون» همخوانی دارد.

تحقیق «وینارد» نیز نشان می‌دهد که در نمونه‌های مورد بررسی، عناصر ابر داده‌ای هسته دابلین به میزان بسیار کم‌تری نسبت به ابر نشانه‌های زبان نشانه‌گذاری فرامتن استفاده شده است (Vineyard 2001). نتایج تحقیقات انجام شده توسط «علیمحمدی» و «پیشوا» و «مجیدی» نشان می‌دهد که در بین ابر نشانه‌های زبان نشانه‌گذاری فرامتن، عناصر «توصیف» و «کلیدواژه» اهمیت بیش‌تری دارند (علیمحمدی ۱۳۸۲؛ پیشوا و مجیدی ۱۳۸۶). نتایج تحقیق حاضر نشان می‌دهد که از مجموعه ابر نشانه‌های زبان نشانه‌گذاری فرامتن، عناصر «عنوان»، «نوع منبع»، «توصیف» و «کلیدواژه» در صفحات بازیابی شده از موتورهای جستجو، بیش‌ترین کاربرد را داشته‌اند. بخشی از نتایج تحقیق حاضر با تحقیقات پیشین مطابقت دارد؛ چراکه این نتایج نیز بیانگر کاربرد عناصر «توصیف» و «کلیدواژه» در موتورهای جستجو می‌باشند؛ اما همان‌طور که مشاهده می‌شود، عناصر «عنوان» و «نوع منبع» کاربرد بیش‌تری نسبت به عناصر «توصیف» و «کلیدواژه» داشته‌اند.

از جمله دلایل کاربرد کم عنصر ابر داده‌ای هسته دابلین در طراحی صفحات وب، پیچیدگی این عناصر است. پیچیدگی این عناصر باعث می‌شود طراحان صفحات وب از این عناصر کم‌تر استفاده کنند. بنابراین، پیشنهاد می‌شود عناصر موجود در طرح‌های ابر داده‌ای تا حد امکان ساده شوند تا توسط طراحان صفحات وب، قابل درک و کاربرد باشند. با توجه به پیچیدگی عناصر ابر داده‌ای نیاز است تا به افرادی به‌منظور طراحی عناصر

برای صفحات، آموزش داده شود، که این امر خود می‌تواند زمان‌بر و پرهزینه باشد. در نتیجه طراحان صفحات وب اغلب ترجیح می‌دهند که کم‌تر از این عناصر استفاده کنند. همچنین، یکی دیگر از دلایل کاربرد کم این عناصر، عدم پشتیبانی موتورهای جستجو از عناصر ابر داده‌ای هسته دابلین است. چنانچه گزارش موتور جستجو نشان داد، هیچیک از موتورهای جستجو از این عناصر پشتیبانی نمی‌کنند: آن‌ها این عناصر را نمایه‌سازی نمی‌کنند، یا امکانی برای جستجو در فیلدهای هسته دابلین فراهم نمی‌آورند. با توجه به این که مطالعات انجام شده نشان می‌دهند که عناصر ابر داده‌ای هسته دابلین به میزان بسیار کمی در صفحات باز یابی شده استفاده شده‌اند، و با توجه به اهمیت این عناصر در باز یابی اطلاعات، پیشنهاد می‌شود طراحان موتورهای جستجو در الگوریتم‌های باز یابی و رتبه‌بندی خود، به این عناصر توجه بیش‌تری داشته باشند. طراحی فیلدهایی برای جستجو از طریق عناصر ابر داده‌ای می‌تواند باعث افزایش امکان باز یابی صفحات مرتبط، و بهبود رتبه آن‌ها در سیاهه نتایج جستجو شود.

۸. فهرست منابع

- پیشوا، پگاه، و موسی مجیدی. ۱۳۸۶. مطالعه میزان به کارگیری ابر داده‌های زبان نشانه‌گذاری فرامتن در مجله‌های دسترسی آزاد وب. فصلنامه علوم و فناوری اطلاعات ۲۲ (۳): ۱۸-۱.
- سعادت، علیرضا. ۱۳۸۵. مطالعه عناصر ابر داده‌ای دابلین کور در سایت‌های وب کتابخانه‌های ملی جهان. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه شیراز.
- شریف، عاطفه. ۱۳۸۵. بررسی میزان اثربخشی عناصر ابر داده‌ای بر رتبه‌بندی صفحات وب توسط موتورهای کاوش. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه تهران.
- صفوی، مهدی. ۱۳۸۲. سنجش میزان اثربخشی عناصر ابر داده‌ای دابلین کور در باز یابی صفحات وب: مطالعه صفحات وب IRANIAN INTERNATIONAL JOURNAL OF SCIENCE. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه تهران.
- علیمحمدی، داریوش. ۱۳۸۲. ارزیابی ابر نشانه‌های وب سایت‌های ایران. پایان‌نامه کارشناسی ارشد کتابداری و اطلاع‌رسانی، دانشگاه تهران.
- واگان، لیون. ۲۰۰۵. روش‌های آماری برای متخصصان کتابداری و اطلاع‌رسانی: رویکرد کاربردی و ساده برای درک، استفاده و تفسیر علم آمار. ترجمه محمدرضا قانع، کیوان کوشا. ۱۳۸۴. تهران: چاپار.

- Bar-Ilan, J. 2005. Comparing rankings of search results on the web. *Information Processing and management* 41(6): 1511-1519.
- Baykan, E., S. Cstelberg, and M. Henzinger. 2000. A comparison of techniques for sampling web pages. <http://www.ra.ethz.ch/CDstore/www2006/iiweb2006.cs.uiuc.edu/3.pdf> (accessed June 17, 2007).
- Craven, T. C. 2000. Features of description meta-tags in public home pages. *Journal of Information Science* 26(5): 303-311.
- Craven, T. C. 2001a. Changes in meta-tag descriptions over time. *First Monday* 6(10) <http://firstmonday.org/issues/issue6-10/craven/index.html> (accessed Sep 24, 2009).
- Craven, T. C. 2001b. Description meta-tags in public home and linked pages. *LIBRES: Library and Information Science Research Electronic Journal* 11(2). <http://libres.curtin.edu.au/LIBRE11N2/craven.htm> (accessed Sep 24, 2009).
- Craven, T. C. 2004. Variations in use of Meta tag descriptions by Web pages in different languages. *Information Processing and Management* 40(3): 479-493.
- Henshaw, R. 1999. The first Monday metadata project. *Libri*49: 125-133.
- Henshaw, R., and E. J. Valauskas. 2001. Metadata as a catalyst: Experiments with metadata and search engines in the internet journal, *First Monday*. *Libri*51: 86-101.
- Jansen, B. J., A. Spink, and T. Saracevic. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing and management*, 36(2): 207-227.
- Mohamed, Kh. 2006. The impact of metadata in resources discovering. *Online information review* 30(2): 155-168.
- Noruzi, A. 2007. A study of HTML title tag creation behavior of academic web sites. *Journal of academic librarianship* 33(4): 501-506.
- Rasmevichington, P., D. M. Pennock, S. Lawrence and C. L. Giles. 2001. Methods for sampling pages uniformly from the World Wide Web. In *processing of the AAAI Fall Symposium on using uncertainty within computation*. <http://citeseer.ist.psu.edu/455785.HTML> (accessed Jun 17, 2007).
- Thelwall, M. 2000. Commercial web sites: lost in cyberspace. *Internet Research: Electronic Networking and Applications Policy* 10(2): 150-159.
- Vineyard, P. 2001. *An analysis of embedded metadata usage on the World Wide Web*. A Master's paper for the M.S. in L.S. degree. <http://neoref.ils.unc.edu/2698.pdf> (accessed Jun 20, 2007).
- Zhang, J. and A. Dimitroff. 2004. Internet search engine response to metadata Dublin Core implementation. *Journal of Information Science* 30(4): 310-320.
- Zhang, J. and A. Dimitroff. 2005a. The impact of webpage content characterization on webpage visibility in search engine results (Part I). *Information Processing and Management*41: 665-690.
- Zhang, J. and A. Dimitroff. 2005b. The impact of metadata implementation on webpage visibility in search engine results (Part II). *Information Processing and Management*41: 691-715.

