

## بررسی میزان اثربخشی عناصر ابر داده‌ای بر رتبه‌بندی صفحات وب توسط موتورهای کاوش عمومی

عاطفه شریف<sup>۱</sup>

### چکیده

پژوهش حاضر با هدف تعیین میزان اثربخشی عناصر ابر داده‌ای بر رتبه‌بندی صفحات وب توسط سه موتور کاوش عمومی «گوگل»، «ياهو»، و «ام‌اس‌ان» انجام پذیرفت. اثربخشی پنج عنصر ابر داده‌ای نشانه عنوان زبان «اچ.تی.ام.ال»، ابر نشانه‌های کلیدواژه‌ها و توصیف زبان «اچ.تی.ام.ال»، ابر داده‌های عنوان و موضوع از قالب ابر داده‌ای «دابلین کور» که بر بازنمون موضوعی صفحات وب متمرکزند، با روش تجربی آزموده شد. ۸۴ صفحه وب در گروه‌های گواه و آزمون به عنوان جامعه مورد مطالعه تحت دامنه فرعی <http://metadata.irandoc.ac.ir> منتشر شد. از میان موتورهای کاوش انتخابی، دو موتور کاوش ياهو و گوگل، صفحات را نمایه‌سازی کردند. کلیدواژه‌های منحصر به فرد و تکرار شونده که در مرحله طراحی در قالب عناصر ابر داده‌ای به صفحات گروه آزمون افزوده شده بود، بررسی شد و میزان معنی‌داری تفاوت رتبه صفحات دو گروه با استفاده از آزمون غیر پارامتری «یو.من.وایتنی» محاسبه گردید. نتایج، نشانگر اثربخشی دو عنصر «عنوان» و «توصیف زبان اچ.تی.ام.ال» در هر دو موتور کاوش مورد مطالعه است. کلیدواژه‌های منحصر به فرد در ابر نشانه کلیدواژه‌های «اچ.تی.ام.ال» و ابر داده موضوع دابلین کور به بهبود رتبه صفحات در ياهو انجامید. در حالی که تکرار کلیدواژه‌ها در بخش سرآیند صفحه وب، باعث بهبود رتبه در موتور کاوش ياهو شد، گوگل عملکردی معکوس داشت. کلیدواژه‌ها: ابر داده، ابر نشانه‌های اچ.تی.ام.ال، دابلین کور، بازیابی اطلاعات، رتبه‌بندی نتایج کاوش، موتورهای کاوش عمومی، وب، قابلیت دسترسی صفحات وب.

۱. دانشجوی دوره دکترای کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد.

[Atefehsharif@gmail.com](mailto:Atefehsharif@gmail.com)

## مقدمه

همزمان با تولد وب، جهان شاهد تغییرات شگرفی در حوزه ذخیره و بازیابی اطلاعات بوده است. هر روز بر تعداد صفحات وب افزوده می‌گردد و حجم وسیعی از اطلاعات، در بستر آن، به صورت ساختار نیافته<sup>۱</sup> (یالتاقیان<sup>۲</sup>، ۲۰۰۲) و فارغ از کنترل محتوایی و کتاب‌شناختی منتشر می‌شود (زانگ و جاستریم<sup>۳</sup>، ۲۰۰۴؛ گوتلیب و الیوپولوس<sup>۴</sup>، ۲۰۰۳). در چنین وضعیتی، مسئله اساسی، چگونگی کنترل و مدیریت بدنه ساختار نیافته و رشد سریع این بدنه است (اسدی و جمالی مهمویی<sup>۵</sup>، ۲۰۰۴). تاکنون ابزارهای کاوش<sup>۶</sup> از قبیل «موتورهای کاوش»<sup>۷</sup> «ابرموتورهای کاوش»<sup>۸</sup> «راهنماهای موضوعی»<sup>۹</sup> و «نرم‌افزارهای کاوش»<sup>۱۰</sup> محیط مجازی وب را تا اندازه‌ای تحت کنترل و مدیریت خویش درآورده‌اند.

کاربران از میان ابزارهای رایج کاوش، موتورهای کاوش را به عنوان نقطه آغازین ورود به اینترنت تلقی می‌کنند (اسپینک و دیگران<sup>۱۱</sup>، ۲۰۰۱ نقل در دوال و واگان<sup>۱۲</sup>، ۲۰۰۴؛ بار - ایلان<sup>۱۳</sup>، ۲۰۰۵؛ زانگ و دیمیتروف<sup>۱۴</sup>، ۲۰۰۴)، بیش از ۹۵٪ ترافیک کاوش در اینترنت به موتورهای کاوش مربوط است و ۸۰٪ کاربران، اطلاعات مورد نیاز خود را از طریق موتورهای کاوش به دست می‌آورند (هاتلی<sup>۱۵</sup>، ۲۰۰۲ نقل در زانگ و دیمیتروف، ۲۰۰۵ a). یافتن اطلاعات موضوعی ویژه در وب دشواریهایی دارد و هر روز

1. Unstructured.
2. Yaltaghian.
3. Zhang & Jastram.
4. Gotlieb & Eliopoulos.
5. Asadi & Jamali M.
6. Internet search tools.
7. Search engines.
8. Meta search engines.
9. Directories.
10. Search Utilities.
11. Spink et al.
12. Thelwall & Vaughan.
13. Bar\_Ilan.
14. Zhang & Dimitroff.
15. Haltley.

بررسی میزان اثربخشی عناصر ابرداده‌ای بر ... / ۲۴۳

بر حجم این دشواریها افزوده می‌گردد (دروت<sup>۱</sup>، ۲۰۰۰، ص ۲۰۹). تعداد نتایج بازیابی شده موتورهای کاوش، اغلب چنان فراوان است که کاربر عملاً جز مرور چند صفحه نخست نتایج، از سایر صفحات منصرف می‌شود (جانسن، اسپینک، و ساراسویک<sup>۲</sup>، ۲۰۰۰؛ فدایی عراقی<sup>۳</sup>، ۲۰۰۵، ص ۱۳؛ یالتاقیان، ۲۰۰۲) و به ناچار به رتبه‌بندی<sup>۴</sup> ارائه شده موتورهای کاوش اعتماد می‌کند (بار-ایلان، ۲۰۰۵). در این وضعیت، چنانچه صفحه‌ای مرتبط، در رتبه‌های اول جای نگیرد، ممکن است از دید کاوشگر پنهان بماند (زانگ و جاستریم، ۲۰۰۵، ص ۹۲؛ گوتلیب و الیوپولوس، ۲۰۰۳).

از سویی، یکی از اولین دغدغه‌های ناشران وب‌سایتها دستیابی به رتبه‌های برتر در میان وب‌سایتهای مشابه و هم موضوع است. بدین منظور همواره سیاهه‌ای از عناوین<sup>۵</sup>، کلیدواژه‌ها<sup>۶</sup> و توصیفهایی<sup>۷</sup> که احتمال کسب رتبه‌های برتر را دارند، تهیه و در طراحی صفحات لحاظ می‌شود (ریچاردسون، ۲۰۰۳<sup>۸</sup> نقل در زانگ و دیمیتروف، ۲۰۰۵a). طراحی نرم‌افزارهایی چون «تحلیلگر چگالی کلیدواژه»<sup>۹</sup> «ورد ترکر»<sup>۱۰</sup> «وب پزیشن گولد»<sup>۱۱</sup> و شکل‌گیری و گسترش وب‌سایتهایی که خدمات توصیه‌ای و مشاوره‌ای بهینه‌سازی صفحات<sup>۱۲</sup> را به منظور کسب رتبه‌های برتر در موتورهای کاوش ارائه می‌دهند، تأییدی بر حساسیت و توجه به این مسئله است.

نتایج تحقیقات، تفاوت‌های قابل ملاحظه‌ای را میان الگوریتمهای رتبه‌بندی موتورهای کاوش عمومی نشان می‌دهد (بار-ایلان، ۲۰۰۵). اطلاعات مربوط به الگوریتمهای

- 
1. Drott.
  2. Jansen ; Spink.; and Saracevic.
  3. Fadaie Araghi.
  4. Ranking.
  5. Title.
  6. Keyword.
  7. Description.
  8. Richardson.
  9. Keyword Density analyzer.
  10. word tracker.
  11. web poisons gold.
  12. Search Engine Optimization (SEO).

رتبه‌بندی موتورهای کاوش به صورت طبقه‌بندی شده<sup>۱</sup> و به عنوان اسرار تجاری<sup>۲</sup> محافظت می‌شود. حفظ حالت رقابتی و جلوگیری از سوء استفاده طراحان وب‌سایتها از این اطلاعات، از جمله دلایل حفاظت هستند (بار-ایلان، ۲۰۰۵، ص ۱۵۱۲). با وجود ابهامهای موجود در زمینه الگوریتم، رتبه‌بندی موتورهای کاوش، پژوهش پیرامون چگونگی رتبه‌بندی نتایج متوقف نشده است و تلاشهایی در زمینه کشف عوامل اثرگذار و تعیین میزان اثرگذاری آن عوامل، انجام پذیرفته است (ترنر و برک بیل<sup>۳</sup>، ۱۹۹۸؛ زانگ و دیمیتروف، ۲۰۰۴؛ ۲۰۰۵a، ۲۰۰۵b، صفری<sup>۴</sup>، ۲۰۰۵؛ محمد<sup>۵</sup>، ۲۰۰۶؛ هنشاش و والاسکاس<sup>۶</sup>، ۲۰۰۱).

### محدوده بررسی

صفحات وب متشکل از سه جزء اند: ۱. معنا<sup>۷</sup> یا محتوا<sup>۸</sup>، ۲. بستر نحوی<sup>۹</sup> یا ساختار<sup>۱۰</sup> و ۳. پیوندهای فرامتنی<sup>۱۱</sup>. محتوا در بستر نحوی زبانهای نشانه گذاری<sup>۱۲</sup> که قالب یا ساختار ارائه محتوا را فراهم می‌آورند، جای می‌گیرد و ارتباط میان اجزای اطلاعاتی از طریق پیوندهای فرامتنی حاصل می‌شود. هر یک از این عناصر - محتوا، ساختار و پیوندهای فرامتنی - ویژگیهای خاصی دارند که بستر ارزیابی صفحات را فراهم می‌آورد و به طور بالقوه در بهبود کیفی رتبه‌بندی حاصل از کاوش اثرگذار است (کوودو-تررو<sup>۱۳</sup>، ۲۰۰۴).

1. Classified.
2. Trade Secrets.
3. Truner & Brackbill.
4. Safari.
5. Mohamed.
6. Henshaw & Valauskas.
7. Semantic.
8. Content.
9. Syntax.
10. Structure.
11. Hyper- Links.
12. Markup Languages.
13. Quevedo-Torrero.

بررسی میزان اثربخشی عناصر ابر داده‌ای بر ... / ۲۴۵

گوگل بیش از هزار عامل را در نظام رتبه‌بندی نتایج خود مدنظر دارد، اما به دلیل ماهیت تجاری و حفظ یکپارچگی نتایج کاوش، از ذکر جزئیات بیشتر خودداری می‌کند (گوگل<sup>۱</sup>، ۲۰۰۴). توافقی بر سر مؤثرتر بودن یک عامل نسبت به سایر عوامل وجود ندارد (فیشکین<sup>۲</sup>، ۲۰۰۵) و هر یک از موتورهای کاوش، الگوریتم رتبه‌بندی خاصی را دنبال می‌کنند (هنشا و والاسکاس، ۲۰۰۱، ص ۹۲). با وجود این، ساختار ابر داده‌ای<sup>۳</sup>، محتوای صفحه<sup>۴</sup> و (عوامل داخلی)، و وضعیت ارجاعات فرامتنی<sup>۵</sup> و- (عوامل خارجی) - از جمله عوامل مؤثر بر رتبه‌بندی ذکر شده است (زانگ و جاستریم، ۲۰۰۵).

در این پژوهش، از میان عوامل مؤثر شناخته شده بر رتبه‌بندی نتایج کاوش، تمرکز بر عناصر ابر داده‌ای<sup>۶</sup> است و از میان عناصر ابر داده‌ای، پنج عنصر به عنوان معیار ربط فنی<sup>۷</sup> انتخاب شده و میزان اثرگذاری هر یک (به صورت منحصر به فرد و در صورت تکرار) بر رتبه صفحات بازیابی شده توسط موتورهای کاوش عمومی، بررسی و تجزیه و تحلیل شده است. عناصر ابر داده‌ای مورد بررسی در این پژوهش عبارتند از: نشانه عنوان<sup>۸</sup> از زبان «اچ.تی.ام.ال»، ابر نشانه‌های کلیدواژه‌ها<sup>۹</sup> و توصیف<sup>۱۰</sup> از زبان «اچ.تی.ام.ال»، دو ابر داده عنوان<sup>۱۱</sup> و موضوع<sup>۱۲</sup> از قالب ابر داده‌ای دابلین کور.

اگر چه نشانه عنوان زبان «اچ.تی.ام.ال»، عنصر ابر داده‌ای محسوب نمی‌شود؛ اما به واسطه اهمیت ویژه اش (سالیوان<sup>۱۳</sup>، ۲۰۰۲؛ نوروزی<sup>۱۴</sup>، ۲۰۰۵) در میان سایر نشانه‌های «اچ.تی.ام.ال»، در کنار چهار عنصر ابر داده‌ای دیگر بررسی شده است.

1. google.
2. Fishkin.
3. Metadata Structure.
4. Webpage Content.
5. Hyperlink Cited status.
6. Metadata elements.
7. Technical Relevance.
8. <title> </title>
9. <meta name="description" content=" ">
10. <meta name="keywords" content=" ">
11. <meta name="DC.Title" content=" ">
12. <meta name="DC.Subject" content=" ">
13. Salivan
14. Noruzi

پنج عنصر منتخب، تنها تعدادی از عناصر ابر داده‌ای موجود در قالب ابر داده‌ای دابلین کور و زبان «اچ.تی.ام.ال» هستند. این عناصر بر بازنمون موضوعی مدرک متمرکزند و در برخی منابع، کاربرد آنها توصیه شده است (زانگ و دیمتروف، ۲۰۰۴؛ سالیوان، ۲۰۰۲؛ لی - اسملتزر<sup>۱</sup>، ۲۰۰۰، ص ۲۰۶؛ والکی، فریر<sup>۲</sup>، ۲۰۰۱، ص ۲۷۲).

### پرسشهای اساسی

پژوهش حاضر فاقد فرضیه است و پرسشهای اساسی آن بدین قرار است:

۱. حضور کلیدواژه مورد کاوش در هر یک از عناصر مورد بررسی (نشانه عنوان، ابر نشانه‌های کلیدواژه‌ها و توصیف زبان نشانه گذاری فرامتن، ابر داده‌های عنوان و موضوع قالب ابر داده‌ای دابلین کور) تا چه میزان بر رتبه‌بندی صفحات در سه موتور کاوش مورد آزمون اثر گذار است؟
۲. تکرار کلیدواژه مورد کاوش در دو تا پنج عنصر یاد شده در بخش سرآیند<sup>۳</sup> صفحه و صرف نظر از نوع عنصر، تا چه میزان بر بهبود رتبه صفحه اثر گذار است؟

### روش‌شناسی

پژوهشهایی که به تعیین میزان اثربخشی عناصر ابر داده‌ای بر رتبه‌بندی صفحات وب در موتورهای کاوش پرداخته‌اند، با تفاوت‌هایی اندک، از روش تجربی بهره برده‌اند. در این پژوهش نیز با آگاهی از وجود سایر عواملی که بر رتبه‌بندی صفحات اثر گذارند، جهت اعمال متغیرهای مستقل، کنترل متغیرهای دخیل، و مشاهده تغییرات در متغیر وابسته، از روش تجربی استفاده کرده‌ایم.

جامعه مورد پژوهش، نشریه ای الکترونیکی با ۸۴ صفحه وب است که پژوهشگر آن را طراحی کرده است. این نشریه با روندی که در ادامه می‌آید، طراحی شد و سپس در معرض نمایه سازی موتورهای کاوش قرار گرفت.

---

1. Lee-Smeltzer.  
2. Valqui & Freire.  
3. Header.

### ۱. مراحل طراحی صفحات

در مرحله نخست، تعداد چهارده مقاله در موضوعات وب معنایی<sup>۱</sup>، هستی‌شناختی<sup>۲</sup>، وب‌سنجی<sup>۳</sup> و ابر داده‌های دابلین کور از میان مقالات منتشر شده در نشریات رایگان حوزه کتابداری و اطلاع‌رسانی که در فهرست راهنمای نشریات پیوسته رایگان<sup>۴</sup> معرفی شده است، انتخاب گردید.

کلیدواژه‌های نمایه‌ای این مقالات به شیوه کنترل نشده (زبان طبیعی) و از بستر عنوان، کلیدواژه‌ها (در صورت وجود) چکیده، و در برخی موارد متن مقاله برگزیده شد. پژوهش بر دو دسته کلیدواژه متمرکز است: الف) کلیدواژه‌های منحصر به فرد و ب) کلیدواژه‌های تکرار شونده. کلیدواژه‌های منحصر به فرد آن دسته از کلیدواژه‌هایی هستند که منحصراً به یک عنصر ابر داده‌ای اختصاص یافته‌اند؛ با دیگر کلیدواژه‌های مربوط به سایر عناصر، همپوشانی و شباهت ندارند و پاسخگویی به نخستین پرسش را ممکن می‌سازند. کلیدواژه‌های تکرار شونده، کلیدواژه‌هایی هستند که در تمام عناصر ابر داده‌ای به یک شیوه و ترتیب، قبل از کلیدواژه‌های منحصر به فرد و به منظور سنجش اثربخشی تکرار کلیدواژه‌ها در دو تا پنج عنصر ابر داده‌ای (پرسش دوم) افزوده شده‌اند.

به طور نمونه، کلیدواژه‌های منحصر به فرد و تکرار شونده مقاله هشتم با عنوان «A Metadata Registry for the Semantic Web» در جدول ۱ آمده است.

جدول ۱. نمونه‌ای از کلیدواژه‌های منحصر به فرد و تکرار شونده

Element	Keywords	
	Unique	Repeated
E1 <sup>۵</sup>	---	Semantic web, Metadata Registry
E2	Metadata schemas, DCMI vocabulary	Semantic web, Metadata Registry
E3	Automated sharing of metadata, Dublin core metadata Initiative, DCMI Registry	Semantic web, Metadata Registry
E4	---	Semantic web, Metadata Registry
E5	Registry applications	Semantic web, Metadata Registry

1. Semantic web.

2. Ontology.

3. Webometrics.

4. Directory of Open Access Journals , Available at: [www.doaj.org](http://www.doaj.org).

۵. عنصر اول: عنوان اچ. تی. ام. ال

پس از تعیین کلیدواژه‌های مربوط به هر یک از عناصر مورد بررسی در مرحله بعد، با توجه به پرسشهای اساسی پژوهش، گروههای گواه و آزمون تعیین گردید. پاسخگویی به پرسشها با وجود شش گروه صفحه امکان‌پذیر است. گروه‌بندی صفحات در جدول ۲ آمده است. به منظور حفظ هم‌ترازی در دو گروه گواه و آزمون، تعداد صفحات گروه گواه با تعداد صفحات گروه آزمون، برابر است و محتوای صفحه آزمون در هر مقاله جز در مورد عناصر ابر داده‌ای افزوده شده، مشابه صفحه گروه گواه خود است. بدین ترتیب، چهارده مقاله منتخب در شش شماره نشریه تکرار شده اند (۸۴ = ۶ \* ۱۴).

جدول ۲. گروه‌بندی صفحات

گروه	حاوی کلیدواژه منحصر به فرد در	حاوی تکرار کلیدواژه‌ها در
اول	----	----
دوم	----	نشانه عنوان «اچ.تی.ام.ال» (عنصر ۱)
سوم	ابر نشانه کلیدواژه‌های «اچ.تی.ام.ال» (عنصر ۲)	نشانه عنوان و ابر نشانه کلیدواژه‌های «اچ.تی.ام.ال» (عناصر ۱ و ۲)
چهارم	ابر نشانه توصیف «اچ.تی.ام.ال» (عنصر ۳)	نشانه عنوان، ابر نشانه‌های کلیدواژه‌ها و توصیف «اچ.تی.ام.ال» (عناصر ۱، ۲، و ۳)
پنجم	----	نشانه عنوان، ابر نشانه‌های کلیدواژه‌ها و توصیف «اچ.تی.ام.ال»، ابر داده عنوان دابلین کور (عناصر ۱، ۲، ۳، و ۴)
ششم	ابر داده موضوع قالب ابر داده‌ای دابلین کور (عنصر ۵)	نشانه عنوان، ابر نشانه‌های کلیدواژه‌ها و توصیف «اچ.تی.ام.ال»، ابر داده‌های عنوان و موضوع دابلین کور (عناصر ۱، ۲، ۳، ۴، و ۵)

سپس کلیدواژه‌های منتخب در بستر نحوی «اچ.تی.ام.ال» نشانه گذاری و در کد منبع<sup>۱</sup> مقالات گنجانده شد. به طور نمونه، کد گذاری مقاله هشتم چنین است: در این مثال، کلیدواژه‌های تکرار شونده به صورت توپر نمایش داده شده است. گروه نخست، فاقد کلیدواژه است.

## در گروه دوم

<TITLE>A Metadata Registry for the Semantic Web</TITLE>

1. Source code.

**در گروه سوم**

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata
Registry, Metadata schemas, DCMI vocabulary">
```

**در گروه چهارم**

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata
Registry, Metadata schemas, DCMI vocabulary">
<META name="description" content="Semantic web, Metadata
Registry, Automated sharing of metadata, Dublin core metadata
Initiative, DCMI Registry">
```

**در گروه پنجم**

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata
Registry, Metadata schemas, DCMI vocabulary">
<META name="description" content="Semantic web, Metadata
Registry, Automated sharing of metadata, Dublin core metadata
Initiative, DCMI Registry">
<META name="DC.Title" content="A Metadata Registry for the
Semantic Web">
```

**در گروه ششم**

```
<TITLE>A Metadata Registry for the Semantic Web</TITLE>
<META name="keywords" content="Semantic web, Metadata
Registry, Metadata schemas, DCMI vocabulary">
<META name="description" content="Semantic web, Metadata
Registry, Automated sharing of metadata, Dublin core metadata
Initiative, DCMI Registry">
<META name="DC.Title" content="A Metadata Registry for the
Semantic Web">
<META name="DC.Subject" content="Semantic web, Metadata
Registry, Registry applications">
```

**۲. انتشار صفحات طراحی شده**

بدین ترتیب، ۸۴ صفحه ایستا<sup>۱</sup> به کمک واژه پرداز Word از مجموعه مایکروسافت آفیس ۲۰۰۳،<sup>۲</sup> طراحی و از ۲۴ خرداد تا ۸ شهریور ۱۳۸۵<sup>۳</sup> در دامنه فرعی «ابرداده» از سایت

1. Static.
2. Microsoft Office Word 2003.
3. June 14, 2006 till Aug 30, 2006.

«پژوهشگاه اطلاعات و مدارک علمی ایران»<sup>۱</sup> منتشر شد و در معرض نمایه سازی سه موتور کاوش گوگل<sup>۲</sup>، یاهو<sup>۳</sup>، ام اس ان<sup>۴</sup> - سه موتور کاوش پر استفاده به استناد «الکسا»<sup>۵</sup>، «سرچ اینجین و اج»<sup>۶</sup>، «سرچ اینجین گاید»<sup>۷</sup> و سایتهای مشابه<sup>۸</sup> - قرار گرفت. در دهمین روز انتشار صفحات، موتور کاوش گوگل، تمامی صفحات منتشر شده را نمایه سازی نمود؛ این در حالی است که یاهو روندی کندتر داشت و پس از گذشت بازه زمانی مشخص شده، ۷۶ صفحه مربوط به مقالات نشریه نمایه شد. اما از میان سه موتور کاوش انتخابی، «ام.اس.ان» به دلیل عدم نمایه سازی تعداد قابل قبولی از صفحات در مدت زمان تعیین شده، از مجموعه پژوهش حذف شد و ادامه پژوهش با تمرکز بر صفحات نمایه شده، در دو موتور کاوش یاهو و گوگل انجام پذیرفت.

### ۳. گردآوری داده‌ها

به منظور تعیین میزان اثربخشی عناصر ابر داده‌ای در این مرحله از پژوهش، کاوش کلیدواژه ای در کادر محاوره‌ای ساده<sup>۹</sup> دو موتور کاوش گوگل و یاهو انجام پذیرفت. از آنجا که دو گروه گواه و آزمون، در وب سایت طراحی شده گنجانده شده است و نیازی به تعیین رتبه صفحات در میان سایر صفحات موجود در وب نیست، کاوش کلیدواژه ای در دامنه وب سایت طراحی شده محدود گردید. فرمول کاوش در موتورهای کاوش گوگل و یاهو به ترتیب، چنین است:

Site:<http://metadata.irandoc.ac.ir> "Keywords"

Site:[metadata.irandoc.ac.ir](http://metadata.irandoc.ac.ir) "Keywords"

یا Domain:[metadata.irandoc.ac.ir](http://metadata.irandoc.ac.ir) "Keywords"

1. <http://metadata.irandoc.ac.ir>
2. [www.google.com](http://www.google.com)
3. [www.yahoo.com](http://www.yahoo.com)
4. [www.msn.com](http://www.msn.com)
5. Jun, 11, 2006 From: <http://www.alexa.com/site/ds/top-500>
6. [www.searchenginewatch.com](http://www.searchenginewatch.com)
7. [www.searchengineguid.com](http://www.searchengineguid.com)
8. <http://www.seoconsultants.com/search-engines/>
9. Simple Search Box.

بررسی میزان اثربخشی عناصر ابرداده‌ای بر ... / ۲۵۱

برای تعیین میزان اثربخشی هر یک از عناصر ابرداده‌ای (پرسش نخست پژوهش) دو مجرای آزمون وجود دارد:

• کاوش کلیدواژه‌های منحصر به فرد که تنها به عنصر ابرداده‌ای ویژه‌ای اختصاص یافته است و مقایسه رتبه صفحه حاوی آن کلیدواژه با صفحه گروه گواه مورد نظر.

• کاوش کلیدواژه‌های تکرار شونده؛ چنانکه بیان شد، به ترتیب افزایش شماره‌های نشریه، یک عنصر ابرداده‌ای به عناصر قبلی افزوده می‌گردد و تنها تفاوت میان دو مقاله مشابه در دو نشریه با شماره‌های متوالی، حضور کلیدواژه‌های تکرار شونده و منحصر به فردی است که در عنصر ابرداده‌ای بعدی افزوده شده است. لذا هر یک از شماره‌های نشریه در حالی که حاوی مقالات گروه آزمون نشریه شماره قبل خود است، حاوی مقالات گروه گواه نشریه شماره بعد نیز هست. بدین ترتیب، به استثنای عناصر عنوانی (عنوان اچ.تی.ام.ال و عنوان دابلین کور) دو دسته آزمون برای تعیین اثربخشی حضور هر یک از عناصر ابرداده‌ای (پرسش نخست) قابل انجام شدن است.

رتبه حاصل از هر یک از کاوشهای انجام شده در موتورهای کاوش، به تفکیک موتور کاوش و گروه گواه و آزمون با توجه به سؤالهای پژوهش در نرم‌افزار آماری SPSS<sup>1</sup> وارد شد و مورد تجزیه و تحلیل قرار گرفت.

با توجه به این که نتایج حاصل از کاوش در موتورهای کاوش از نوع داده‌های رتبه‌ای است (ترنر و برک بیل، ۱۹۹۸؛ صفری<sup>۲</sup>، ۲۰۰۵)، میزان معنادار بودن تفاوت دو گروه گواه و آزمون (دو گروه مستقل) در بازیابی صفحات وب، با استفاده از آزمون غیر پارامتری یومن وایتنی<sup>۳</sup> تعیین شد.

در مجموع، تحت شرایط زمانی و مکانی یکسان، ۳۶۳ کاوش در موتور کاوش گوگل و ۲۵۹ کاوش در موتور کاوش یاهو انجام شد و رتبه‌های کسب شده در دو گروه صفحات آزمون و گواه ثبت گردید. تجزیه و تحلیل آماری نیز بر مبنای داده‌های جمع‌آوری شده، با نرم‌افزار آماری «اس.پی.اس.اس.» انجام پذیرفت.

1. Statistical Package for Social Science.

2. Safari.

3. Mann-Whitney U.

## تجزیه و تحلیل یافته‌ها

۱. پاسخگویی به نخستین پرسش پژوهش: میزان اثربخشی هر یک از عناصر ابر داده‌ای بر رتبه‌بندی به وسیله دو موتور کاوش یاهو و گوگل.

برای درک بهتر و امکان مقایسه و تحلیل نتایج، جدول ۳ که نمایی کلی از مقدار پی محاسبه شده توسط نرم‌افزار آماری «اس.پی.اس.اس» در آزمون «یو من وایتنی» است، ارائه می‌شود. وجود تفاوت معنادار در هر عنصر، با قلم‌درشت تر و توپر نشان داده شده است. تفاوت میان میانگین رتبه‌های کسب شده در هر گروه نشانگر آن است که رتبه به نفع کدام گروه تغییر وضعیت داده است؛ لذا در هر مورد از عناصر، به تفکیک موتور کاوش، گروه صفحاتی که بهبود رتبه داشته (آزمون یا گواه) با اصطلاح «به نفع» مشخص شده است.

جدول ۳. نتایج آزمون «یو من وایتنی» در پاسخ به پرسش نخست

عنصر مورد آزمون	نوع کلیدواژه	گوگل		ياهو	
		به نفع	Asymp. Sig. (2-tailed)	به نفع	Asymp. Sig. (2-tailed)
عنوان اچ.تی.ام.ال	کلیدواژه منحصر به فرد	--	--	--	--
	کلیدواژه تکرار شونده	.000	آزمون	.001	آزمون
کلیدواژه‌های اچ.تی.ام.ال	کلیدواژه منحصر به فرد	.462	آزمون	.017	آزمون
	کلیدواژه تکرار شونده	.472	آزمون	.172	آزمون
توصیف اچ.تی.ام.ال	کلیدواژه منحصر به فرد	.015	آزمون	.000	آزمون
	کلیدواژه تکرار شونده	.016	گواه	.000	آزمون
عنوان دابلین کور	کلیدواژه منحصر به فرد	--	--	--	--
	کلیدواژه تکرار شونده	.826	آزمون	.807	آزمون
موضوع دابلین کور	کلیدواژه منحصر به فرد	.110	آزمون	.023	آزمون
	کلیدواژه تکرار شونده	.038	گواه	.050	گواه

یافته‌های حاصل از بررسی اثربخشی دو ابرنشانه کلیدواژه‌ها و توصیف در این پژوهش، با یافته‌های «ترنر» و «برک بیل» (۱۹۹۸) مبنی بر اثربخشی ابرنشانه کلیدواژه‌ها در مقایسه با ابرنشانه توصیف، تفاوت دارد. یافته‌ها در پژوهش حاضر نشان داد که ابرنشانه توصیف در مقایسه با ابرنشانه کلیدواژه‌ها اثر بیشتری در بهبود رتبه صفحات دارد.

بررسی میزان اثربخشی عناصر ابر داده‌ای بر ... / ۲۵۳

یافته‌های «صفری» (۲۰۰۵) مبنی بر عدم تأثیر عنصر عنوان دابلین کور بر رتبه صفحات، در این پژوهش نیز تأیید می‌شود. یافته‌های این پژوهش در راستای نتایج زانگ و دیمیتروف (۲۰۰۴) تأثیر عنصر موضوع دابلین کور را در بهبود رتبه صفحات نشان می‌دهد. این یافته با یافته‌های صفری مبنی بر عدم اثربخشی عنصر موضوع دابلین کور، متفاوت است.

۲. **پاسخگویی به پرسش دوم:** میزان اثربخشی دو تا پنج بار تکرار کلیدواژه در بخش سرآیند صفحات وب بر رتبه‌بندی صفحات وب در دو موتور کاوش یاهو و گوگل جدول شماره ۴، نمایی کلی از مقدار پی محاسبه شده توسط نرم‌افزار آماری «اس.پی.اس.اس» در آزمون «یومن وایتنی» ارائه می‌کند. در پاسخ به پرسش دوم پژوهش، تأثیر حضور کلیدواژه‌های تکرار شونده در دو تا پنج عنصر تحت بررسی، تجزیه و تحلیل شد.

جدول ۴. نتایج آزمون «یومن وایتنی» در پاسخ به پرسش دوم

تعداد تکرار مورد بررسی	گوگل		ياهو	
	Asymp.Sig.(2-tailed)	به نفع	Asymp.Sig.(2-tailed)	به نفع
۲ بار تکرار	۰/۵۸۰	آزمون	۰/۱۷۲	آزمون
۳ بار تکرار	۰/۰۰۸	گواه	۰/۰۱۶	آزمون
۴ بار تکرار	۰/۰۶۲	گواه	۰/۰۱۹	آزمون
۵ بار تکرار	۰/۰۰۰	گواه	۰/۱۷۵	آزمون

عملکرد دو موتور کاوش گوگل و یاهو در مقابل عناصر ابر داده‌ای در این پژوهش بررسی شد. به طور خلاصه، عملکرد دو موتور کاوش مورد بررسی، چنین است:

- موتورهای کاوش نسبت به حضور عناصر ابر داده‌ای مورد بررسی، بی‌اعتنا نیستند.
- عنصر عنوان «اچ.تی.ام.ال» در مقایسه با سایر عناصر مورد بررسی، اثرگذارترین عنصر در هر دو موتور کاوش مورد بررسی است.
- از میان ابر نشانه‌های «اچ.تی.ام.ال»، ابر نشانه توصیف، تأثیر بیشتری در مقایسه با ابر نشانه کلیدواژه‌ها دارد.
- توجه به عناصر ابر داده‌ای زبان «اچ.تی.ام.ال» بیش از عناصر ابر داده‌ای قالب دابلین کور است.

- توجه موتورهای کاوش مورد بررسی به عنصر موضوع قالب ابر داده‌ای دابلین کور، بیش از عنصر عنوان آن است.
- دو موتور کاوش گوگل و یاهو، عملکرد متفاوتی نسبت به کلیدواژه‌های تکرار شونده و منحصر به فرد دارند.
- در هر دو موتور کاوش، کلیدواژه‌های منحصر به فرد در مقایسه با کلیدواژه‌های تکرار شونده، تأثیر بیشتری دارد.
- موتور کاوش گوگل بیش از یاهو نسبت به کلیدواژه‌های تکراری در بخش سرآیند صفحات وب، حساس است و به نظر می‌رسد که این تکرارها را نوعی تقلب به شمار می‌آورد.

### نتیجه‌گیری

با وجود تمامی پیشرفتهای حاصل شده در فناوری موتورهای کاوش، هنوز روش اصلی نمایه سازی در موتورهای کاوش، محتوا محور است و بر مبنای کلیدواژه‌ها انجام می‌پذیرد (یو و دیگران<sup>۱</sup>، ۲۰۰۲). نمایه سازی تمام متن، کاوش کلیدواژه‌ای بدون توجه به بافت متن و کاستیهای موجود در رتبه‌بندیها، به جامعیت بالا و مانعیت پایین منجر می‌شود (لی - اسملتزر، ۲۰۰۰، ص ۲۰۶). تحقیقات فراوانی در زمینه مشکلات نمایه‌سازی خودکار در موتورهای کاوش انجام گرفته است (چانگ و لی<sup>۲</sup>، ۱۹۹۸، ص ۱۴۹). آنچه پژوهشگران در باب آن به اتفاق نظر رسیده‌اند آن است که نمایه‌سازی غیرمتمرکز و توزیع شده، به گونه‌ای که در موتورهای کاوش کنونی در جریان است، معماری قابل تأییدی در بازیابی اطلاعات وب نیست (لی - اسملتزر، ۲۰۰۰، ص ۲۰۶).

چنانچه محتوای صفحه به درستی در فیله‌های ابر داده‌ای باز نموده شود و موتورهای کاوش از این فیله‌های ابر داده‌ای در رتبه‌بندی و بازیابی نتایج بهره ببرند، مانعیت افزایش و

---

1. Yu et al.  
2. Chung & Lee.

\_\_\_\_\_ بررسی میزان اثربخشی عناصر ابر داده‌ای بر ... / ۲۵۵

نتایج بازیابی به سطحی قابل مدیریت کاهش می‌یابد. به علاوه، رتبه‌بندی صفحات صحت بیشتری خواهد داشت (زانگ و جاستریم، ۲۰۰۵، ص ۱) و آن بخش از ویژگیها که شناسایی آنها به درک انسانی نیاز دارد، در فرایند نمایه‌سازی منظور خواهد شد؛ اما همچنان پاسخ به این سؤال که آیا جامعه بهره‌گیر، به اهمیت وجود ابر داده پی برده است یا خیر، آن را تأیید می‌کند و یا اینکه به درستی از آن بهره می‌برد یا نه، یک معماست (زانگ و جاستریم، ۲۰۰۵، ص ۲۲).

از نظر «تیم برنرلی»<sup>۱</sup> و سایر پیشگامان وب معنایی، وب کنونی، اطلاعاتی انسان فهم دارد. صفحات وب به زبان اچ.تی.ام.ال نگاشته می‌شود و مرورگرها بر نشانه‌هایی تمرکز دارند که نه بر محتوای معنایی، بلکه بر ساختار متمرکزند. چنانچه وب معنایی تحقق یابد، صفحات وب آینده همانگونه که انسان فهم هستند، ماشین فهم<sup>۲</sup> نیز خواهند بود (کمپل<sup>۳</sup>، ۲۰۰۴، ص ۱۹۷)، در این شیوه اطلاعات از طریق نشانه‌های معنادار ارائه می‌شود و ابر داده‌های لازم را «آر.دی.اف» به صفحات وب می‌افزاید. پدیدآورندگان وب فردا، امکان جستجوپذیر کردن سایتهای خود را با مانعیت بیشتری عملی خواهند ساخت و کارگزاران هوشمند<sup>۴</sup> آینده، بسیار قدرتمندتر و متکامل‌تر از موتورهای کاوش امروزی عمل خواهند کرد (کمپل، ۲۰۰۴، ص ۱۹۹).

### منابع

-Asadi, S., & Jamali M., H.R. (2004). "Shifts in search engine development: A review of past, present and future trends in research on search engines". *Webology*, 1(2). Retrieved Feb, 18, 2005, From <http://www.webology.ir/2004/v1n2/a6.html>

- 
1. Tim Burners Lee.
  2. Machine understandable.
  3. Campbell.
  4. Intelligent Agents.

-Bar-Ilan, J. (2005). "Comparing Rankings of Search Results on the Web". *Information Processing & Management*, 41. Retrieved Oct, 12, 2005, From Elsevier Database.

-Campbell, D.G. (2004). "The Metadata – Bibliographic Organization nexus" In *Metadata Applications and Management*. London: Facet Publishing, 185-203.

-Chung, S.M. & Lee, J.Y. (1998). "Information Discovery on the Internet". In *Encyclopedia of Library and Information Science*. vol. 62, (supple. 25) pp.146-157.

-Drott, M.C. (2002). "Indexing aids at Corporate websites: the use of robot.txt and meta tags". *Information Processing and Management*, 38(2), 209-219. Retrieved June 10, 2006, From Elsevier database.

-Fadaie, G. (2005). "User Satisfaction through better indexing". *Cataloging & Classification Quarterly*, 40(2), 5-17.

-Fishkin, R. (2005). "Search Engine Ranking Factors" Retrieved Dec, 17, 2005, From <http://www.seomoz.org/articles/search-ranking-factors.php#1>

-Global Top 500 (2006) Retrieved Jun, 11, 2006, From [http://www.alexa.com/site/ds/top\\_500](http://www.alexa.com/site/ds/top_500)

-Google (2004). "Information for webmasters". Retrieved Oct, 11, 2005, From <http://www.google.com/webmasters/4.html>

-Gotlieb, C. ; Eliopoulos, D. (2003). "Evaluating Web Search Results Ranking". *Online*, 27(2). Retrieved Oct, 11, 2005, From Ebsco Computer Science Database.

-Henshaw, R. & Valauskas, E.J. (2001). "Metadata as a catalyst: experiments with metadata and search engines in the internet journal, *First Monday*". *Libri*, 51(2), 86-101pp. Retrieved Oct, 23, 2005, From [www.librijournal.org/pdf/1999-3pp125-131.pdf](http://www.librijournal.org/pdf/1999-3pp125-131.pdf)

-Jansen, B.J. ; Spink, A.; and Saracevic, T. (2000). " Real life, real users, and real needs: a study and analysis of user queries on the web" . Information Processing & Management, 36(2), 207-227. Retrieve May, 8, 2005, From Elsevier Database.

-Lee-Smeltzer, K.H. (2000). " Finding the needle: Controlled Vocabularies, Resource Discovery, and Dublin core". Library collections, Acquisitions, & Technical Services, 24 , 205-215. Retrieve May, 8, 2005, From Elsevier Database.

-Lewis , E. (2005). Top Ten Search Engines - Top 10 SEs . Retrieved Apr, 22, 2006, From <http://www.seoconsultants.com/search-engines/>

-List of Top Search Engines (2006). Retrieved Apr, 22, 2006, From <http://capmex.biz/resources/top-search-engines>

-Noruzi, A. (2005). "Editorial" .Webology, 2(4). Retrieved Apr, 20, 2006, From <http://www.webology.ir/2005/v2n4/editorial6.html>

-Quevedo\_Torrero, J.U. (2004). " Improving Web Retrieval by Mining the HTML tags for Keywords and Exploring the Hyperlink Structures of Web Pages" [Abstract] doctoral Dissertation. University of Houston. Retrieved Oct, 10, 2005, From <http://www.lib.umi.com/dissertations/fullcit/3156028>

-Safari, M. (2005). "Search Engines and Resource Discovery on the web: Is Dublin Core an Impact Factor?". Webology, 2(2). Retrieved Apr, 20, 2006, From <http://www.webology.ir/2005/v2n2/a13.html>

-Sullivan, D. (2002). "How to Use HTML Meta tags". Retrieved Apr, 20, 2006, From <http://searchenginewatch.com/webmasters/article.php/2167931>

-Thelwall, M.; Vaughan, L. (2004). " New Versions of PageRank employing alternative web document models" . ASLIB Proceedings, 56(1), 24-33. Retrieved Feb, 19, 2005, From [www.scit.wlv.ac.uk/~cm1993/papers/2004\\_new\\_pagerank\\_preprint.pdf](http://www.scit.wlv.ac.uk/~cm1993/papers/2004_new_pagerank_preprint.pdf)

-Turner, T.P.; Brackbill, L. (1998). "Rising to the Top: Evaluating the use of html meta tag to improve retrieval of world wide web documents through internet search engines" . Library Resources and Technical Services, 24(4), 258-271.

-Yaltaghian , B. (2002). "Improving the ranking of Search Engine Output: A Network Analysis Approach". [Abstract] doctoral Dissertation. University of Toronto. Retrieved Feb, 21, 2006, From [www.cosc.canterbury.ac.nz/open/seminars/old/2002.html](http://www.cosc.canterbury.ac.nz/open/seminars/old/2002.html)

-Yu et al (2002). "Patterns in Unstructured Data: Discovery, Aggregation, and Visualization". Retrieved Jan, 21, 2005, From [http://javelina.cet.middlebury.edu/lisa/out/cover\\_page.htm](http://javelina.cet.middlebury.edu/lisa/out/cover_page.htm)

-Zhang, Z. ; Dimitroff, A. (2005b). " The Impact of Metadata Implementation on Webpage Visibility in Search Engine Results (Part II)" . Information Processing & Management, 41 (3), 691-715. Retrieved Feb, 18, 2005, From Elsevier Database.

-Zhang, Z. ; Dimitroff, A. (2004). "Internet Search engines response to metadata Dublin core implementation" . Journal of Information Science , 30(4), 310-320. Retrieved Oct, 11, 2005, From Elsevier Database.

-Zhang, Z. ; Dimitroff, A. (2005a). " The Impact of Webpage Content Characteristics on Webpage Visibility in Search Engine Results (Part I)" . Information Processing & Management, 41 (3), 665-690. Retrieved Feb, 18, 2005, From Elsevier Database.

-Zhang, Z. ; Jastram, I. (2005). " A Study of the metadata creation behavior of different user groups on the Internet" Information Processing and management. Retrieved Oct, 11, 2005, From Elsevier Database.