

طراحی موتور جستجوی هوشمند تخصصی

سیدرضا سرآبادانی*

کارشناس ارشد مهندسی فناوری اطلاعات،
دانشگاه صنعتی مالک اشتر

حسین شیرازی^۱

دانشیار،

دانشگاه صنعتی مالک اشتر

دریافت: ۱۳۸۹/۰۱/۲۱ | پذیرش: ۱۳۸۹/۰۶/۲۱

فصلنامه علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا(چاپی) ۱۷۳۵-۵۲۰۶
شاپا(الکترونیکی) ۲۰۰۸-۵۵۸۳
نمایه در LISA، SCOPUS و ISC
http://jst.irandoc.ac.ir
دوره ۲۶ | شماره ۴ | صص ۹۰۳-۹۲۴
تابستان ۱۳۹۰
نوع مقاله: پژوهشی

چکیده: امروزه موتورهای جستجوی شناخته شده و پر کاربرد زیادی در سطح دنیا مطرح هستند. این موتورهای جستجو نمی توانند به خوبی نیاز کاربران را در ارائه اطلاعات تخصصی برآورده نمایند. در این مقاله یک "موتور جستجوی هوشمند تخصصی" طراحی و معرفی شده است. همچنین، معماری سیستم و روش کار و اجزای تشکیل دهنده این موتور جستجو و ویژگی های هوشمند آن تشریح گردیده است. برای این ربات هوشمند، یک "فهرست راهنما" طراحی شده است که هدف آن، ساخت پایگاه دانش تخصصی به منظور دسترسی آسان و سریع کاربران به اطلاعات آن است. ابتدا تمام کلیدواژه های تخصصی در این فهرست تعریف شدند و بر اساس آن سایت ها نمایه سازی شدند. ۸۶ درصد از جامعه آماری مورد آزمایش معتقد بودند که زمان صرف شده برای جستجوی اطلاعات در این موتور جستجو کمتر از سایر موتورهای جستجوی عمومی شناخته شده است، همچنین ۸۰ درصد معتقد بودند که دقت و کیفیت نتایج جستجو در اینجا بیشتر از سایر موتورهای جستجوی عمومی بوده است.

کلیدواژه ها: موتورهای جستجو، ربات جستجوگر، فهرست راهنما، موتورهای جستجوی هوشمند، موتورهای جستجوی هوشمند تخصصی

*rezas1388@yahoo.com
1. hossein_shirazi@yahoo.com

۱. مقدمه

امروزه، دسترسی به اطلاعات تخصصی برای سازمان‌های علمی و تحقیقاتی بسیار مهم و حیاتی است. به‌عنوان مثال، در یک سازمان نظامی، دسترسی به آخرین دستاوردهای موشکی و یا مشخصات فنی موتور موشک X و یا موارد اینچنینی به‌آسانی از طریق گوگل امکان‌پذیر نیست، چون در حال حاضر موتورهای جستجوی فعلی آنقدر هوشمند نیستند که بخواهند ارزش محتوای اطلاعات را بفهمند و تجزیه و تحلیل کنند. بنابراین، گاهی اوقات ارزش برخی اطلاعات موجود در اینترنت برای موتورهای جستجو به‌دلیل استفاده از الگوریتم‌های مبتنی بر تحلیل پیوند، بسیار پایین است (Allan et al. 2005)، ولی همان اطلاعات ممکن است از دید یک سازمان علمی بسیار مهم و ارزشمند باشد. این مسأله باعث می‌شود تا اطلاعات تخصصی به‌سختی در موتورهای جستجوی امروزی پیدا شود و در بیشتر مواقع، اطلاعات غیرمرتبط نمایش داده می‌شود. حتی با ظهور نسل جدید موتورهای جستجوی گول‌پیکر امروزی مانند گوگل^۱ و یاهو^۲ که مخاطبان زیادی را به‌خود اختصاص داده است، همچنان مشکلات اساسی در این موتورهای جستجو به‌قوت خود باقی است. برخی از مشکلات مهم و اساسی موتورهای جستجوی فعلی در خصوص دسترسی به اطلاعات تخصصی به‌شرح ذیل است:

- **پایین بودن دقت اطلاعات.** یکی از مشکلات مهم و اساسی کاربران، دستیابی به اطلاعات تخصصی است. گاهی، کلیدواژه‌های وارد شده در موتورهای جستجو مربوط به یک حوزه دانشی خاص است. بنابراین موتورهای جستجوی جاری با مراجعه به بانک نمایه^۳ خود سعی می‌کنند اطلاعات مورد نظر کاربر را به نمایش بگذارند، اما از آنجا که این اطلاعات به‌دلیل تخصصی بودن وجود ندارد، بنابراین موتور جستجو اطلاعات مشابهی که با درخواست کاربر مغایرت دارد نمایش می‌دهد که این از مشکلات فعلی موتورهای جستجو است (Yajun Du et al. 2004).

- **عدم رضایت کاربران در دسترسی به اطلاعات تخصصی.** به‌عنوان مثال، بیشتر موتورهای جستجو فقط قادر به نمایه‌سازی بخش بسیار کوچکی از صفحات وب موجود در شبکه جهانی هستند. بیشتر روش‌های نمایه‌سازی موجود فقط بر ابرنشان‌های^۴ درون مدارک استوار است؛ دوره روزآمدسازی در آنها به‌نسبت طولانی است؛ ... و به‌احتمال همه این مسائل به‌بازیابی ناقص منابع منجر می‌شود و کاربران فقط بخش بسیار کوچکی از اطلاعات مورد نظرشان را دریافت می‌کنند (Lawrence et al. 1999 quoted in Isfandyari Moghaddam 2007).

1. Google

2. Yahoo

3. Index

4. Metatag

• **زمان صرف شده بسیار بالا برای جستجوی اطلاعات در اینترنت.** امروزه، اگر یک کاربر

اینترنت بخواهد به اطلاعات تخصصی دسترسی پیدا کند علاوه بر نبود اطلاعات مناسب به دلیل مسائل بیان شده باید ساعت ها وقت صرف کند تا به برخی اطلاعات دسترسی پیدا کند. در آخر هم به دلیل ماهیت موتورهای جستجوی فعلی، اطلاعات به دست آمده در بعضی موارد فاقد کیفیت و غیر قابل استفاده است (Yajun Du et al. 2004).

برای رفع مشکلات اساسی مطرح شده و با توجه به جدید بودن تحقیق، طراحی و ساخت موتورهای جستجوی تخصصی برای سازمان ها و مراکز علمی بسیار لازم و ضروری است. بنابراین، "طراحی موتور جستجوی هوشمند تخصصی" به عنوان یک پروژه پایان نامه کارشناسی ارشد در دانشگاه صنعتی مالک اشتر تعریف گردید. پس از یک سال تلاش و کوشش، سرانجام این موتور جستجو با موفقیت مورد آزمون قرار گرفت که نتایج آن در این مقاله ارائه گردیده است.

در این مقاله، جزئیات پیاده سازی "موتور جستجوی هوشمند تخصصی" که برای مجموعه دفاع کشور طراحی شده، بیان گردیده است. همچنین، معماری سیستم و روش کار و اجزای تشکیل دهنده این موتور جستجو شرح داده شده است. این موتور جستجو به اختصار SISE¹ نامیده شده است. ساختار موتور جستجوی هوشمند تخصصی با ربات های جستجوگر موجود همچون گوگل تا حدودی متفاوت است. SISE ابتدا کلیدواژه های نظامی را از برنامه فهرست راهنما² که برای آن طراحی شده است دریافت و بر اساس آن اطلاعات موجود سایت ها را نمایه سازی می کند. بنابراین، اطلاعات زاید و اضافی خارج از کلیدواژه های تعریف شده در این پایگاه قرار نمی گیرد. این ربات جستجوگر، به طور کامل، تخصصی نظامی است.

۲. فرضیه ها و سؤال های پژوهش

فرضیه های این تحقیق عبارتند از:

۱. استفاده از SISE، سرعت و دقت نتایج جستجو توسط متخصصان دفاع را افزایش می دهد.
۲. استفاده از فهرست راهنما در SISE، باعث بهبود کیفی و کمی جستجو توسط متخصصان می گردد.

در همین راستا، سؤالات تحقیق به شرح ذیل مطرح می شود. سؤالات تحقیق در واقع مسایلی هستند که محقق در طول انجام تحقیق به دنبال پاسخگویی به آنهاست. سؤالات ذیل به عنوان سؤالات اصلی تحقیق مد نظر قرار گرفته اند.

۱. ویژگی های یک موتور جستجوگر هوشمند چیست؟
۲. یک موتور جستجوگر هوشمند تخصصی باید دارای چه ویژگی هایی باشد؟

1. Special Intelligent Search Engine

2. directory

۳. اهداف پژوهش

۱. طراحی SISE به منظور کمک به کارشناسان و متخصصان مجموعه دفاع کشور.
۲. دسترسی به اطلاعات تخصصی نظامی در زمان کم و با دقت و کیفیت بالای اطلاعات با استفاده از SISE.

۴. پیشینه پژوهش

در اینجا پژوهش‌های انجام شده در خصوص موتورهای جستجو، در داخل و خارج کشور بررسی می‌شود. برای بررسی دقیق پیشینه پژوهش، موتورهای جستجوی طراحی شده در دنیا با توجه به ماهیت آن به دو بخش اصلی موتورهای جستجوی هوشمند عمومی و موتورهای جستجوی هوشمند تخصصی تقسیم بندی شده‌اند. منظور از موتور جستجوی عمومی، موتور جستجویی است که اطلاعات کلی از مجموع علوم مختلف را در خود ذخیره نماید. این اطلاعات می‌تواند ورزشی، اقتصادی، اخبار، فیلم، موزیک و ... باشد. منظور از موتور جستجوی اختصاصی، موتور جستجویی است که خاص یک نوع اطلاعات یا علم باشد؛ به عنوان مثال، موتور جستجوی ورزشی، موتور جستجوی نظامی، موتور جستجوی پزشکی، موتور جستجوی فیزیک، و موتور جستجوی شیمی.

۴-۱. پیشینه موتورهای جستجوی هوشمند عمومی

سال ۲۰۰۱، برند اولین ربات جستجوگر هوشمند را به نام ربات جستجوگر "هوش مصنوعی" معرفی کردند. این ربات هوشمند با استفاده از الگوریتم‌های هوش مصنوعی اطلاعات موجود سایت‌ها را نمایه می‌کرد و در بانک اطلاعاتی خود ذخیره می‌نمود (Thomas and Wolff 2001). سال ۲۰۰۲، پنت و منزر ربات هوشمند خود را به نام "وب کرالر" مطرح نمودند (Pant and Menczer 2002) برای بالا بردن سرعت ربات جستجوگر هوشمند، کینیاک و سوئل به توسعه آن پرداختند (Shkapenyuk and Suel 2002). در ادامه، ربات "هاپفیلد نت" در شبکه‌های عصبی توسعه پیدا کرد (Song et al. 2002) که الگوریتم این ربات‌ها در بعضی موارد در موتورهای جستجوی گوگل و یاهو مورد استفاده قرار گرفت.

علاوه بر موارد اشاره شده، از سال ۲۰۰۲ به بعد، موتورهای جستجوی متن باز^۴ بسیاری در سراسر دنیا به عنوان پروژه‌های تحقیقاتی و علمی دانشجویی تولید شده است که بعضی بسیار موفق بوده‌اند. از جمله، دو موتور جستجوی متن باز "ناچ"^۵ که با زبان برنامه‌نویسی "جاوا"^۶ و "اسفیدر

1. AI spider 2. Web Crawlers 3. Hopfield net 4. Open source
5. Nutch (<http://lucene.apache.org/nutch>) 6. Java

پلاس^۱ که با زبان برنامه نویسی "پی.اچ.پی."^۲ طراحی شده‌اند. این دو موتور جستجو، مقالات و مطالب علمی بسیاری را در اینترنت به خود اختصاص داده‌اند، به‌خصوص موتور جستجوی "ناچ" که قرار بود جایگزین گوگل شود ولی به دلیل مشکلات مالی این امر عملی نشد. در قسمت روش تحقیق، به‌طور کامل ویژگی‌های ۱۰ موتور جستجوی متن باز بررسی شده شرح داده می‌شود.

۴-۲. پیشینه موتورهای جستجوی هوشمند تخصصی

مطالعات انجام شده نشان می‌دهد که پژوهش‌های اندکی در جهت ساخت موتورهای جستجوی هوشمند تخصصی انجام شده است که دلیل این امر خود می‌تواند به‌عنوان یک تحقیق توسط محققان بررسی گردد. بیشتر موتورهای جستجوی بیان شده مانند گوگل و یاهو جنبه عمومی دارند و تخصصی نیستند. به‌عنوان نمونه از موتورهای جستجوی تخصصی طراحی شده خارجی می‌توان به موتور جستجوی "اسپورتیو"^۳ اشاره نمود. این موتور جستجو که یک موتور جستجوی ورزشی است به‌عنوان یک کار تحقیقاتی توسط "گاک‌ود، نینگلی، و زوماتایو"^۴ در سال ۲۰۰۶ تهیه شده است.

مبنای اصلی این موتور جستجو از موتور جستجوی متن باز "اسفیدر"^۵ شکل گرفته است. در واقع، با نصب و راه‌اندازی موتور جستجوی "اسفیدر" و قرار دادن نشانی سایت‌های ورزشی در اختیار این ربات، کلیه اطلاعات موجود در وب‌سایت‌های ورزشی نمایه می‌شود. مشکل این موتور جستجو این است که علاوه بر ذخیره‌سازی اطلاعات ورزشی سایت‌ها، اطلاعات غیرورزشی زیادی در بانک اطلاعاتی این موتور جستجو ذخیره می‌شود. ولی، در مجموع نسبت به موتورهای جستجوی عمومی تخصصی‌تر است.

از کارهای مشابه داخلی می‌توان به موتور جستجوی ریسمون^۶ محصول شرکت رادکام اشاره کرد. این جستجوگر همه وب‌سایت‌های منبع را در دوره‌های زمانی یک ماهه می‌پیماید و محتویات و مطالب آنها را نمایه‌سازی می‌کند و در بانک اطلاعاتی خود جهت ارائه خدمات جستجو به مراجعه‌کنندگان، نگهداری می‌نماید. در ریسمون می‌توان جستجوهای تخصصی ایجاد کرد. هم‌اکنون بخش جستجوی "شعر و ادب" در ریسمون به‌عنوان یک خروجی تخصصی جستجو عمل می‌نماید.

پس از بررسی پیشینه موتورهای جستجوی هوشمند، ویژگی‌های آنها در جهت پاسخ به سؤال اول تحقیق به شرح ذیل ارائه می‌گردد.

-
1. Sphider plus (<http://www.sphider.eu/>)
 2. PHP
 3. Sportivo
برای کسب اطلاعات بیشتر فایل مربوط را از نشانی <http://the-designing.com/contents/paper/sportivo.pdf> دانلود نمایید.
 4. Gaikwad, Ningli, and Zhumatayev
 5. Sphider
 6. www.rismoon.com

ویژگی‌های مهم یک ربات جستجوگر هوشمند شامل موارد ذیل است (Yajun Du et al. 2004):

۱. به‌طور هوشمندانه نیازمندی‌های کاربران را نگهداری نماید.
کلیدواژه‌های جستجو شده به‌طور معمول ارتباط نزدیکی با سابقه دانشی کاربر دارد. بنابراین، یک ربات جستجوگر هوشمند باید در هنگام فرآیند نمایه‌سازی اطلاعات از صفحات وب، به سابقه دانشی کاربران اهمیت دهد و اطلاعات مربوط را در خود نگهداری نماید تا بتواند نیاز کاربر را به‌خوبی پاسخگو باشد.
 ۲. کاربر زمان بسیار کمی را صرف جستجو در اینترنت نماید.
نتایج جستجو باید نیاز کاربران را برطرف نماید. در حال حاضر، بسیاری از موتورهای جستجوی فعلی قادر نیستند اطلاعات را در کمترین، کامل‌ترین، دقیق‌ترین، و بلا درنگ در اختیار کاربران قرار دهند.
 ۳. به‌طور لحظه‌ای و آنی سوابق نیازمندی کاربران را با لغات کلیدی درخواست شده پاسخ دهد.
برای اینکه موتور جستجو بتواند نیازمندی‌های کاربران را با لغات کلیدی درخواست شده به‌طور لحظه‌ای و آنی پاسخ دهد، باید برای هر کاربر فضای اطلاعاتی خاص در اینترنت ایجاد شود. برای ایجاد فضای اطلاعاتی خاص در اینترنت می‌باید درخت علاقه‌مندی^۱ ایجاد شود. چون ربات جستجو می‌تواند با پیمایش در این درخت، اطلاعات مورد نظر را به کاربر ارسال نماید (Yajun Du et al. 2004).
- از آنجاکه هوشمندی امری نسبی است، در طراحی SISE موارد دوم و تا حدودی سوم برای این پروژه مد نظر قرار گرفته است و مورد اول می‌تواند در پژوهش‌های آتی توسط سایر محققان پیاده‌سازی گردد. در مجموع، از بررسی پژوهش‌های انجام شده پیداست که مجال تحقیق و کنکاش در زمینه موتورهای جستجوی تخصصی وجود دارد. در ادامه جزئیات طراحی و پیاده‌سازی SISE شرح داده می‌شود.

۵. روش پژوهش

۵-۱. نوع و روش پژوهش

پژوهش حاضر از نوع کاربردی است که با ترکیبی از روش‌های تحقیقی کتابخانه‌ای و پیمایشی انجام شده است. ابزارهایی که در روش کتابخانه‌ای استفاده شده‌اند عبارتند از منابع الکترونیکی و نشریات و مقالات در موضوع موتورهای جستجو که به‌نوعی مرتبط با موضوع تحقیق بوده‌اند. ابزارهای مورد استفاده در روش پیمایش، پس از مشاهده و تهیه نرم‌افزار موتور جستجو، پرسشنامه بوده است.

1. generate the interest tree

در واقع در این پژوهش دو گام تحقیقاتی انجام شده است که به ترتیب عبارتند از: ۱- طراحی نرم افزار SISE؛ ۲- ارزیابی نرم افزار SISE. بنابراین در ابتدا طراحی نرم افزار SISE و ویژگی های آن شرح داده شده و سپس برای ارزیابی موتور جستجوی طراحی شده، به منظور اثبات فرضیه های تحقیق از پرسشنامه و جامعه آماری استفاده شده است.

۲-۵. گام اول: طراحی نرم افزار SISE

پس از مطالعات میدانی بر روی چندین موتور جستجوی متن باز مربوط به سال های ۲۰۰۵ به بعد^۱، شروع پروژه SISE با استفاده از "ناچ"^۲ که یک موتور جستجوی متن باز مبتنی بر جاواست، آغاز شد. نصب این برنامه و بررسی نتایج آن زمان بر بود. با وجود قدرتمند بودن این نرم افزار، تغییر کدهای آن ساده نبود، بنابراین استفاده از PHP crawler که یک ابزار متن باز مبتنی بر PHP & MySQL است و بررسی و تغییر کدهای آن آسان است، در دستور کار قرار گرفت. در هنگام استفاده از PHP crawler این نتیجه حاصل شد که این ابزار از ثبات کافی برخوردار نیست و اشکالات زیادی دارد. گاهی اوقات نیز در هنگام پیمایش^۳ با خطا روبرو می شد. اگرچه نسخه جدید PHP Crawler ارتقا یافته و از قابلیت نمایه سازی برخوردار است، همچنان از کارایی لازم برخوردار نیست و مکانیسم مرتب سازی^۴ اطلاعات و رتبه بندی را شامل نمی شود. در ادامه از یک ابزار متن باز دیگر به نام Sphider-plus استفاده شد. این ابزار، مبتنی بر PHP & MySQL است؛ نصب و استفاده از آن آسان است؛ قدرتش نیز از PHP Crawler بیشتر است. مهمترین نکته در مورد Sphider-plus قابلیت نمایه^۵، رتبه بندی^۶، مرتب نمودن، و به روز بودن آن است، در نتیجه Sphider-plus به عنوان ابزار اولیه پروژه SISE انتخاب شد.

۲-۵-۱. جزئیات پیاده سازی پژوهشگاه علوم انسانی و مطالعات فرهنگی

- پس از انتخاب Sphider-plus به عنوان یک ابزار اولیه جهت تهیه نرم افزار SISE، در ابتدا، این برنامه بر روی یک کامپیوتر پنتیوم طبق مراحل زیر نصب و راه اندازی شد:
- برنامه Wamp را بر روی یک کامپیوتر معمولی پنتیوم نصب نمایید. این برنامه تحت سیستم عامل ویندوز اجرا می شود. با نصب این برنامه نرم افزارهای PHP و Apache و MySQL بر روی کامپیوتر نصب می گردد.
 - اگر برنامه در مسیر (C:\) نصب شده است، باید برنامه Sphider در مسیر C:\wamp\www کپی شود.

1. Datapark, Sphider Plus, PHP crawler, Indri, Nutch, MG4J, Namazu, Omega, SWISH-E, SWISH++, Terrier, Zettair. 2. Nutch 3. Crawling
4. Sort 5. Index 6. Ranking

از مسیر `C:\wamp\www\sphider-plus\settings`، فایل `database.php` باز و موارد ذیل تنظیم

می شود:

```
$database
$mysql_user
$mysql_password
$mysql_host
```

● برنامه `install_all.php` از مسیر `http://localhost/sphider-plus/admin/install_all.php` اجرا می شود.^۱

۲-۲-۵. اجرای برنامه Sphider

از مسیر `./admin/admin.php` به بخش مدیریت نرم افزار وارد و سایت `designation-systems` جهت نمایه به موتور جستجوی Sphider معرفی می شود. ربات جستجوگر شروع به پیمایش پیوندهای موجود در سایت مورد نظر می نماید و صفحات پیمایش شده را نمایه و در بانک اطلاعاتی خود ذخیره می نماید.

۲-۲-۳. خلاصه نتیجه بررسی Sphider

پس از راه اندازی برنامه Sphider بر روی یک کامپیوتر شخصی و آزمون این برنامه جهت نمایه سازی یک سایت نظامی، وضعیت اطلاعات ذخیره شده در بانک اطلاعاتی این برنامه در جدول ۱ بیان شده است.

جدول ۱. نمایش آمار ۶۰ کلیدواژه با اهمیت

موضوع	نظامی	غیر نظامی
تعداد لغات	۷	۵۳
درصد پوشش اطلاعات	٪۱۲	٪۸۸

در جدول ۱، منظور از کلیدواژه های با اهمیت، کلیدواژه هایی است که دارای معنی بوده و بیشترین تکرار را در صفحات وب سایت داشته اند. بنابراین، به طور تصادفی ۶۰ کلیدواژه ذخیره شده و با اهمیت بررسی شدند. مشاهده اطلاعات جدول نشان می دهد که بیشترین لغات کلیدی ذخیره شده از یک سایت نظامی نمایه شده، غیر نظامی است. همین مسأله باعث می شود در خروجی نتایج جستجو برای کاربران، اطلاعات غیر مرتبط زیادی نمایش داده شود، بنابراین برنامه Sphider

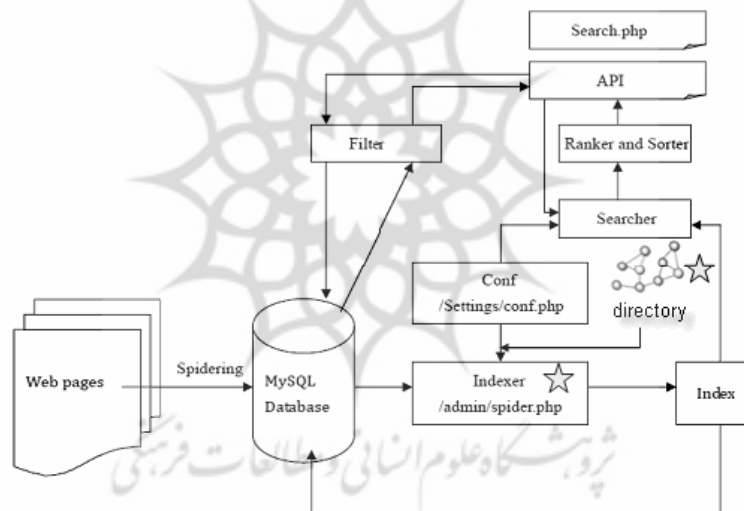
۱. جهت کسب اطلاعات بیشتر به سایت <http://www.sphider.eu> مراجعه نمایید.

به تنهایی برای SISE مناسب نیست. برای رفع مشکلات بیان شده تغییراتی در معماری Sphider به وجود آورده شد و این معماری برای SISE در نظر گرفته شد که به شرح ذیل ارائه می گردد.

۴-۲-۵. طراحی معماری SISE

همان طور که در شکل ۱ ملاحظه می شود دو تغییر اساسی در معماری SISE نسبت به معماری موتورهای جستجوی عمومی از جمله Sphider وجود دارد. این تغییرات با علامت ستاره مشخص شده است که عبارتند از (۱) تغییر روش نمایه سازی^۱ در Spider؛ (۲) طراحی فهرست راهنمای تخصصی. در ادامه، اجزای اصلی این معماری شرح داده می شود:

MySQL: به عنوان پایگاه اطلاعاتی، برای ذخیره تمامی اطلاعات از جمله اسناد و مدارک پیمایش شده، ایندکس کلیدواژه ها، لغات جستجو شده، و پیوندهای پیمایش شده، استفاده می شود.



شکل ۱. معماری SISE

Indexer: به پویش مدارک می پردازد و آنها را به صورت کلیدواژه می سازد و در پایگاه اطلاعاتی MySQL ذخیره می نماید، به تعبیر دیگر برنامه ای که برای یک سند، شاخص تولید می کند. پارامترهای نمایه سازی^۲ در فایل پیکربندی (/admin/conf.php) قرار می گیرند. به عنوان مثال، برای تعیین اینکه فایل های pdf را ایندکس کنیم یا انواع فایل های دیگر را، پارامترها را در فایل conf.php تغییر می دهیم.

1. Indexer

2. Index table

Searcher: جدول نمایه^۱ را بررسی می کند و به دنبال نتایج منطبق با وزن های مختلف است. علاوه بر این، جستجوگر با رشته های جستجو در ارتباط است و آنها را با suffix stripping که یک الگوریتم برای ریشه یابی کلمه^۲ است، پردازش می کند.

Ranker and Sorter: نتایج به دست آمده را بر اساس وزنشان رتبه بندی می کند. وزن با استفاده از فرمول رنکینگ (شکل ۶) محاسبه می شود. همچنین، نتایج براساس اسامی "دامنه"^۳ مرتب می شوند.

Filter: کلماتی که کاربران در فیلد جستجو وارد می کنند از این فیلتر عبور می کند و تجزیه و تحلیل می شود. اگر از شرطی که برای جستجوی اطلاعات تعریف شده است عبور کند، در نتایج جستجو ظاهر می شود. در غیر این صورت، اتفاقی نمی افتد فقط نتایجی که بیشترین تکرار را دارند، برگردانده می شوند.

۵-۲-۵. مقایسه ربات "جستجوگر"^۴ موتور جستجوی عمومی و SISE

تغییراتی در کد برنامه Spider به وجود آمده است تا SISE برخلاف ربات Sphider و یا سایر موتورهای جستجوی عمومی، صفحات وب را به صورت هدفمند نمایه کند.

۵-۲-۵-۱. رفتار ربات جستجوگر در موتور جستجوی عمومی

موتورهای جستجوی عمومی، کلمات رایج (مانند am, is, the, in, on و غیره) را که به "کلمات ممنوع"^۵ معروف هستند در فرآیند نمایه سازی قرار نمی دهند. این کلمات به طور معمول، در یک فایل متنی قرار می گیرند. فرض کنید این کلمات در فایل متنی common.txt قرار دارد.

```
Common.txt=Set of common keywords
such as (am ,is,the,to,...)

{X |X=Set Of Indexed Keywords}

X Search in Common.txt
IF (X don't Appear in common.tx)
{
Then save X to database
}
```

General Search Engine

شکل ۲. نمایش بخشی از الگوریتم ربات جستجوگر عمومی

1. Index table
3. domain

2. word stemming
4. Spider

5. Stop Words

در شکل ۲، بخشی از الگوریتم ربات جستجوگر عمومی نمایش داده شده است. در این شکل، مجموعه کلمات نمایه شده از صفحات وب با حرف X نمایش داده شده است. ربات جستجوگر کلمات نمایه شده را با تمامی کلمات موجود در فایل common.txt مقایسه می کند. چنانچه X در فایل common.txt وجود نداشته باشد در بانک اطلاعاتی ذخیره می شود، در غیر این صورت کلمه نمایه شده جزو کلمات ممنوعه محسوب شده و ذخیره نمی شود.

۲-۵-۲-۵. رفتار ربات جستجوگر در SISE

محتوای فایل common.txt در SISE با موتورهای جستجوی عمومی متفاوت است. در SISE، کلمات تخصصی که "مخفف" کلیدواژه‌ها هستند در فایل common.txt قرار می گیرند، در صورتی که در موتورهای جستجوی عمومی "کلمات ممنوع" در این گونه فایل‌ها ذخیره می شوند. با مراجعه به شکل ۳، روش کار ربات جستجوگر در SISE مشاهده می شود.



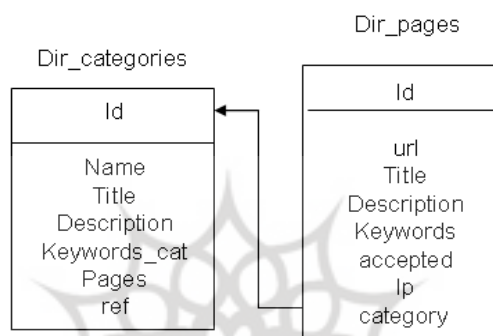
شکل ۳: نمایش بخشی از الگوریتم ربات جستجوگر SISE

در ربات جستجوگر SISE دو تغییر اساسی انجام شده است: (۱) کلمات عمومی بیان شده به کلمات تخصصی تبدیل شده اند؛ (۲) تغییر در کد برنامه Spider است. این کد به گونه ای تغییر می کند که کلمات نمایه شده از صفحات وب را که در شکل ۳ با حرف X نمایش داده شده است با تمامی کلمات موجود در فایل common.txt مقایسه می کند. چنانچه آن کلمه در فایل مربوط وجود داشته باشد در بانک اطلاعاتی ذخیره می شود، در غیر این صورت این کلمه را به عنوان یک کلمه زاید می شناسد و از ذخیره شدن آن در بانک Mysql خودداری می کند. رفتار این ربات جستجوگر، به طور کامل، برعکس موتورهای جستجوی عمومی است.

1. Abbreviation

۵-۲-۶. طراحی فهرست راهنمای تخصصی برای SISE

معماری SISE دارای یک فهرست راهنمای تخصصی است. این فهرست راهنما، SISE را از سایر موتورهای جستجوی عمومی متمایز می‌کند. ساختار داده‌ها در فهرست راهنمای تخصصی (شکل ۴) از دو جدول اصلی تشکیل شده است: یکی برای ذخیره‌سازی "دسته‌بندی‌ها" (Dir_categories) و دیگری محل قرار گرفتن پیوندها و کلیدواژه‌ها (Dir_pages) است.



شکل ۴: معماری داده در فهرست راهنمای SISE

دو نقش اصلی فهرست راهنمای تخصصی در SISE بدین شرح است: (۱) این فهرست راهنما کلیدواژه‌های تخصصی را در فایل common.txt می‌گذارد و سپس این فایل را در اختیار Spider تخصصی قرار می‌دهد. Spider به صورت هدفمند اطلاعات سایت‌ها را بر اساس این کلیدواژه نمایه و در پایگاه اطلاعاتی فهرست راهنما ذخیره و نگهداری می‌نماید؛ (۲) در ادامه، نرم‌افزار فهرست راهنما همه کلیدواژه‌های تخصصی نمایه‌شده را دسته‌بندی می‌نماید تا فهرست راهنمای تخصصی ساخته شود.

۵-۲-۷. آزمون و مقایسه موتور جستجوی عمومی با SISE

در این تحقیق، موتور جستجوی Sphider (موتور جستجوی عمومی) و SISE (موتور جستجوی تخصصی) که توسعه یافته و بهینه‌شده موتور جستجوی Sphider است، مورد آزمون قرار گرفتند. در این آزمون، سایت نظامی designation-systems توسط دو موتور جستجوی اشاره‌شده، نمایه شد که خلاصه نتایج به دست آمده از لحاظ محتوای اطلاعات و نحوه جستجوی اطلاعات به شرح ذیل است:

1. category

• محتوای اطلاعات

محتوای اطلاعات ذخیره شده در SISE، به طور کامل، تخصصی و نظامی است. جدول ۲ اختلاف نوع اطلاعات ذخیره شده برای ۶۰ کلیدواژه مهم^۱ را نشان می دهد.

جدول ۲. مقایسه ۶۰ کلیدواژه با اهمیت ذخیره شده در موتور جستجوی عمومی با SISE

موتورهای جستجو	تعداد کلیدواژه نظامی	تعداد کلیدواژه غیر نظامی	درصد پوشش اطلاعات نظامی	درصد پوشش اطلاعات غیر نظامی
SISE	۶۰	۰	٪۱۰۰	٪۰
Sphider	۷	۵۳	٪۱۲	٪۸۸

وجود اطلاعات غیر تخصصی در موتورهای جستجوی عمومی مانند Sphider باعث می شود تا کاربران در هنگام جستجوی اطلاعات تخصصی ساعت ها وقت صرف کنند و در بیشتر مواقع به اطلاعات مورد نیاز دسترسی پیدا نمی کنند. برعکس، در موتور جستجوی SISE، دسترسی به اطلاعات تخصصی نظامی به خوبی مقدور است به طوری که در ارزیابی SISE در همین مقاله این موضوع به طور کامل به اثبات رسیده است.

• نحوه جستجوی اطلاعات

SISE، اطلاعات تخصصی نظامی را به صورت یک فهرست راهنمای دفاعی برای کاربران ترسیم می کند. این فهرست راهنما به دلیل ساختار تخصصی، ابزار خوبی برای جستجوی اطلاعات نظامی است چون اطلاعات نظامی به صورت یک درخت در لایه ها قرار می گیرد، کاربر با رفتن به لایه ها می تواند به اطلاعات تخصصی دسترسی پیدا کند.

موتورهای جستجوی عمومی، فاقد فهرست راهنمای دفاعی است و بیش از ۸۰ درصد اطلاعات نمایه شده (جدول ۲) غیر نظامی هستند. همین موضوع باعث می شود که در هنگام جستجوی اطلاعات تخصصی نتایج غیر مرتبط زیادی در خروجی جستجو ظاهر شود.

۵-۲-۸ ویژگی های SISE

ویژگی های SISE به دو بخش عمومی و خصوصی تقسیم می شود. ویژگی های عمومی، به ویژگی هایی اطلاق می شود که به طور معمول، به شکل های مختلف در موتورهای جستجوی

1. Top

عمومی همچون گوگل و یاهو وجود دارد. اما، ویژگی‌های خصوصی از ویژگی‌های منحصر به فرد SISE است. ویژگی‌های SISE در جهت پاسخ به سؤال دوم تحقیق به شرح ذیل مطرح می‌گردد:

۵-۲-۸-۱. ویژگی‌های عمومی SISE

(۱) نمایه‌سازی اسناد و پرونده‌هایی با قالب‌های Text، RTF، Microsoft Office، PDF و HTML با پشتیبانی کامل از زبان فارسی؛ (۲) ذخیره‌سازی اطلاعات بر اساس الگوریتم رتبه‌بندی تعریف شده؛ (۳) نمایش آمار تعداد صفحات و لغات کلیدی نمایه شده؛ و (۴) نمایه نمودن سایت‌ها بر اساس عمق تعریف شده.

۵-۲-۸-۲. ویژگی‌های منحصر به فرد SISE

(۱) نمایه‌سازی هدفمند صفحات وب بر اساس کلیدواژه‌های تعریف شده تخصصی؛ (۲) ذخیره‌سازی اطلاعات با استفاده از فهرست راهنمای تعریف شده؛ (۳) استفاده از ماژول هوشمند Re-index به منظور جلوگیری از تکرار صفحات نمایه شده؛ و (۴) جستجوی آسان و سریع با استفاده از الگوریتم تعریف شده در فهرست راهنما.

۵-۲-۹. ماژول‌های هوشمند SISE

۱. الگوریتم ریشه‌یابی: به عنوان مثال، با تایپ کلمه runs و یا running، موتور جستجو با استفاده از الگوریتم هوشمند خود به دنبال ریشه کلمه یعنی run هم می‌گردد. بخشی از این فرمول در شکل ۵ نمایش داده شده است.

```
if (substr($word, -1) == 's') {
  replace($word, 'sses', 'ss')
  OR replace($word, 'ies', 'i')
  OR replace($word, 'ss', 'ss')
  OR replace($word, 's', '');
}
```

شکل ۵. بخشی از الگوریتم ریشه‌یابی

۲. الگوریتم کشف لغات: با تایپ اولین کاراکتر در جعبه جستجو، فهرستی از کلمات موجود در بانک SISE بر اساس حروف لاتین مرتب می‌شود. این الگوریتم کمک قابل توجهی در یافتن اطلاعات به کاربر می‌نماید.

۳. الگوریتم غلط‌یابی: این الگوریتم، کلمه کلیدی تایپ شده را با کلمات کلیدی نمایه شده

در بانک اطلاعاتی خود مقایسه می کند و نزدیک ترین کلمه را از لحاظ شباهت حروف به کلمه کلیدی تایپ شده توسط کاربر نمایش می دهد.

۴. الگوریتم شناسایی علاقه مندی کاربران در کلمات جستجو شده: با استفاده از این الگوریتم می توان فهمید که کاربران دنبال چه اطلاعاتی هستند. در واقع، وقتی کاربری کلمات کلیدی مورد علاقه خود را وارد می کند این کلمات در بانک اطلاعاتی ذخیره می شود و در نهایت، مدیران SISE به این نتیجه می رسند که کاربران به دنبال چه موضوعات علمی هستند. این مسأله کمک می کند به ربات جستجوگر تا در هنگام نمایه سازی سایت ها این نوع اطلاعات را بیشتر مد نظر قرار دهد تا SISE بتواند به خوبی نیازهای اطلاعاتی کاربران را فراهم نماید.

۵. الگوریتم رتبه بندی: ربات جستجوگر، کلیدواژه ها را در ۴ بخش `title`، `domain`، `path` و `meta` جستجو می نماید و در صورت یافتن، آن کلیدواژه را به منظور رتبه بندی در فرمول مربوط مطابق شکل ۶ قرار می دهد.

$$\begin{aligned} \text{\$weight} = & ((\text{\$words_in_page} \\ & + \text{\$word_in_title} * \text{\$title_weight} \\ & + \text{\$word_in_domain} * \text{\$domain_weight} \\ & + \text{\$word_in_path} * \text{\$path_weight} \\ & + \text{\$meta_keyword} * \text{\$meta_weight}) * 10 \\ & / ((0.6 + 0.8 * \text{\$path_depth})) * \text{\$main_url_factor}, \end{aligned}$$

شکل ۶. نمایش فرمول Ranking

۳-۵. گام دوم: ارزیابی نرم افزار SISE

پس از طراحی SISE، عملکرد این موتور جستجو مورد آزمون و ارزیابی قرار گرفت. جامعه آماری پژوهش، مدیران و کارشناسان ارشد متخصص در حوزه دانش نظامی بودند. نمونه مورد نظر به روش تصادفی طبقه ای ساده و به تعداد ۳۰ نفر از بین اعضای جامعه انتخاب شدند. برای انجام این پژوهش، نرم افزار SISE در اختیار هر یک از اعضای جامعه قرار گرفت. از پرسشنامه به عنوان ابزار گردآوری اطلاعات استفاده شد. پرسشنامه به تعداد اعضای نمونه در بین افراد جامعه توزیع گردید. با پیگیری های به عمل آمده هر ۳۰ پرسشنامه توسط اعضای جامعه پاسخ داده و جمع آوری شد. پرسشنامه مورد استفاده از چهار بخش مجزا تشکیل شده است:

بخش نخست، مربوط به ویژگی های جمعیت شناختی پاسخگویان بود. بخش دوم، مربوط به

ارزیابی کمی فهرست راهنما بود؛ تعداد ۹ سؤال پنج گزینه‌ای طراحی گردید. بخش سوم، مربوط به ارزیابی کیفی فهرست راهنما بود؛ تعداد ۸ سؤال سه گزینه‌ای آزمونی و تشریحی طراحی گردید. بخش چهارم، مربوط به ارزیابی و عملکرد موتور جستجو در مقایسه با سایر موتورهای جستجوی پر کاربرد بود؛ تعداد ۵ سؤال پنج گزینه‌ای آزمونی طراحی گردید. در مجموع، پرسشنامه شامل ۲۲ سؤال (جدول ۳) بوده است.

روش‌های متعددی برای تعیین اعتبار ابزار اندازه‌گیری وجود دارد که یکی از مهمترین آنها اعتبار محتواست. اعتبار محتوا، به‌طور معمول، برای بررسی اجزای تشکیل‌دهنده یک ابزار اندازه‌گیری به کار برده می‌شود. اعتبار محتوای یک ابزار اندازه‌گیری به سؤال‌های تشکیل‌دهنده آن بستگی دارد. اگر سؤال‌های ابزار، معرف ویژگی‌ها و مهارت‌های ویژه‌ای باشد که محقق قصد اندازه‌گیری آنها را داشته باشد، آنگاه آزمون دارای اعتبار است (بازرگان، سرمد، و حجازی ۱۳۷۹).

از آنجا که اعتبار روایی یک آزمون توسط افرادی متخصص در موضوع مورد مطالعه تعیین می‌شود، اعتبار روایی به قضاوت داوران بستگی دارد. از این رو، معیارهای بیان‌شده در حوزه SISE در قالب پرسشنامه و مصاحبه حضوری در اختیار ۵ نفر از خبرگان با تخصص‌های علوم نظامی قرار گرفت. در نهایت، از نظرات اساتید راهنما و مشاور و مشاور صنعتی نیز در جهت اصلاح و بهبود پرسشنامه بهره گرفته شد. بدین ترتیب، روایی پژوهش مورد تأیید قرار گرفت.

برای محاسبه ضریب قابلیت اعتماد ابزار اندازه‌گیری (پایایی) ۱، شیوه‌های مختلفی به کار برده می‌شود، از جمله آنها روش آلفای کرونباخ است. این روش برای محاسبه هماهنگی درونی ابزار اندازه‌گیری از جمله پرسشنامه‌ها یا آزمون‌هایی که خصیصه‌های مختلفی را اندازه‌گیری می‌کنند به کار می‌رود و در این گونه ابزار پاسخ هر سؤال می‌تواند مقادیر عددی مختلف را اختیار کند. برای محاسبه ضریب آلفای کرونباخ ابتدا باید واریانس نمره‌های هر زیرمجموعه سؤال‌های پرسشنامه و واریانس کل را محاسبه کرد، سپس با استفاده از فرمول ۱ مقدار ضریب آلفای کرونباخ را محاسبه نمود (بازرگان، سرمد، و حجازی ۱۳۷۹).

$$r_a = \frac{j}{j-1} \left(1 - \frac{\sum S_j^2}{S^2} \right)$$

فرمول ۱

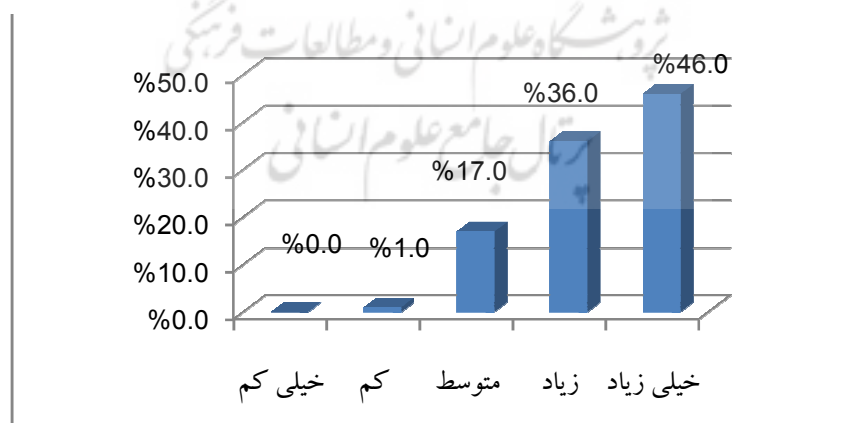
در فرمول ۱، j تعداد زیرمجموعه شاخص‌های ارزیابی پرسشنامه و S_j^2 واریانس زیرآزمون j ام و S^2 واریانس کل آزمون است. بنابراین، اطلاعات پرسشنامه‌ها وارد نرم‌افزار SPSS گردید و سپس ضریب آلفای کرونباخ ($R_a = .925$) به دست آمد که نشان‌دهنده قابلیت اعتماد بالاست.

جدول ۳. ترکیب سؤالات پرسشنامه

سؤالات	شرح	ردیف	شاخص ها
۱ الی ۹	ساختار مناسب فهرست راهنما	۱	ارزیابی کمی فهرست راهنما
۱۰ الی ۱۷	تغییر ساختار فهرست راهنما	۲	ارزیابی کیفی فهرست راهنما
18	مقایسه زمان صرف شده در SISE و سایر موتورهای جستجو (گوگل و یاهو و...)	۳	
۱۹	مقایسه دقت و کیفیت نتایج جستجو در SISE و سایر موتورهای جستجو (گوگل و یاهو و...)	۴	
۲۰	مقایسه سرعت پاسخگویی در SISE و سایر موتورهای جستجو (گوگل و یاهو و...)	۵	ارزیابی SISE
۲۱	مقایسه پیوندهای زاید در SISE و سایر موتورهای جستجو (گوگل و یاهو و...)	۶	
۲۲	مقایسه دسترسی به اطلاعات تخصصی در SISE و سایر موتورهای جستجو (گوگل و یاهو و...)	۷	

۵-۳-۱. ارزیابی فهرست راهنمای طراحی شده در حوزه سامانه‌های موشکی بر اساس شاخص کمی "طراحی ساختار مناسب فهرست راهنما"

نمودار ۱ بیانگر نظرات افراد جامعه مورد مطالعه در خصوص طراحی ساختار مناسب فهرست راهنما در حوزه سامانه‌های موشکی است. در مجموع، ۸۲ درصد پاسخ‌دهندگان، طراحی ساختار فهرست راهنما در حوزه سامانه‌های موشکی را در حد (زیاد و خیلی زیاد) مناسب دانسته‌اند.



نمودار ۱. درصد فراوانی نسبی فهرست راهنما بر اساس شاخص "طراحی ساختار مناسب"

۲-۳-۵. ارزیابی فهرست راهنمای طراحی شده در حوزه سامانه‌های موشکی بر اساس شاخص کیفی "تغییرات ساختاری فهرست راهنما"

در مجموع، از ۸ سؤال کیفی پخش شده بین ۳۰ محقق، فقط در ۴ مورد تغییرات در خصوص اضافه شدن گزینه جدید به فهرست راهنما مطرح شده است. پس از جمع‌بندی و مشورت با سایر محققان این نتیجه به دست آمد که هر ۴ مورد مطرح شده خارج از حوزه سامانه‌های موشکی است و ربطی به این فهرست راهنما ندارد و به بخش‌های دیگر مرتبط است. بنابراین، فهرست راهنما به خوبی موارد مرتبط با سامانه‌های موشکی را پوشش خواهد داد.

۳-۳-۵. ارزیابی SISE "مقایسه عملکرد SISE با سایر موتورهای جستجو"

پس از آنکه SISE در اختیار ۳۰ نفر از اعضای جامعه قرار گرفت از هر یک از آنها خواسته شد در حوزه سامانه‌های موشکی اطلاعات مورد نظر خود را در SISE و سایر موتورهای جستجوی پر کاربرد همچون گوگل و یاهو... جستجو کنند. جدول‌های ۴ و ۵ و نمودار ۱ بیانگر نظرات افراد جامعه مورد مطالعه در خصوص مقایسه عملکرد SISE با سایر موتورهای جستجوی عمومی است. در مجموع، SISE در مقایسه با سایر موتورهای جستجو از شرایط بسیار بهتری برخوردار است.

جدول ۴. ارزیابی و عملکرد SISE در مقایسه با سایر موتورهای جستجو

ردیف	فرآوانی					معیار ارزیابی	مقایسه
	Karfoo	Yahoo	Google	SISE	سایر		
۱	۱		۱	۲۸		زمان صرف شده در دستیابی به اطلاعات تخصصی حوزه موشکی در کدام یک از موتورهای جستجوی اشاره شده کمتر است؟	
۲	۱		۲	۲۷		به نظر شما دقت و کیفیت نتایج جستجو در کدام یک از موتورهای جستجوی اشاره شده بیشتر است؟	
۳		۱	۴	۲۵		سرعت پاسخگویی به درخواست شما در کدام یک از موتورهای جستجوی اشاره شده بیشتر است؟	
۴				۳۰		پیوندهای زاید و گمراه کننده در کدام یک از موتورهای جستجوی اشاره شده کمتر است؟	
۵	۱	۰	۱	۲۸		در مجموع، برای دسترسی به اطلاعات تخصصی حوزه موشکی کدام یک از موتورهای جستجوی اشاره شده را به ما پیشنهاد می‌کنید؟	

جدول ۵. درصد فراوانی نسبی ارزیابی و عملکرد SISE در مقایسه با سایر موتورهای جستجو

نام موتور پیشنهادی	درصد فراوانی نسبی					عامل کلیدی: مقایسه SISE با سایر موتورهای جستجو	ردیف
	سایر	Kartoo	Yahoo	Google	SISE		
		۷		۷	۸۶	زمان صرف شده کمتر است.	۱
		۷		۱۳	۸۰	دقت و کیفیت نتایج جستجو بیشتر است.	۲
			۷	۲۶	۶۷	سرعت پاسخگویی به درخواست کاربر بیشتر است.	۳
					۱۰۰	پیوندهای زاید کمتر است.	۴
		۷	۰	۷	۸۶	پیشنهاد استفاده.	۵

جدول ۶. خلاصه نتایج کسب شده آماری

درصد آرا	شرح	ردیف	شاخص ها
٪۸۲	ساختار مناسب فهرست راهنما در حد (خیلی زیاد و زیاد)	۱	ارزیابی کمی فهرست راهنما
٪۱۸	تغییر ساختار فهرست راهنما	۲	ارزیابی کیفی فهرست راهنما
٪۸۶	در SISE زمان جستجو کمتر است.	۳	ارزیابی موتور جستجوی هوشمند تخصصی در مقایسه با سایر موتورهای جستجو
٪۸۰	در SISE دقت و کیفیت نتایج جستجو بیشتر است.	۴	
٪۶۷	در SISE سرعت پاسخگویی به درخواست کاربر بیشتر است.	۵	
٪۱۰۰	در SISE پیوندهای زاید کمتر است.	۶	
٪۸۶	پیشنهاد استفاده از SISE	۷	

جدول ۶ نشان می‌دهد: (۱) ۸۶ درصد از جامعه آماری معتقدند که زمان صرف‌شده برای جستجوی اطلاعات در SISE کمتر از سایر موتورهای جستجوی عمومی شناخته شده است؛ (۲) ۸۰ درصد معتقدند که دقت و کیفیت نتایج جستجو در SISE بیشتر از سایر موتورهای جستجوی عمومی است؛ (۳) ۶۷ درصد معتقدند که سرعت پاسخگویی به درخواست کاربر در SISE بیشتر از سایر موتورهای جستجوی عمومی است؛ (۴) ۱۰۰ درصد از جامعه آماری معتقدند که پیوندهای زاید در SISE کمتر از سایر موتورهای جستجوی عمومی است. بنابراین نتایج این جدول باعث اثبات فرضیه‌های تحقیق می‌گردد.

۶. نتیجه‌گیری

در این تحقیق، کتاب‌ها و مقالات و موتورهای جستجوی متن باز بسیاری مورد مطالعه و بررسی قرار گرفت. تمرکز تحقیق بر روی تهیه موتور جستجوی هوشمند تخصصی بوده است. نتایج این تحقیق نشان می‌دهد که فهرست راهنمای طراحی شده برای موتور جستجوی هوشمند تخصصی، آن را از سایر موتورهای جستجوی عمومی مانند گوگل و یاهو متمایز نموده و باعث شده است تا کاربران به اطلاعات تخصصی دسترسی پیدا نمایند. آزمایش به‌خوبی نشان داد که اطلاعات زاید و غیرتخصصی در این فهرست راهنما قرار نمی‌گیرد. SISE، اطلاعات تخصصی را در قالب یک فهرست راهنما برای کاربران ترسیم می‌کند. این فهرست راهنما، به‌دلیل ساختار تخصصی، ابزار خوبی برای جستجوی اطلاعات است. از آنجا که اطلاعات تخصصی به‌صورت یک درخت در لایه‌ها قرار می‌گیرند، کاربر با رفتن به لایه‌ها می‌تواند به اطلاعات تخصصی دسترسی پیدا کند. ویژگی‌های مهم SISE به‌منظور پاسخ به سؤال دوم تحقیق به‌شرح زیر است:

۱. دسترسی آسان و سریع به اطلاعات تخصصی نظامی با استفاده از فهرست راهنمای تعریف‌شده؛
۲. صرفه‌جویی در هزینه؛
۳. عدم ذخیره‌سازی اطلاعات زاید؛
۴. عدم نمایه مجدد صفحاتی که اطلاعات آنها تغییری نداشته است، به‌دلیل داشتن ماژول هوشمند reindex؛
۵. قابلیت توسعه فراگیر SISE به‌دلیل استفاده از زبان برنامه‌نویسی متن باز؛
۶. قابل استفاده در شبکه‌های اینترنت و اینترنت؛
۷. قابل نصب بر روی سیستم عامل‌های ویندوز و لینوکس؛
۸. دسته‌بندی اطلاعات نظامی بر اساس فهرست راهنمای تعریف‌شده؛ و
۹. نمایه هدفمند صفحات بر اساس لغات کلیدی تعریف‌شده؛

۱۰. با استفاده از این فهرست راهنما می‌توان با تغییر لغات کلیدی از نظامی به غیر، دانش‌های تخصصی برای سایر علوم ایجاد نمود. بنابراین، محتوای فهرست راهنمای ساخته شده قابل تغییر است، یعنی اگر در این فهرست راهنما به جای کلیدواژه‌های نظامی، لغات پزشکی قرار گیرد پایگاه دانش پزشکی ساخته می‌شود. در نتیجه، با تغییر کلیدواژه‌ها می‌توان پایگاه‌های دانشی مختلفی را ایجاد کرد.

بر این اساس، به‌عنوان یک پیشنهاد پژوهشی، انتظار می‌رود که در پژوهش‌های آتی، از یک سو، اهمیت امکانات بررسی شده در این تحقیق از نظر کاربران مورد توجه قرار گیرد و از سوی دیگر، با توجه به متن باز بودن این نرم‌افزار، موارد زیر توسط محققان علاقه‌مند در این حوزه به این موتور جستجو اضافه گردد:

- توسعه ربات SISE برای نمایه‌سازی تصاویر و فیلم‌ها (در حال حاضر، SISE قادر نیست تصاویر و فیلم‌های موجود در صفحات وب را نمایه کند)؛
- توسعه ماژول جستجوی پیشرفته مطابق موتور جستجوی گوگل (جستجو به تفکیک نوع فایل و فیلم‌های مختلف)؛
- توسعه ربات جستجوگر به‌منظور جمع‌آوری و ذخیره‌سازی اطلاعات مورد علاقه کاربران از وبسایت‌ها؛ و
- توسعه طراحی ساختار فهرست راهنما در حوزه غیر نظامی.

۷. منابع

- بازرگان، عباس، زهره سرمد، و الهه حجازی. روش‌های تحقیقی در علوم رفتاری. تهران: انتشارات آگاه، ۱۳۷۹.
- Allan, B., G. O. Roberts, J. S. Rosenthal, and P. Tasparas. 2005. Links analysis ranking algorithms, theory, and experiments. March 25, 2004. *Journal ACM Transactions on Internet Technology (TOIT)* 5 (1): 1-7.
- Isfandyari-Moghaddam, A. 2007. Web metasearch engines: a comparative study on search capabilities using an evaluation checklist. *Online Information Review* 31 (3): 300-309.
- Liu, J. 1999. Guide to meta-search engines. www.indiana.edu/~libresd/search/meta.html (accessed 10 Nov. 2009).
- Middleton, C., and R. Baeza-Yates. 2005. A comparison of open source search engines. <http://wrg.upf.edu/WRG/dctos/Middleton-Baeza.pdf> (accessed 7 Nov. 2007).
- Pant, G., and F. Menczer. 2002. My spiders: evolve your own intelligent Web crawlers. *Autonomous Agents and Multi-Agent Systems* 5 (2): 221-229.
- Shkapyuk, V., and T. Suel. 2002. Design and implementation of a high-performance distributed Web crawler: Data engineering. In *Proceedings of 18th International Conference on Data Engineering (ICDE), San Jose, CA Feb. 26-March 1, 357-368*, San Jose, California: IEEE CS Press.
- Song, J. P., Y. C. Wang, W. Ten, and H. Q. Xu. 2002. Improvement of the robot search algorithm. *Journal of Information Science* 21 (2): 130-133.

- Thomas, B., and C. Wolff. 2001. Ubiquitous web information agents, workshop on artificial intelligence in mobile systems, ECAI 2000. In *Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, Germany, August 20-25, 2000*. IOS Press 2000. Werner Horn (Ed.), 599-603. Berlin: ECAI 2000.
- Yajun Du, Yang Xu, Liangzhong Yi, and Zheng Pei. 2004. Research on personal intelligent search engine based on keywords Dept. of Appl. Math., Southwest Jiaotong Univ., Chengdu, China. In *Communications and Information Technology, 2004. ISCIT 2004. IEEE International Symposium on Issue Date: 26-29 Oct. 2004*, Y. Mi Yanaga (Ed.), 228 – 231. Los Alamitos, CA, U.S.A.



Special Intelligent Search Engine

Reza Sarabadani*

Ph.D in Artificial Intelligence,
Professor Assistant, Malek Ashtar University

Hossein Shirazi¹

Master student in Information Technology,
Malek Ashtar University

Information
Sciences
& Technology

Abstract: In this paper Special Intelligent Search Engine (SISE) design has been introduced. Also System architecture and methods of this component of its intelligent search engine and features have been described. For this intelligent robot, a "Directory" designed that aims to build professional knowledge base for quick and easy access to information for user. First of all specialized keywords were defined in the list based on the sites were indexed. 86% of the population tested, believed that the time spent searching for information SISE was lower than other general search engines and 80% believed that quality and accuracy of search results is SISE were more than other search engines.

Keywords: search engines, robot explorer, web directory, intelligent search engines, special intelligent search engines

Iranian Research Institute
For Science and Technology
ISSN 1735-5206
eISSN 2008-5583
Indexed in LISA, SCOPUS & ISC
Vol.26 | No.4 | pp: 903-924
summer 2011

پژوهشگاه علوم انسانی و مطالعات فرهنگی
رتال جامع علوم انسانی

*Corresponding author rezas1388@yahoo.com
1. hossein_shirazi@yahoo.com