

بررسی تأثیر استفاده از ابربرچسب در بهبود بازیابی اطلاعات از وب

منصوره دمرچی لو*

دانشجوی کارشناسی ارشد علوم کتابداری و اطلاع‌رسانی،
دانشگاه علوم پزشکی

دریافت: ۱۳۸۷/۱۰/۱۳ پذیرش: ۱۳۸۸/۰۴/۲۷

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۵۲۰۶-۱۷۳۵
شاپا(الکترونیکی) ۵۵۸۳-۲۰۰۸
نمایه در SCOPUS و ISC
<http://jlist.irandoc.ac.ir>
دوره ۲۶ | شماره ۴ | صص ۹۲۵-۹۴۰
تابستان ۱۳۹۰

نوع مقاله: مروری

چکیده: این مقاله با مرور و بررسی مقالات پژوهشی، تأثیر استفاده از ابربرچسب را در بهبود بازیابی اطلاعات از وب جهان‌گستر مورد مطالعه قرار داده است. نتایج نشان داد که بیشتر پژوهش‌های انجام‌شده، بر تأثیر کاربرد ابربرچسب در بازیابی اطلاعات تأکید می‌کند. اما، از آنجا که در نتایج یکی از پژوهش‌های اخیر، اختلاف معناداری در استفاده و عدم استفاده از ابربرچسب یافت نشد، نیاز به پژوهش‌های بیشتر را در این زمینه الزامی کرد. آنچه مسلم است این ابزار به نویسندگان و طراح وبسایت کمک می‌کند تا مقداری کنترل روی نحوه نمایه‌سازی وبسایت خود توسط موتورهای کاوش داشته باشند و این امتیازی نیست که بتوان به راحتی از آن گذاشت.

کلیدواژه‌ها: ابربرچسب‌ها، موتورهای کاوش، نمایه‌سازی، بازیابی اطلاعات

*mansoureh.damirchi@gmail.com

۱. مقدمه

میلیون‌ها صفحه با اطلاعات مفید و غیرمفید در وب وجود دارد. مقدار زیادی از این اطلاعات درست یا غلط توسط میلیون‌ها انسان در سراسر جهان هر روز جستجو و استفاده می‌شوند. لورنس و گیلز^۱ تخمین زده‌اند که حدود ۸۵ درصد از استفاده‌کنندگان اینترنت از موتور کاوش استفاده می‌کنند. همچنین، پژوهش زونا^۲ نشان می‌دهد که ۷۷ درصد از وقت استفاده‌کنندگان صرف موتور کاوش می‌شود. بنابراین، موتوهای کاوش نقش مهمی در فرایند بازیابی اطلاعات در شبکه جهانی وب بازی می‌کنند (Alimohammadi 2003).

اما، رشد مدام در مقدار محتوای قابل دسترس در محیط وب، مشکلات و مسائلی را برای کاربران ایجاد کرده است. آنها معتقدند که امروزه، نه تنها پیدا کردن اطلاعات سخت است، اطلاعاتی که به دست می‌آید آن قدر بی‌شمار و کلی است که انتخاب کلیدواژه‌های دقیق برای جستجو مشکل شده است (Kruger 1999).

در واقع از یک سو، تعداد زیادی صفحه هر روز منتشر می‌شود و از سوی دیگر، ابزارها و موتورها و روش‌های جستجو و بازیابی اطلاعات در این محیط در حال تغییر و چالش، نمی‌توانند با این تعداد زیاد منابع اطلاعاتی به‌طور مؤثر مقابله کنند (Alimohammadi 2005).

کروگر بیان می‌کند که وب فاقد بسیاری از ساختارهاست تا استفاده‌کنندگان را قادر سازد که به‌طور مؤثر و کارآمد کار کنند (Kruger 1999) و کریستین سن معتقد است این مسأله تاحدی به این دلیل است که اطلاعات موجود در اینترنت به‌روش منابع کتابخانه‌ای فهرست نمی‌شود (Christensen 1999) و علیمحمدی به نبود یک ساختار نظام‌مند در وب که جستجوکنندگان اطلاعات بتوانند دلخواه خود را در کمترین زمان تهیه کنند، اشاره می‌کند (Alimohammadi 2003).

با توجه به اینکه بیشتر استفاده‌کنندگان اینترنت، از موتورهای کاوش به‌عنوان ابزار اولیه برای یافتن منابع اینترنتی استفاده می‌کنند، ناشران وب تلاش می‌کنند تا دسترس‌پذیری صفحه وب خود را در فهرست نتایج جستجوی موتورهای افزایش دهند (Zang and Dimitroff 2005). موتورهای کاوش، به‌عنوان ابزاری برای سازماندهی و بازیابی اطلاعات در اینترنت، به کمک یک ربات نرم‌افزاری به نام عنکبوت^۳ یا خزنده وب^۴، به‌طور مرتب وب را مرور و سایت‌ها را بازدید می‌کنند و آنها را می‌خوانند. سپس، اطلاعات سایت را جمع‌آوری کرده، به موتور کاوش ارائه می‌دهند (داورپناه ۱۳۸۵، ۹۲). در واقع کار خزنده، بازیابی کدهای HTML صفحات است.

1. Lawrence and Giles
3. Web spider

2. Zona
4. Web crawler

اگر یک صفحه وب توسط موتورهای نمایه‌سازی نشود و یا اگر نمایه‌سازی انجام شود، ولی بهینه‌سازی وب به خوبی انجام نگردد؛ تعداد محدودی از جستجوگران اینترنت به آن دسترسی می‌یابند. هر چند موتورهای کاوش، به طور گسترده، از نمایه‌سازی خودکار تمام متن صفحات وب استفاده می‌کنند، این امر در بازیابی مجموعه‌های کوچک مؤثر است و در اینترنت با حجم زیاد و متنوع و پویای اطلاعات، مؤثر نیست. با این توصیف، رشد کمی اطلاعات در وب و ضعف ابزار جستجوی اینترنت، باعث ایجاد یک مشکل جدی در بازیابی اطلاعات شده است و نیاز به یک استاندارد ساده، سازگار، انعطاف‌پذیر، و مناسب را برای توصیف اطلاعات اینترنت ایجاد می‌کند تا نمایه‌سازی خودکار اطلاعات موجود در اینترنت را به طور مؤثر و کارآمد تسهیل کند. ابر داده شاید بتواند چنین استانداردی باشد. ابر داده تلاش می‌کند تا درک، شناسایی، توصیف، استفاده، و بازیابی منابع و محتوای اطلاعات اینترنت را تسهیل کند (Zang and Dimitroff 2005).

به طور کلی، ابر داده برای سازماندهی اطلاعات استفاده می‌شود؛ اطلاعات یک منبع اطلاعاتی را توصیف می‌کند؛ راهنمایی برای دسترسی به یک منبع اطلاعاتی است. به طور اساس، ابر داده یک فرایند مؤثر برای توصیف و مکان‌یابی داده‌هایی است که در یک موضوع به خصوص مرتبط هستند. آنها شناسایی و استخراج اطلاعات را از اینترنت تسهیل می‌کنند (Zang and Dimitroff 2004). اصطلاح ابر داده، اصطلاح جدیدی نیست، کتابداران سال‌هاست که در کتابخانه‌ها به فهرست‌نویسی کتاب‌ها و مجلات می‌پردازند، فهرست کتابخانه خود نوعی ابر داده است که برای یافتن کتاب‌ها و مجلات در زمینه‌های موضوعی خاص و بازیابی آنها از قفسه کتابخانه مورد استفاده قرار می‌گیرد. در واقع، همان‌طور که فهرست‌ها در کتابخانه‌ها برای جستجو و سازماندهی و بازیابی منابع کتابی و غیر کتابی مورد استفاده قرار می‌گیرد، ابر داده نیز همچون فهرستی برای جستجو و سازماندهی و بازیابی منابع الکترونیکی موجود در شبکه جهانی اینترنت استفاده می‌شود. بنابراین، با توجه به تعاریف بالا روشن می‌گردد که ابر داده در واقع نوعی روش و ابزار برای سازماندهی و بازیابی منابع الکترونیکی شبکه اینترنت است (محمدی ۱۳۸۳).

اطلاعات ابر داده‌ای در مورد یک منبع الکترونیکی، برای مثال در محیط وب، شامل اطلاعات مربوط به عنوان سایت یا منبع، موضوع، پدیدآورنده سایت، ناشر، تاریخ نشر یا ایجاد سایت، جزئیات مالکیت معنوی اثر، زبان، پوشش، توصیف و کلیدواژه‌های قابل جستجو برای هر منبع و کدها و نشانه‌هایی برای رده‌بندی محتوایی مدارک است. لازم به اشاره است که نشانه‌های ابر داده‌ای در شرایط عادی در مرورگرهای وب نشان داده نمی‌شوند، برای مشاهده آنها می‌توان با انتخاب گزینه source از منوی view، در هر کدام از مرورگرهای وب نظیر اینترنت اکسپلورر^۱ و یا نت اسکپ^۲ می‌توان ابر داده هر سایتی را ملاحظه نمود.

1. Internet Explorer

2. Netscape

امروزه، ابر داده متداول‌ترین مرجع برای توصیف اطلاعات وب جهان‌گستر و سایر منابع الکترونیکی است. چنانچه بپذیریم که زبان HTML یک ابزار برای انتشار سریع اطلاعات از طریق وب است، ابر داده بخشی از فرایند انتشار در اینترنت است که امکان دسترس‌پذیری اطلاعات دیجیتال را افزایش می‌دهد و ابربرچسب که یکی از روش‌های سازماندهی ابر داده است، به فرایند کشف منبع کمک می‌کند (Henshaw and Valauskas 2001). در واقع، ابربرچسب‌ها یکی از انواع ابر داده‌ها هستند که کاربرد آنها در این مقاله مورد بررسی قرار گرفته است. ابربرچسب‌ها، به‌عنوان یکی از شیوه‌های سازماندهی و دسترس‌پذیرسازی اطلاعات دیجیتالی، در برخی از سایت‌ها به کار می‌روند (علیمحمدی ۱۳۸۱) و آنچه در این مقاله بررسی می‌شود، تأثیر استفاده از ابربرچسب در بهبود بازیابی اطلاعات، با استفاده از مرور پژوهش‌های انجام‌شده در این زمینه است.

۲. ابربرچسب چیست و چه کاری می‌کند؟

کدهای صفحات اینترنتی از دو قسمت Head و Body تشکیل شده‌اند. قسمت مربوط به Head در میان دو کد <head> و </head> و قسمت مربوط به Body در میان دو کد <body> و </body> قرار می‌گیرند. کدهایی که در میان کد Body قرار می‌گیرند محتویات یک صفحه اینترنتی را در هنگام نمایش ایجاد می‌کنند (همه چیز درباره متاتگ‌ها، ۱۳۸۷)، ولی به‌جز برچسب عنوان، اطلاعات بخش Head صفحه وب در مرورگر دیده نمی‌شود (Sullivan 2007). در عوض، اطلاعات این بخش، موتورهای کاوش را در شناختن مطالب داخل سایت یاری می‌نماید (متاتگ چیست؟، بی تا). ابربرچسب‌ها اطلاعاتی هستند که در بخش Head صفحه وب قرار می‌گیرند.

تعریف رسمی کنسرسیوم وب از HTML، ابربرچسب را نیز شامل می‌شود (Craven 2001). ابربرچسب‌ها، برچسب‌های پنهان و غیر آشکاری هستند که صاحبان و نویسندگان سایت‌ها را قادر می‌سازند تا حدی بر نحوه نمایه‌شدن صفحه وب کنترل داشته باشند. "ابر" در ابربرچسب از واژه "ابرداده" ناشی شده است که به معنی داده‌ای درباره دیگر داده‌هاست. برای مثال، یک ابربرچسب ممکن است محتوای یک صفحه وب را شرح دهد که داده‌ای درباره دیگر داده‌ها، یعنی صفحه وب است. هر صفحه وب می‌تواند مجموعه متنوعی از ابر داده‌ها را به خود اختصاص دهد (علیمحمدی ۱۳۸۱).

در واقع ابربرچسب‌ها روشی برای سازماندهی ابر داده‌ها هستند که به فرایند کشف منابع کمک می‌کنند (Henshaw 1999) و به تعبیر کریستین سن ساده‌ترین شکل ابر داده هستند (Christensen 1999).

ابربرچسب‌ها دستوراتی را در مورد صفحات اینترنتی به موتورهای کاوش و کاربران ارائه

می‌کنند که در صورت آگاهی از کاربرد و استفاده صحیح از آنها، موقعیت وبلاگ و یا وبسایت در موتورهای کاوش مانند گوگل به شدت بهبود می‌یابد (همه چیز درباره متاتگ‌ها ۱۳۸۷). ابربرچسب‌ها می‌توانند برای شناسایی درست یک سند استفاده شوند (به عنوان مثال، نویسنده، تاریخ انقضا، کلیدواژه و...) و ارزش‌ها یا اعتبارها را برای مشخصه‌ها تعیین کنند؛ آنها کمک می‌کنند نویسنده و ناشر مطمئن شوند که منابع آنها در صورت انجام یک جستجوی مناسب، در نتایج جستجو مشاهده خواهد شد. بر اساس نظر سالیوان^۱، ابربرچسب‌ها به عنوان شیشه‌های بزرگ‌نماکننده استفاده می‌شوند و به موتورهای کاوش کمک می‌کنند که روی مهمترین بخش‌های صفحه تمرکز کنند. بنابراین، در صورت تعریف کردن ابربرچسب برای یک صفحه، موتورهای کاوش تمام صفحات را پیدا خواهند کرد و اگر بخواهیم که یک صفحه به رتبه بالایی برسد، لازم است که برای آن عنصر، ابربرچسب را تعریف کنیم (Alimohammadi 2003).

۳. انواع ابربرچسب

دو نوع اصلی ابربرچسب وجود دارد:

۳-۱. نام

این نوع از ابربرچسب برای تعیین اطلاعات داخل صفحه وب استفاده می‌شود که شناسایی صفحه وب شما را برای موتورهای کاوش تسهیل می‌کند. به عنوان مثال، زبان این صفحه، نام طراح، کلمات کلیدی این صفحه و...

مثال: `<Meta name="Keywords" content="Persian macromediax dreamweaver">`

این ابربرچسب از نوع name است؛ مقدار name مساوی با keywords است یعنی این ابربرچسب، کارش شناساندن کلمات کلیدی سایت به موتورهای کاوش است. Content در ابربرچسب، کارش معرفی مقداری است که ابربرچسب با یکی از دو صفت name و یا HTTP-EQUIV در خود دارد (متاتگ چیست؟، بی‌تا).

ابربرچسب نام انواع متفاوتی دارد که در ادامه به برخی از آنها اشاره خواهد شد (همه چیز درباره متاتگ‌ها ۱۳۸۷):

۳-۱-۱. ابربرچسب تولیدکننده^۲

استفاده از این ابربرچسب اجباری نیست ولی وجودش باعث می‌شود تا صفحه شما حرفه‌ای‌تر و معتبرتر باشد. این برچسب مشخص می‌کند که شما با قدرت چه سیستمی صفحه اینترنتی‌تان را اجرا کرده‌اید.

1. Sullivan

2. GENERATOR Meta tag

نمونه کد:

<Meta name="generator" content="نام و آدرس صفحه ارائه‌دهنده وبلاگ و یا نام نسخه سیستم ارائه‌دهنده و یا نام زبان برنامه‌نویسی">

۳-۱-۲. ابربرچسب نوع منبع^۱

برای افزایش ارزش محتوای سایت وب در جریان نمایه‌سازی باید از این نوع ابربرچسب استفاده شود.

نمونه کد:

<META NAME="resource type" CONTENT="document">

۳-۱-۳. ابربرچسب روبات^۲

با کمک این ابربرچسب به موتورهای کاوش دستور می‌دهید:

- صفحه مورد نظر را index کنند. "index"
- صفحه مورد نظر را index نکنند. "noindex"
- پیوندهای موجود در صفحه را index و از آنها پیروی کنند. "follow"
- پیوندهای موجود در صفحه را index و از آنها پیروی نکنند. "nofollow"
- صفحه مورد نظر را به همراه پیوندهای آن index و از پیوندها پیروی کنند. "all"

نمونه کد:

<meta name="robots" content="در این قسمت یکی از مقادیر اشاره‌شده را با توجه به

درخواست‌تان بنویسید"

مثال: <meta name="robots" content="follow">

۳-۱-۴. ابربرچسب توزیع^۳

این ابربرچسب دامنه استفاده وبسایت شما را مشخص می‌کند، به این معنا که مطالب و محتوای شما مخاطبان جهانی دارد و یا مختص به محدوده خاصی است.

- اگر محتوای نوشته‌های وبسایت، مخاطبان جهانی دارد از مقدار "global" استفاده کنید.
- اگر محتوای نوشته‌های وبسایت مخاطب خاص دارد از مقدار "local" استفاده کنید.

نمونه کد:

<meta name="distribution" content="در این قسمت، یکی از مقادیر بالا را استفاده کنید"

1. Resource type Meta tag

2. Robots Meta tag

3. DISTRIBUTION Meta tag

۳-۱-۵. ابربرچسب حق مؤلف^۱

استفاده از این ابربرچسب در صفحاتی رایج است که متون و یا تصاویری را در بردارند که حق تولیدشان مربوط به مسئول وبسایت است.
نمونه کد:

```
<meta name="copyright" content="نام شما و یا نام شرکت مربوطه | تاریخ">
```

۳-۱-۶. ابربرچسب نویسنده

این ابربرچسب، نویسنده وبسایت و یا وبلاگ را مشخص می‌کند.
نمونه کد:

```
<meta name="author" content="نام نویسنده">
```

۳-۱-۷. ابربرچسب طبقه‌بندی^۲

با استفاده از این ابربرچسب، موضوع وبسایت یا وبلاگ را مشخص می‌کنید. استفاده از این کد باعث می‌شود تا موتورهای کاوش، وبسایت شما را در طبقه و موضوع اصلی‌اش به‌نمایش بگذارند.
نمونه کد:

```
<meta name="classification" content="موضوع وبسایت یا وبلاگ">
```

۳-۲. قرینه مقاله‌نامه انتقال فرامتن^۳

این ابربرچسب باعث می‌شود که اطلاعات اضافی همراه صفحه درخواستی شما در Head به مرورگر ارسال شود. به عنوان مثال، می‌توانید تاریخ انقضای اطلاعات را تعیین کنید یا بدون دخالت کاربر، صفحه وب را با یک ابربرچسب خاص پس از چند ثانیه به صفحه دیگری بفرستید (متاتگ چیست؟، بی‌تا).

ابربرچسب قرینه مقاله‌نامه انتقال فرامتن انواع مختلفی دارد که تعدادی از آنها عبارتند از (همه چیز درباره متاتگ‌ها ۱۳۸۷):

۳-۲-۱. ابربرچسب زبان^۴

این ابربرچسب زبان صفحات اینترنتی را مشخص می‌کند. برای زبان فارسی می‌توانید از یکی از دو اسم Farsi و یا Persian استفاده کنید و یا نام زبان مورد نظرتان را از فهرست زبان‌ها انتخاب کنید:

1. COPYRIGHT Meta tag
3. HTTP-EQUIV

2. CLASSIFICATION Meta tag
4. Language Meta tag

نمونه کد:

```
<meta http-equiv="language" content="نام زبان مربوطه را وارد کنید">
```

۲-۲-۳. ابربرچسب اقلام بی اعتبار^۱

با استفاده از این ابربرچسب، برای صفحه مورد نظرتان تاریخ مصرف تعیین می کنید و طبق تاریخی که شما تعیین کرده اید محتویات صفحه مورد نظر پنهان خواهد شد.

نمونه کد:

```
<meta http-equiv="expires" content="تاریخی که برای انقضای مطالب در نظر دارید">
مثال: <meta http-equiv="expires" content="Sun, 09 Jan 2008 00:00:01 GMT">
```

۳-۲-۳. ابربرچسب تجدید کردن^۲

با استفاده از این ابربرچسب می توانید بینندگان سایت یا وبلاگ را به آدرس دیگری انتقال دهید. مدت زمانی که طول می کشد تا انتقال صورت بگیرد بر حسب ثانیه قابل تنظیم است و عددی را که به جای حرف n در کد زیر وارد می کنید مشخص می کند که بعد از گذشت چند ثانیه، بازدیدکننده به صفحه مورد نظر منتقل شود. این کد Optional URL Redirect هم نامیده می شود.

نمونه کد:

```
<meta http-equiv="refresh" content="n:url">
باید به آن منتقل شود"
```

۴-۲-۳. ابربرچسب پاسخ (برقراری ارتباط)^۳

این ابربرچسب، مشخصاتی از نویسنده را در اختیار بازدیدکنندگان قرار می دهد و به نوعی اطلاعات صفحه را تکمیل می کند تا در صورتی که کسی مایل بود نظری را برای نویسنده ارسال کند، یک نشانی از او داشته باشد.

نمونه کد:

```
<meta http-equiv="reply-to" content="you@yourweb.com">
نام نویسنده"
البته، ابربرچسب های توصیف و کلیدواژه از مهمترین ابربرچسب های نام هستند که در نمایه سازی مورد استفاده قرار می گیرند. در بخش بعد، به طور مفصل، به آنها اشاره خواهد شد.
```

۴. مهمترین ابربرچسب ها چه هستند؟

تعداد زیادی ابربرچسب در HTML وجود دارد و تخمین زده شده که بیش از ۵۰ و حتی ۱۲۳

1. DOCUMENT EXPIRES Meta tag
3. Reply to Meta tag

2. REFRESH Meta tag

نوع ابربرچسب موجود است، اما ابربرچسب‌های مورد توجه حوزه علوم کتابداری مواردی هستند که در زمینه سازماندهی و نمایه‌سازی اطلاعات کاربرد دارند و استفاده کننده را قادر می‌سازند تا جای اطلاعات را پیدا کند. مهمترین ابربرچسب‌ها که توسط موتورهای کاوش با هدف نمایه‌سازی استفاده می‌شوند ابربرچسب‌های "توصیف" و "کلیدواژه" هستند (Alimohammadi 2003).

۴-۱. ابربرچسب توصیف^۱

این ابربرچسب برای توصیف محتوای صفحه وب به کار می‌رود و کاربرد این توصیف را به‌هنگام بازیابی وب‌سایت مشاهده خواهد کرد (علیمحمدی ۱۳۸۱). اگر هیچ پاراگراف توصیفی در ابتدای سند وجود نداشته باشد، ابربرچسب توصیف باید استفاده شود. اما، اگر یک پاراگراف توصیفی وجود دارد بهتر است که ابربرچسب توصیف حذف شود، چون ممکن است هنگام به‌روزشدن یکی از موارد فراموش شود. به‌خصوص، این ابربرچسب برای مدارکی که یک متن کوتاه هستند، مفید است. در ضمن بهتر است کلیدواژه‌های مهم در ابتدای توصیف قرار گیرند (Alimohammadi 2003).

بدون این ابربرچسب، موتور کاوش، به‌طور معمول، دو خط اول متن روی سایت را در نظر می‌گیرد و از آن به‌عنوان توصیف استفاده می‌کند. پیشنهاد می‌شود که توصیف به ۲۰-۲۵ واژه یا ۱۵۰-۲۰۰ حرف از متن محدود شود (Christensen 1999). توصیف می‌تواند یک کلمه، جمله یا حتی پاراگراف باشد. در هر مورد، باید به‌طور معقولانه و منطقی، کوتاه و مختصر و شامل همه نکات مهم باشد. البته نباید آنقدر خلاصه و چکیده باشد که نتواند انعکاس مناسبی از محتوا ارائه دهد (Alimohammadi 2003). سایت‌های ونکور وب‌پیج^۲ و اورپوینت^۳ بیان می‌کنند که ابربرچسب توصیف مهم است، به‌خصوص اگر سند شما یک متن خیلی کوتاه دارد. در ارتباط با طول توصیف، منابع مختلف آن را تا ۲۰۰ کاراکتر محدود می‌کنند یا اخطار می‌دهند که تعداد زیادی از موتورهای کاوش فقط ۲۰۰ کاراکتر اولیه را در بازیابی‌ها استفاده می‌کنند. متا مدیک^۴، که کنترل برچسب‌ها را انجام می‌دهد، در صورتی که یک برچسب توصیف بیشتر از ۲۰۰ کاراکتر باشد، اخطار می‌دهد (اگرچه پیشنهاد می‌دهد که ابربرچسب کلیدواژه بین ۳۰۰ تا ۵۰۰ کاراکتر باشد) (Craven 2001).

مثال: ابربرچسب توصیف وب‌سایت کتابخانه ملی پزشکی آمریکا (NLM)

<meta name="description" content="The National Library of Medicine (NLM), on the campus of NIH in Bethesda, Maryland, is the world's largest medical library. It collects materials in all major areas of the health sciences and to a lesser degree in such

1. Description Meta tag

2. Vancouver-Webpages.com

3. Everpoint.com

4. META MEDIC (Northernwebs.com)

areas as chemistry, physics, botany and zoology. The collections stand at 5 million items-books, journals, technical reports, manuscripts, microfilms, photographs and images. Housed within the Library is one of the world's finest medical history collections of old (pre-1914) and rare medical texts, manuscripts, and incunabula." />

۴-۲. ابربرچسب کلیدواژه^۱

ابربرچسب کلیدواژه مشخص می‌کند که موتور کاوش، زیر چه کلیدواژه‌هایی باید جستجوپذیر باشد (Henshaw 1999) و کلیدواژه‌ها به موتور کاوش کمک می‌کنند تا وبسایت را طبقه‌بندی موضوعی کند و این امکان را به کاربران می‌دهد تا صفحات وب را سریع‌تر پیدا کنند (Alimohammadi 2003).

بدون این ابربرچسب موتور کاوش، کلمات را از عنوان و متن سایت انتخاب خواهد کرد (Henshaw 1999). بهتر است کلیدواژه‌هایی مورد استفاده قرار گیرند که ممکن است توسط کاربر برای یافتن اطلاعات در پنجره جستجوی وبسایت تایپ شوند. همچنین، واژه‌های مترادف یا کلمات مرتبط یا کلمات ترکیبی باید مورد استفاده قرار گیرند. اگر یک کلیدواژه که بارها اشتباه نوشته شده است استفاده می‌شود، اشتباه‌های املائی معمول نیز باید در ابربرچسب کلیدواژه لحاظ شود. البته، بهترین پیشنهاد انتخاب کلیدواژه‌ها بر اساس اصطلاح‌نامه است (Alimohammadi 2003). چون بیشتر کاربران هنگام جستجو از حروف بزرگ استفاده نمی‌کنند، بهتر است که کلیدواژه‌ها با حروف کوچک تایپ و توسط یک ویرگول و فاصله از هم جدا شوند (Henshaw 1999). کلیدواژه‌ها باید کمتر از ۱۰۰۰ کاراکتر باشند و این رقم ویرگول‌ها و فاصله‌ها را نیز شامل می‌شود. هرچند بیشتر موتورها در مورد تعداد کلیدواژه‌های قابل مشاهده، محدودیت گذاشته‌اند، بهتر است که همیشه در نظر داشته باشیم که کلیدواژه‌ها به اندازه ممکن خلاصه و خاص باشند (Alimohammadi 2003).

مثال: ابربرچسب کلیدواژه وبسایت کتابخانه ملی پزشکی آمریکا (NLM)

```
<meta name="keywords" content="National Library of Medicine, MEDLINE, MEDLARS, LocatorPlus, MedlinePlus, NLM, Grateful Med, Index Medicus, UMLS, Unified Medical Language System, Metathesaurus, Mesh, Medical Subject Headings, AIDSLINE, Healthstar, Visible Human, PubMed, Clinical Trials" />
```

۵. تولید ابربرچسب

تولید و تهیه ابربرچسب با وجود سایت‌هایی که کار تهیه برچسب‌ها را انجام می‌دهند کار سختی نیست و هر کسی می‌تواند برای سایت یا وبلاگ خود با استفاده از این سایت‌ها ابربرچسب

1. Keyword Meta tag

تولید کند. فقط کافی است که در یکی از موتورهای کاوش کلیدواژه‌های "meta tag generator" را تایپ کند تا فهرستی از سایت‌هایی را که برای این منظور طراحی شده‌اند، مشاهده نماید. برای مثال کدی که در ادامه می‌آید، ابربرچسبی است که برای سایت کتابخانه ملی ایران از طریق سایت <http://www.submitexpress.com> تهیه شده است:

```
<HTML>
<HEAD>
<TITLE>national library and archives of Iran ایران ملی کتابخانه ملی ایران
</TITLE>
<META NAME="Keywords" CONTENT="national library, Iran, OPAC, book,
librarianship, national archive, ISSN, CIP,
گنجینه، جستجوی منابع کتابخانه ملی، کتاب، کتابداری، ایران، فیبا، شاپا"
<META NAME="Description" CONTENT=>
امکان جستجوی همه مدارک موجود در کتابخانه ملی و دسترسی به فایل تمام متن مقالات
مجله‌های فصلنامه کتاب و گنجینه، فصلنامه تحقیقات تاریخی و مطالعات آرشیوی را میسر و سایر
خدمات کتابخانه را نیز معرفی می‌کند.
<META NAME="Author" CONTENT="pria@nlai.ir">
<!-- META Tags generated by http://submitexpress.com/metatag.html -->
</HEAD>
```

۶. تأثیر کاربرد ابربرچسب در بازیابی اطلاعات

مطالعات زیادی در زمینه استفاده از ابربرچسب در وبسایت‌ها انجام شده است. نگاهی به پژوهش‌های انجام شده در این زمینه نشان می‌دهد که بیشتر مطالعات به اندازه‌گیری یا کشف حضور ابربرچسب‌های کلیدواژه یا توصیف در وبسایت‌ها مربوط هستند و در مقابل، تعداد کمی از آنها به ارزیابی تأثیر و کارایی استفاده ابربرچسب‌های کلیدواژه یا توصیف، در بهبود بازیابی اطلاعات در وب پرداخته‌اند. فرض گروه اول این است که کارایی ابربرچسب‌ها / ابرداده در بهبود بازیابی اطلاعات مسلم است و هدف گروه دوم، امتحان و آزمون این فرضیه است (Alimohammadi 2005). در ادامه به نتایج پژوهش‌هایی که در دسته دوم قرار دارند، یعنی بررسی تأثیر و کارایی این ابزار در بازیابی اطلاعات، اشاره می‌شود:

تورنر و براکیل برای ارزیابی تأثیر ابربرچسب‌های کد صفحات وب روی بهبود بازیابی اطلاعات، ۲۰ صفحه در گروه‌های زیر طراحی کردند:

۱. صفحاتی بدون ابربرچسب

۲. صفحاتی با ابربرچسب کلیدواژه

۳. صفحاتی با ابربرچسب توصیف

۴. صفحاتی با هر دو ابربرچسب کلیدواژه و توصیف

این صفحه‌ها به موتورهای آلتاویستا^۱، هات‌بات^۲ و اینفوسیک^۳ ارائه شد. هات‌بات از نمایه کردن این صفحات در پایگاه خود خودداری کرد، بنابراین صفحه‌ها با استفاده از کلیدواژه‌هایی از طریق آلتاویستا و اینفوسیک جستجو شدند. نتایج پژوهش نشان داد که استفاده از ابربرچسب کلیدواژه، دسترس‌پذیری صفحات وب را افزایش و بهبود می‌بخشد (Turner and Brackbill 1998).

هنشاو و والوسکاس پژوهشی به منظور اندازه‌گیری تأثیر ابر داده در بهبود بازیابی اطلاعات انجام دادند. آنها ۳۰ شماره از نشریه اینترنتی فرست ماندی^۴ را به عنوان بافت پژوهش، انتخاب کردند. ابتدا، بر اساس کلیدواژه‌های اقتباس شده از عنوان و متن مقالات انتخابی و نقدهای کتاب و نمایه‌های ۳۰ شماره، چند جستجو توسط آلتاویستا، اکسایت^۵، گوگل، هات‌بات، اینفوسیک، لیکوس^۶ و نورترین لایت^۷ انجام و نتایج یادداشت شد. سپس، ابر داده دوبلین کور و ابربرچسب‌ها به کد این صفحات اضافه شد. ۵ ماه بعد دوباره جستجو تکرار و نتایج یادداشت شد. تحلیل و مقایسه میان نتایج نشان داد که ابر داده دوبلین کور و ابربرچسب‌ها، رتبه مقالات انتخاب شده را در میان نتایج بازیابی شده افزایش می‌دهند (Henshaw and Valauskas 2001).

زانگ و دیمیتروف عملکرد موتورهای کاوش را در ارتباط با دو گروه از صفحات وب بررسی کردند: صفحاتی که در کد صفحه وب، ابر داده داشتند و صفحاتی که فاقد آن بودند. فرضیه آنها این بود که معرفی و ارائه ابر داده دوبلین کور به خصوص عنصر موضوع (معادل ابربرچسب کلیدواژه) نباید قابلیت بازیابی و به تبع آن، قابلیت مشاهده صفحه وب را بهبود بخشد. برای انجام بررسی، صفحات طراحی شده به یک حوزه عمومی وب ارسال شدند تا موتورهای کاوش به آنها امکان دسترسی داشته باشند و آدرس این صفحات به ۷ موتور کاوش یا هو، آلتاویستا، لیکوس، گوگل، اینتایر وب^۸، آلد وب^۹، فست^{۱۰} و ایفواسپیس^{۱۱} ارائه شد. چند جستجو انجام شد و نتایج یادداشت گردید. تحلیل یافته‌ها نشان داد که معرفی و ارائه ابر داده‌ها با محتوای موضوعی به صفحات وب، قابلیت مشاهده را در بیشتر موتورهای کاوش و البته نه در همه آنها بهبود می‌بخشد (Zang and Dimitroff 2004).

1. Altavista.com
4. First Monday
7. Northernlight
10. Fast

2. Hot bot
5. Excite
8. EntireWeb
11. Infospace

3. Infoseek
6. Lycos
9. AlltheWeb

زانگ و دیمیتروف در پژوهش دیگری، به شناسایی عوامل دارای نفوذ و قدرت اجرا و پیاده‌سازی ابر داده که بر بهینه‌سازی موتور کاوش تأثیر می‌گذارد، پرداختند و تأثیر آنها را روی موتورهای جستجو تحلیل کردند و روش‌های عملی و کاربردی برای بهبود قدرت و نفوذ موتورهای کاوش را از لحاظ پیاده‌سازی ابر داده پیشنهاد دادند. برای این هدف، ۴۶ صفحه وب مشتق و اصلاح شده از یک صفحه وب اصلی انتخاب و به اینترنت ارسال شد. آدرس URL این صفحات پردازش شده، به ۱۹ موتور کاوش ارائه گردید. پژوهشگران از یک سؤال فرمول-بندی شده ساده استفاده کردند تا ۱۹ موتور را به صورت هفتگی (به مدت ۲۱ هفته) جستجو کنند. نتایج به دست آمده از هر کدام از موتورها یادداشت شد. تحلیل یافته‌ها نشان داد که ابر داده مکانیسم خوبی برای بهبود قابلیت مشاهده یک صفحه در فهرست نتایج موتور کاوش است. فیلد موضوع ابر داده، مهمترین نقش را نسبت به سایر فیلدهای ابر داده بازی می‌کند که معادل ابربرچسب کلیدواژه است. در انتها، پیشنهاد شده است که کلیدواژه‌ها از عنوان یا متن صفحه مورد نظر انتخاب شوند و اگر کلیدواژه‌های انتخابی، هم در عنوان و هم در متن قید شده باشند، مؤثرتر است (Zang and Dimitroff 2005).

سالیوان بعد از انتشار کار ترور و براکیل^۱ در سال ۱۹۹۸ و هنشاو و والوسکاس^۲ سال ۲۰۰۱، ادعا می‌کند که ابربرچسب کلیدواژه دیگر به خاطر تأثیرش در بهبود بازیابی اطلاعات استفاده نمی‌شود (Sullivan 2007)، اما علیمحمدی به این نکته اشاره می‌کند که باید این مسأله را در نظر داشته باشیم که این ادعا توسط هیچ پژوهشی پشتیبانی نمی‌شود (Alimohammadi 2005). از طرف دیگر، زانگ و دیمیتروف بر اساس پژوهش‌هایی که در سال‌های ۲۰۰۴ و ۲۰۰۵ انجام داده‌اند، نشان می‌دهند که اجرای ابر داده دوبلین کور، به خصوص عنصر موضوع (معادل ابربرچسب کلیدواژه)، به طور مؤثری می‌تواند جایگاه صفحه را در فهرست نتایج موتورهای کاوش بهبود بخشد. با این حال، محمد با پژوهشی درباره تأثیر ابر داده در بازیابی منابع، تلاش می‌کند تا تأثیر استفاده ابر داده را در رتبه‌بندی صفحه برای سه موتور گو^۳، التاویستا و هات‌بات بررسی کند. این موتورها ادعا می‌کنند که استفاده ابر داده در رتبه‌بندی صفحه تأثیر می‌گذارد و در مورد استفاده از کلیدواژه‌های نامرتب در ابر داده اخطار می‌دهند. مطالعه نشان داد که اختلاف معناداری در رتبه صفحاتی که شامل ابر داده (ابربرچسب و دوبلین کور) هستند و آنهایی که نیستند، وجود ندارد. البته در بیشتر موارد، صفحاتی که شامل ابربرچسب هستند در سطح بالاتری نسبت به صفحاتی که ابربرچسب ندارند، رتبه‌بندی می‌شوند، اما این تفاوت آنقدر کم است که تأثیر قابل توجهی را در استفاده از ابر داده نشان نمی‌دهد. فرضیه مطالعه این بود که افزودن عناصر ابر داده برای تفسیر

1. Turner and Brackbill

2. Henshaw and Valauskas

3. Go

صفحات وب در رتبه‌بندی صفحه تأثیر خواهد گذاشت. نتایج نشان داد که این فرضیه در تمام موارد درست نیست. البته مقدار کمی تأثیر مشاهده شد، به خصوص وقتی ابربرچسب‌ها مورد استفاده قرار گرفتند (Mohamed 2006).

۷. کاستی‌های ابربرچسب

فقدان یک استاندارد مشخص و واحد برای تولید ابربرچسب، از مهمترین کاستی‌های این ابزار است. البته ابربرچسب‌ها کم و بیش ساختار مشخصی دارند، اما زبان نشانه‌گذاری فرامتن مشخص نمی‌کند که به طور دقیق، چه ابربرچسب‌هایی وجود دارد و چگونه باید آنها را به کار گرفت (علیمحمدی ۱۳۸۱).

از طرفی، ابربرچسب‌ها هرگز یک راه تضمینی برای به دست آوردن یک رتبه‌بندی بالا در موتورهای کاوش مبتنی بر خزنده‌ها نبوده‌اند (Sullivan 2007). در واقع، نباید انتظار داشت که با افزودن ابربرچسب، وبسایت مورد نظر در نتایج موتورهای کاوش به عنوان مثال، در ردیف ۱۰ سایت برتر قرار می‌گیرد. برای این منظور باید علاوه بر ابربرچسب، از راهبردهای دیگری مثل نسخه‌های خاص صفحه‌های وب که متناسب با ویژگی‌های هریک از موتورهای کاوش باشد، استفاده کرد (Henshaw 1999).

علاوه بر این، باید توجه داشت که تمام موتورهای کاوش ابربرچسب را مورد توجه قرار نمی‌دهند. برخی وبسایت‌ها مثل یاهو و اکسایت مشخص کرده‌اند که ابربرچسب‌ها را به هیچ وجه در نظر نمی‌گیرند چون یاهو به طور کامل، به نمایه‌سازی‌های انسانی وابسته است و اکسایت از نمایه‌سازی‌های تمام‌متن استفاده می‌کند (Mohamed 2006).

یک مشکل جدی که در مورد ابربرچسب کلیدواژه وجود دارد، مشکلی است که اسپم ایندکسینگ^۱ یا اسپمینگ^۲ نامیده می‌شود که در واقع استفاده نادرست از ابربرچسب کلیدواژه است. منظور از استفاده نادرست، مواردی چون تکرار کلیدواژه‌ها به دفعات یا استفاده از کلیدواژه‌هایی است که سایت چیزی درباره آنها ندارد اما مورد توجه و علاقه مردم هستند، مانند سکس. نقطه نظرات مختلفی در مورد بیشترین تکرار کلیدواژه وجود دارد. برخی سه بار و برخی دیگر ۶ مرتبه را پیشنهاد می‌دهند و در نهایت برخی معتقدند که آنها می‌توانند تا ۷ بار تکرار شوند. همچنین، این نکته را باید در نظر داشت که اگرچه ترافیک سایت با استفاده از کلیدواژه‌های نامربوط زیاد خواهد شد، انزجار و تنفر کاربران از این سایت هم به مرور زیاد خواهد شد. البته موتورهای کاوش بزرگتر، در حال حاضر اسپم‌ایندکس را به وسیله نادیده انگاشتن کل برچسب حل می‌کنند (Alimohammadi 2004).

1. Spam indexing

2. Spamming

دیگر نکته قابل توجه، رویکرد انسانی به ابربرچسب‌ها در نمایه‌سازی صفحات وب است. استفاده از ابربرچسب، زمانی می‌تواند اثربخش باشد که تهیه‌کنندگان آن متخصصان اطلاع‌رسانی و آگاه بر مسائل ذخیره و بازیابی اطلاعات دیجیتال باشند، به‌ویژه داشتن تخصص موضوعی اهمیت فراوان دارد. پیش‌بینی دقیق نیازهای اطلاعاتی کاربران یک سایت و درک صحیح فرایند نمایه‌سازی اطلاعات دیجیتالی و توانایی بهره‌گیری از واژگان کنترل‌شده در تعیین مرتبط‌ترین کلیدواژه‌ها، از نکات اساسی در استفاده از ابربرچسب‌هاست (علیمحمدی ۱۳۸۱).

۸. نتیجه‌گیری

نتایج بررسی و مرور مطالعاتی که در زمینه تأثیر کاربرد ابربرچسب در بازیابی اطلاعات انجام شده است، بر تأثیر مثبت استفاده از ابربرچسب موضوع (کلیدواژه) در بازیابی اطلاعات تأکید می‌کند. بدون ابربرچسبی که محتوای سایت را شرح دهد، موتور کاوش از دو خط اول متن صفحه به‌عنوان توصیفگر استفاده می‌کند و این روش منجر به بازیابی نتایج زیادی خواهد شد که به‌احتمال در میان آنها رکوردهای نامربوط زیادی هم وجود دارد (Christensen 1999). البته پژوهش‌های اخیر که توسط محمد در سال ۲۰۰۶ انجام شده است و نتایج، اختلاف معناداری در استفاده و عدم استفاده از ابربرچسب نشان نمی‌دهد، نیاز به پژوهش‌های بیشتر در این زمینه را الزامی می‌کند. به‌رحال ابربرچسب هنوز موضوعی است که عقاید متفاوتی در موردش وجود دارد و به‌همین دلیل مطالعات بیشتری لازم است که در این زمینه انجام شود تا تأثیر و کارایی ابربرچسب در بهبود بازیابی اطلاعات، مورد ارزیابی دقیق‌تری قرار گیرد و این موضوع روشن‌تر شود (Alimohammadi 2004). آنچه مسلم است این است که این ابزار به نویسنده و طراح وب‌سایت کمک می‌کند تا مقداری کنترل روی نحوه نمایه‌سازی وب‌سایت خود توسط موتورهای کاوش داشته باشند و این امتیازی نیست که بتوان به‌راحتی از آن گذشت.

به‌نظر می‌رسد جامعه کتابداری و اطلاع‌رسانی می‌تواند ضمن آشنایی با این پدیده به بهبود شیوه‌های به‌کارگیری ابربرچسب برای ذخیره و بازیابی هرچه بهتر اطلاعات بیاورد (علیمحمدی ۱۳۸۱). کتابداران و اطلاع‌رسانان همواره در طول تاریخ با امر سازماندهی و ذخیره و بازیابی اطلاعات درگیر بوده‌اند و امروزه با این مسأله مواجه هستند که چگونه می‌توان اطلاعات دیجیتال را ذخیره و بازیابی کرد. شاید استفاده از ابربرچسب با تمام کاستی‌هایی که دارد، تا حدی بتواند در رفع این عیب مؤثر باشد.

۹. منابع

داورپناه، محمدرضا. ۱۳۸۵. جستجوی اطلاعات علمی و پژوهشی در منابع چاپی و الکترونیکی. تهران: دبیزش.

علیمحمدی، داریوش. ۱۳۸۱. ابررچسبها و کاربرد آنها در نمایه‌سازی صفحات وب. در مجموعه مقالات ششمین همایش کتابداران سازمان مدیریت و برنامه‌ریزی کشور، ۲۶۱-۲۸۰. تهران: سازمان مدیریت و برنامه‌ریزی کشور، معاونت امور پشتیبانی، مرکز مدارک علمی و انتشارات.

محمدی، حسین. ۱۳۸۳. ابر داده، مفاهیم و کاربردها. مجله الکترونیکی مرکز اطلاعات و مدارک علمی ایران ۲ (۳) http://www.irandoc.ac.ir/data/e_j/vol12/metadata.htm (دسترسی در ۱۳۸۷/۸/۱۴).

همه چیز درباره متاتگها (Meta tags). ۱۳۸۷. وبلاگ شقایق.

<http://shaghayegh.blogspot.com/1387/08/22/post-118> (دسترسی در ۱۳۸۷/۸/۱۴).

متاتگ چیست؟ Meta tag. بی تا. Parsigold Forum. <http://forum.parsigold.com/showthread.php?t=3380>. (دسترسی در ۱۳۸۷/۸/۲۹).

- Alimohammadi, D. 2003. Meta-tag: a means to control the process of web indexing. *Online Information Review* 27 (4): 238-242.
- Alimohammadi, D. 2004. Measurement of the presence of keywords and description Meta-tags on a selected number of Iranian Web Sites. *Online Information Review* 28 (3): 220-223.
- Alimohammadi, D. 2005. Meta-tags: still a matter of opinion. *The Electronic Library* 23 (6): 625-631.
- Christensen, D. 1999. Golden retrievers. *School Library Journal* 45 (11): 38-41.
- Craven, T. C. 2001. 'Description' Meta tags in locally linked web pages. *Aslib Proceedings* 53 (6): 203-216.
- Henshaw, R. 1999. The first Monday metadata project. *Libri* 49 (3): 125-131.
- 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 (2): 86-101.
- Kruger, P. 1999. Meta tags that page. *The Electronic Library* 17 (2): 89-90.
- Mohamed, K. A. F. 2006. The impact of metadata in web resources discovering. *Online Information Review* 30 (2): 155-167.
- Sullivan, D. 2002. Death of a Meta tag. *The Searchenginewatch*. <http://searchenginewatch.com/2165061> (accessed 22 Nov. 2008).
- Sullivan, D. 2007. How to use HTML Meta tags. *The Search engine watch*. <http://searchenginewatch.com/2167931> (accessed 25 Nov. 2008).
- Turner, T.P., and L. Brackbill. 1998. Rising to the top: evaluating the use of the html meta-tag to improve retrieval of World Wide Web documents through Internet search engines. *Library Resources and Technical Services* 42 (4): 258-271.
- Zang, J., and A. Dimitroff. 2004. Internet search engines' response to metadata Dublin core implementation. *Journal of Information Science* 30 (4): 310-320.
- Zang, J., and A. Dimitroff. 2005. The impact of metadata implementation on Web page visibility in search results (part II). *Information Processing and Management* 41 (3): 691-715.

Evaluation of the Impact of Using Meta Tag on the Improvement of Information Retrieval of the World Wide Web

Information
Sciences
& Technology

Mansoureh Damerchiloo*

Master student in Library and Information Science,
Iran University of Medical Science

Iranian Research Institute
For Science and Technology
ISSN 1735-5206
eISSN 2008-5583
Indexed in LISA, SCOPUS & ISC
Vol.26 | No.4 | pp: 925-940
summer 2011

Abstract: This article has evaluated the impact of using Meta tag on the improvement of information retrieval of the World Wide Web. So it reviewed the research papers that have examined Meta tag as a means to enhance information retrieval. Results showed there were some researches that have confirmed the using of Meta tags was efficient for improving information retrieval of the web, but results of a research that has done by Mohamed (2006) indicated no significant difference in the rank of pages that include metadata and those that don't include any metadata. Therefore more studies should be done in this field. In general, Meta tags can provide site owners and authors with a degree of control over how a web page is indexed and this is a considerable advantage that it shouldn't be ignore easily.

Keywords: Meta tags, search engines, indexing, retrieval information

*Corresponding author mansoureh.damirchi@gmail.com