



ارزیابی خودکار جویش گره‌های ویدئویی حوزه وب فارسی بر اساس تجمیع آرا

محمد مهدی یداللهی، فرزاد زرگری* و مژگان فرهودی

گروه سکویهای فناوری اطلاعات، پژوهشکده فناوری اطلاعات، مرکز تحقیقات مخابرات ایران، تهران، ایران

چکیده

امروزه رشد بسیار سریع اینترنت و نفوذ روزافزون آن در زندگی افراد باعث شده تا کاربران بسیاری برای رفع نیازهای روزمره خود به جویش گره‌ها مراجعه کنند و این جویش گره‌ها به توسعه و بهبود مستمر نیاز دارند. از این رو ارزیابی جویش گره‌ها برای تعیین کارایی آنها اهمیت به‌سزایی دارد. در ایران نیز همانند سایر کشورها پژوهش‌های گسترده‌ای در زمینه ایجاد جویش گره‌های خاص‌منظوره بومی انجام شده است. یکی از مهم‌ترین جویش گره‌های خاص‌منظوره ایجاد شده، جویش گر ویدئویی است که وظیفه بازیابی ویدئوها از سطح وب را برعهده دارد. برای ارزیابی کیفیت این جویش گره‌ها و بهبود مستمر آنها باید سطح خدمات‌دهی هر کدام از جویش گره‌ها در مقایسه با دیگر جویش گره‌های موجود مورد ارزیابی قرار گیرد. از آنجا که سرعت ارزیابی نقش مهمی در تعیین روند اصلاحات مورد نیاز دارد، بحث ارزیابی خودکار جویش گره‌ها بسیار پراهمیت خواهد شد. در این مقاله روشی مبتنی بر تجمیع آرا به‌منظور ارزیابی خودکار جویش گره‌های ویدئویی ارائه شده است. تمرکز اصلی این روش بر روی حوزه وب فارسی بوده و با توسعه روشی نوین برای شباهت‌سنجی مبتنی بر محتوا براساس بردارهای حرکت ویدئوها، سعی در ارزیابی جویش گره‌های ویدئویی دارد. برای محک‌زدن روش معرفی شده، سازوکاری طراحی شد تا نتایج به‌دست آمده با نتایج حاصل از ارزیابی انسانی مورد مقایسه قرار گیرد. نتایج به‌دست آمده نشان‌دهنده میزان همبستگی بیش از ۹۴٪ دو روش است که قابل اتکابودن روش خودکار ارزیابی پیشنهادی را بیان می‌کند.

واژگان کلیدی: ارزیابی خودکار، جویش گره‌های ویدئویی، وب فارسی.

Automatic Evaluation of Video Search engines in Persian Web domain based on Majority Voting

Mohammad Mehdi Yadollahi, Farzad Zargari* & Mojgan Farhoodi

Group of Information Technology at Iran Telecommunication Research Center, Tehran, Iran

Abstract

Today, the growth of the internet and its high influence in individuals' life have caused many users to solve their daily needs by search engines and hence, the search engines need to be modified and continuously improved. Therefore, evaluating search engines to determine their performance is of paramount importance. In Iran, as well as other countries, extensive researches are being performed on search engines. To evaluate the quality of search engines and continually improve their performance, it is necessary to evaluate search engines and compare them to other existing ones. Since the speed plays an important role in the assessment of the performance, automatic search engine evaluation methods attracted grate attention. In this paper, a method based on the majority voting is proposed to assess the video search engines. We introduced a mechanism to assess the automatic evaluation method by comparing its results with the results obtained by human search engine evaluation. The results obtained, shows 94 % correlation of the two methods which indicate the reliability of automated approach.

In general, the proposed method can be described in three steps.

*Corresponding author

* نویسنده عهده‌دار مکاتبات

Step 1: Retrieve first k retrieve results of n different video search engines and build the return result set for each written query.

Step 2: Determine the level of relevance of each retrieved result from the search engines

Step 3: Evaluating the search engines by computing different evaluation criteria based on decisions on relevance of the retrieved videos by each search engine

Clearly, the main part of any evaluation system with the goal of evaluating the accuracy of search engines is the second step. In this paper, we have tried to present a new solution based on the aggregation of votes in order to determine whether a result is relevant or not, as well as its level of relevance. For this purpose, for each query the return results from different search engines are compared with each other, and the result returned by more than m of the search engines ($m < n$) will be considered as a relevant result. It should be noted that we evaluate similarity of results at two levels: the web page address and the content of that page. In the similarity measure component, the results are compared at the first level based on the URL address; and the result of which their URLs (after the normalization) are similar to the normalized URL from the $m-1$ of the other search engines, are considered as the relevant results. At the second level, the retrieved results will be compared in terms of content. In this way, after calculating the address-like similarity, all the results are transmitted to the motion vector extraction component to extract and store the motion vector.

In the content based similarity algorithm, the set of motion vectors is initially considered as a sequence of motion vector. We, then, try to find the greatest similarity of the smaller sequence with the larger sequence. After this step, we will report the maximum similarity of the two videos. The process of finding the maximum similarity is that we consider a window with a smaller video sequence length. In this window we calculate and hold the similarity of two sequences. In the proposed method, after identifying the similarity between the return results of different search engines, their level is ranked at three different levels: "unrelated" (0), "quantitatively related" (1) and "related" (2). Since Google's search engine is currently the world's largest and best-performing search engine, and most search engines have been compared to it, and are also trying to achieve the same function, the first five Google search engine results are get the minimum relevance, by default, "slightly related". Then the similarity module is used to evaluate the similarity of the retrieved n results of the tested search engines.

Keywords: Automatic assessment, Video search engine, Persian Web

این معیار اهمیت بالایی دارد که با استفاده از شاخص دقت سنجیده می‌شود.

روش‌های مختلفی برای ارزیابی جویش‌گرها وجود دارد که با توجه به عامل ایجاد مجموعه قضاوت به دو دسته مهم روش‌های ارزیابی خودکار [21-3] و انسانی [2, 25-22] قابل تقسیم هستند. در واقع یک سامانه ارزیابی جویش‌گر شامل سه بخش اصلی است که عبارتند از واحد تولید پرس و جوها، واحد قضاوت و واحد محاسبه معیارها. ایجاد واحد قضاوت مهم‌ترین و هزینه‌برترین بخش در یک سامانه ارزیابی است. مبنای خودکار یا انسانی بودن روش ارزیابی نیز به نحوه پیاده‌سازی بخش قضاوت‌های مرتبط برمی‌گردد.

موضوع مهمی که باید توجه داشت آن است که اغلب روش‌های ارزیابی انسانی دقت بالاتری دارند؛ اما به دلیل هزینه‌بر و زمان‌بر بودن به‌کارگیری آنها، استفاده از روش‌های خودکار نقش مهمی در خودارزیابی جویش‌گرها دارد؛ به‌نحوی که توسعه‌دهندگان آنها به‌منظور امکان ارزیابی تأثیر تغییرات اعمال‌شده و نسخه‌های جدید نسبت به نسخه‌های پیشین از این روش‌ها به‌صورت گسترده‌ای استفاده می‌کنند.

۱- مقدمه

امروزه با پیشرفت روزافزون شبکه جهانی اینترنت و همچنین ارتقای سرعت آن، غالب محتوای منتشرشده در این شبکه به سمت داده‌های چندرسانه‌ای سوق پیدا کرده است. مهم‌ترین داده چندرسانه‌ای، فایل‌های ویدئویی است. موتور جستجوی ویدئویی ابزار تحت وبی است که صفحات وب را برای پیدا کردن فایل‌های ویدئویی مورد بررسی قرار می‌دهد. برخی از موتورهای جستجو مانند گوگل، یاهو و بینگ، فایل‌های ویدئویی را بر روی پایگاه ارائه‌دهنده آن پویش، اما برخی دیگر مانند Blinkx، فایل‌های ویدئویی را بر روی سرورهای خود نگهداری می‌کنند. همچنین در بعضی از این موتورهای جستجو قابلیت جستجو براساس فرمت فایل ویدئویی، تاریخ بارگذاری و طول زمان آن امکان‌پذیر است.

با توجه به اهمیت جویش‌گرها به‌عنوان درگاه ورود هشتاد درصد کاربران به وب و همچنین نقش آنها در بازیابی اطلاعات از وب، مبحث ارزیابی جویش‌گرها اهمیت دارد. مهم‌ترین عامل رضایت کاربران از یک موتور جستجو، مرتبط بودن نتایج ارائه‌شده توسط آن است. بنابراین ارزیابی

۴ نتایج و تحلیل‌های مرتبط ارائه خواهد شد و در نهایت مقاله در بخش ۵ جمع‌بندی می‌شود.

۲- کارهای مرتبط

بررسی‌های انجام‌شده در حوزه ارزیابی جویس گرهای نشان می‌دهد که تا کنون پژوهش‌های جامعی برای ارزیابی جویس گرهای ویدئویی به صورت خودکار انجام نشده است؛ در نتیجه با توجه به ضرورت بررسی کارهای انجام‌شده و همچنین مشخص شدن ابعاد مسئله، پژوهش‌های موجود در زمینه ارزیابی جویس گرهای عام مانند جویس گرهای متنی مورد بررسی قرار گرفته‌اند. با الهام از پژوهش‌های انجام‌شده در این حوزه می‌توان تخمینی از ابعاد موجود در ارزیابی جویس گرهای ویدئویی داشت.

برای این منظور کارهای انجام‌شده در حوزه ارزیابی جویس گرهای متنی دسته‌بندی شده تا درک بهتری از آن‌ها به دست آید. به طور کلی کارهای صورت‌گرفته را در حوزه ارزیابی جویس گرهای می‌توان در چهار دسته انسانی، نیمه خودکار (مبتنی بر بازخورد کاربر)، مبتنی بر رتبه‌بندی مجدد و مبتنی بر تجمیع آرا جای داد.

روش‌های انسانی از متداول‌ترین و قابل اعتمادترین انواع ارزیابی هستند. در این روش‌ها نتایج بازگشتی از جویس گرهای متفاوت برای یک پرس‌وجو به کاربر انسانی نشان داده می‌شود و کاربر انسانی هر کدام از نتایج را بررسی و سطح مرتبط بودن هر نتیجه، را گزارش می‌کند. در پایان از اطلاعاتی که کاربر در اختیار سامانه ارزیابی قرار داده است، معیارهای ارزیابی محاسبه و جویس گرهای نسبت به یکدیگر مقایسه می‌شوند.

دسته دوم روش‌های نیمه خودکاری هستند که بر مبنای بازخورد کاربر بنا شده‌اند. در [13] اطلاعات موجود در فایل کلیک کاربران به عنوان نشانه‌هایی دال بر ترجیحات نسبی اسناد نسبت به یکدیگر در نظر گرفته شده‌اند که متفاوت با کارهای قبلی است که کلیک شدن اسناد را به عنوان مرتبط بودن مطلق آن‌ها در نظر می‌گیرد. روش مطرح در [14] روی پرس و جوهای پیمایشی تمرکز کرده است. در این مقاله برای ایجاد مجموعه قضاوت از ترکیبی از رفتار ضمنی کاربر و کلیک روی نتایج استفاده شده است. پاسخ متناظر با یک پرس‌وجوی پیمایشی، سندی در فهرست نتایج متناظر با آن است که بیشتر از سایر اسناد موجود در این فهرست، توسط کاربران کلیک شده است. یعنی کاربران چنین نتیجه‌ای را، نتیجه مناسبی برای پرس‌وجوی خود یافته‌اند و علاقه‌مند به

از نگاه سازوکار ایجاد مجموعه قضاوت و داده‌های ورودی مورد استفاده آنها روش‌های ارزیابی خودکار جویس گرهای، در دو دسته مهم روش‌هایی مبتنی بر بازخورد کاربر [16-13، 18] و روش‌های مبتنی بر تجمیع آرا و رتبه‌بندی مجدد نتایج موتورهای جستجو [1]، [3]، [5]، [8]، [12]، [19]، [20]، [21] قابل ارائه هستند. روش‌های مبتنی بر بازخورد کاربران از اطلاعاتی نظیر کلیک کاربر یا رفتار ضمنی کاربر حین جستجو استفاده می‌کنند. این دسته از روش‌ها با توجه به بهره‌برداری از مشارکت غیرمستقیم کاربران از دقت بالایی برخوردار هستند؛ اما محدودیت‌هایی برای استفاده و پیاده‌سازی آنها وجود دارد. دسته دوم روش‌ها تحت عنوان "روش‌های مبتنی بر تجمیع آرا و رتبه‌بندی مجدد نتایج موتورهای جستجو" از نتایج رتبه‌بندی شده توسط جویس گرهای به عنوان داده ورودی اولیه استفاده می‌کند. دو سیاست متداول استفاده‌شده در این زمینه عبارتند از استفاده از اجماع جویس گرهای و توافق آنها روی نتایج مشابه و یا رتبه‌بندی مجدد نتایج جویس گرهای (مجموع شده در یک مخزن) با استفاده از شاخص‌ها و سازوکارهای متنوع. برای پرس و جوهای پیمایشی نیز علاوه بر روش‌های یادشده دسته دیگری از روش‌ها تحت عنوان روش‌های مبتنی بر بخش‌های شناخته شده [4]، [10]، [11] نیز مطرح هستند.

موضوع مورد بررسی در این مقاله پیاده‌سازی یک روش ارزیابی خودکار برای ارزیابی جویس گرهای ویدئویی بومی و استفاده از آن برای مقایسه جویس گرهای بومی با چند جویس گر بین‌المللی نظیر گوگل و بینگ است. واضح است که به دلیل عدم دسترسی به جزئیات پیاده‌سازی سامانه‌های جویس گرهای، ارزیابی آنها به روش جعبه سیاه انجام می‌شود. همچنین به منظور اطمینان از صحت عملکرد روش پیاده‌سازی شده، همبستگی روش با روش ارزیابی انسانی نیز مقایسه می‌شود. روش پیشنهادی در دسته دوم روش‌ها یعنی رتبه‌بندی مبتنی بر تجمیع آرا قرار می‌گیرد. مهم‌ترین ویژگی روش پیشنهادی که آن را با سایر روش‌های ارائه‌شده تا کنون متمایز می‌کند، عدم وابستگی به روش رتبه‌بندی مورد استفاده جویس گرهای است و تنها پیش‌فرضی که در این روش وجود دارد، عدم همبستگی بین جویس گرهای مورد ارزیابی است؛ چون از اجماع جویس گرهای بر اساس شباهت نتایج URL ها یا محتوای نتایج ارائه‌شده توسط آنها استفاده می‌کند.

به همین منظور در ادامه در بخش ۲ ابتدا مروری بر روش‌های ارزیابی خودکار جویس گر ویدئویی خواهیم داشت؛ سپس در بخش ۳ راهکار پیشنهادی ارائه شده و در بخش

مشاهده آن بوده‌اند. روش مطرح‌شده در [16] یک روش نیمه‌خودکار است. در این روش با ارائه فهرستی از ترکیب نتایج جویش‌گرهای مختلف، تنها از نیروی انسانی خواسته می‌شود مطابق با رفتار معمولی که در استفاده از یک جویش‌گر دارد، بر روی اسنادی که در فهرست نتایج هستند و آن‌ها را پسندیده است کلیک کند. بنابراین نحوه به‌دست‌آوردن داده‌های کلیک به‌صورت بی‌طرفانه و از طریق واسط یکسان است.

در [18] روشی دیگر برای ارزیابی خودکار در نظر گرفته شده است که اصطلاحاً از بازخوردهای ضمنی برای این موضوع استفاده می‌کند. در این مقاله ارزیابی سه‌سطحی برای تعیین مرتبط بودن نتایج استفاده شده است. معیارهای مرتبط بودن عبارتند از رونوشت گرفتن از صفحه^۱، اضافه کردن به صفحات مورد علاقه^۲، نشانه‌گذاری صفحه^۳، چاپ کردن^۴، ذخیره کردن^۵ و Scroll کردن صفحه^۶. انجام هر یک از این اعمال در حقیقت می‌تواند یک بازخورد ضمنی از سمت کاربر تلقی شود؛ در غیر این صورت سند بازگشتی می‌تواند به‌عنوان نامرتبط در نظر گرفته شود. روش‌های موجود در دسته سوم به نوعی قصد دارند تا با رتبه‌بندی مجدد نتایج به‌دست آمده از جویش‌گرهای متفاوت آن‌ها را مورد ارزیابی قرار دهند. از روش‌های بارز این دسته می‌توان به موارد زیر اشاره کرد.

روش AWSEEM از جمله روش‌های تجمیع رتبه در [12] معرفی شده است. در این روش پرس‌وجوها به هشت جویش‌گر از جمله یاهو و ایم‌اس‌ان ارسال شده‌اند؛ سپس دویست نتیجه نخست هر یک از این جویش‌گرها به یک مجموعه‌ی نهایی افزوده شده و با توجه به میزان شباهت‌شان با پرس و جوی ارسالی مورد رتبه‌بندی قرار گرفته‌اند. از جمله نقاط ضعف این روش این است که موتورهای جستجو در هنگام رتبه‌بندی اسناد پارامترهای متعددی را مدنظر قرار می‌دهند؛ درحالی‌که روش AWSEEM تنها به محتوای متنی صفحات وب توجه می‌کند. این روش به‌عنوان مبنای کار [3] قرار گرفته است. برای بهبود این روش پارامترهای دیگری نظیر پیج‌رنک^۷ و الکسانرنک^۸ برای انجام ارزیابی به این روش افزوده شدند. روش مطرح‌شده در [19] روشی بر اساس تجمیع رتبه‌ها است. تجمیع رتبه‌ها روشی است که در آن رتبه‌بندی‌های صورت‌گرفته توسط چند روش مختلف با یکدیگر به نحوی ترکیب می‌شوند تا رتبه‌بندی‌ای که حاصل

می‌شود، کیفیت بالاتری داشته باشد. در روش مطرح‌شده در این مقاله مهم‌ترین گام، گام یادگیری است که در آن قواعد رتبه‌بندی به‌کمک چهار روش (۱) الگوریتم PageRank، (۲) معیار شباهت دودویی (بازیابی دودویی)، (۳) مدل فضای برداری و (۴) بازخوردهای ضمنی کاربران استخراج می‌شود. دسته آخر از روش‌های ارزیابی جویش‌گرها، روش‌های مبتنی بر تجمیع آرا است. مبنای این روش‌ها تعداد تکرار یک نتیجه در جویش‌گرهای متفاوت است. روش [8] مبتنی بر تجمیع آرا است. ابتدا پرس‌وجو به هریک از جویش‌گرها تحت بررسی داده می‌شود؛ از نتایج بازگشتی پرس‌وجوی فعلی، n سند دارای بالاترین رتبه در نظر گرفته می‌شوند؛ سپس، بررسی می‌شود که هر سند در نظر گرفته‌شده برای هر جویش‌گر، چند بار در مجموعه اسناد در نظر گرفته‌شده از سایر جویش‌گرها تکرار شده است. مجموع تکرارها برای هر سامانه بازیابی اطلاعات به‌عنوان امتیاز آن سامانه لحاظ می‌شود. جویش‌گرها بر اساس امتیازات محاسبه شده مورد رتبه‌بندی قرار می‌گیرند.

۳- راه‌کار پیشنهادی

هدف از این پژوهش ارائه راه‌کاری نوین مبتنی بر تجمیع آرا برای ارزیابی خودکار جویش‌گرهای ویدئویی در حوزه وب فارسی است. لازم به ذکر است که روش‌های ارزیابی جویش‌گرهای متنی از آن جهت که محتوای متنی صفحات را مورد بررسی قرار می‌دهند، نمی‌توانند برای ارزیابی جویش‌گرهای ویدئویی مورد استفاده قرار گیرند. در شکل (۱) شمای کلی روش پیشنهادی ارائه شده است. به‌طور کلی روش پیشنهادی را در سه گام کلی می‌توان بیان کرد.

(۱) گام نخست: بازیابی $k_{retrieve}$ نتیجه از n جویش‌گر ویدئویی متفاوت و ساخت مجموعه نتایج بازگشتی به‌ازای هر پرس و جوی نوشتاری

(۲) گام دوم: مشخص کردن سطح مرتبط بودن هر یک از نتایج بازگشتی از جویش‌گرها

(۳) گام سوم: محاسبه معیارهای ارزیابی متفاوت واضح است که بخش اصلی هر سامانه ارزیابی با هدف ارزیابی دقت جویش‌گرها، گام دوم است. در این مقاله، سعی شده است، راه‌کاری نوین مبتنی بر تجمیع آرا برای مشخص کردن مرتبط بودن یا نبودن یک نتیجه و همچنین

⁵ Saving

⁶ Scrolling

⁷ Page Rank

⁸ Alexa Rank

¹ Copying

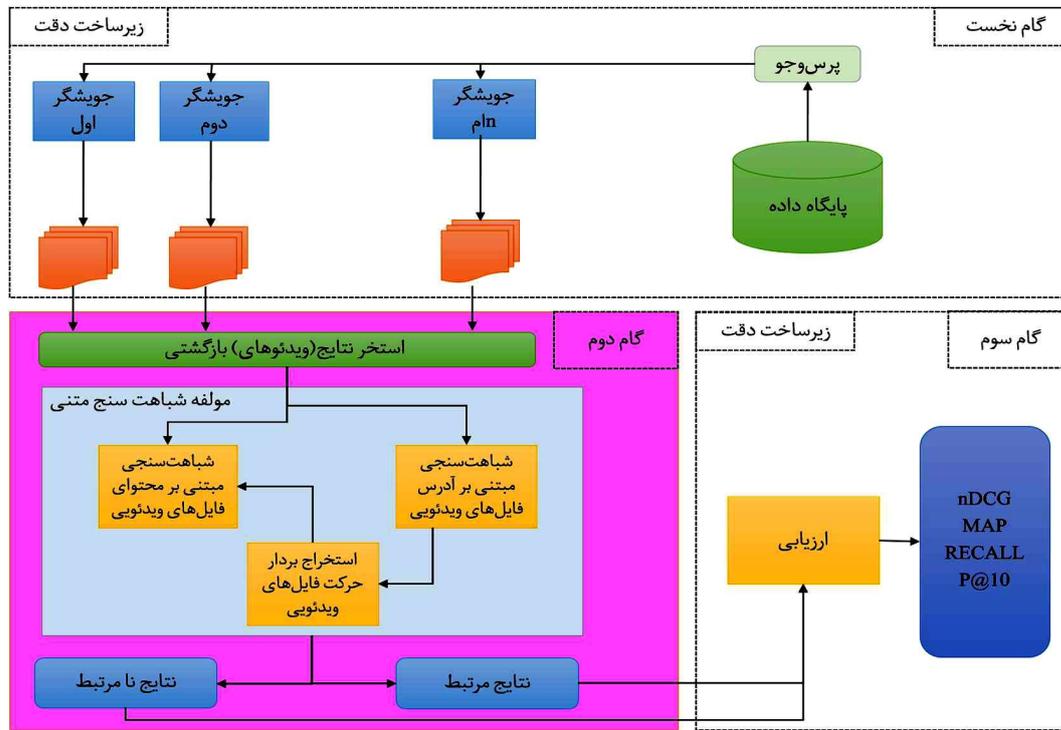
² Adding to Favorite

³ Bookmarking

⁴ Printing

خواهد شد. لازم به ذکر است که مقایسه نتایج از طریق شباهت سنجی در دو سطح نشانی صفحه وب و همچنین محتوای آن صفحه صورت خواهد گرفت.

سطح مرتبط بودن آن ارائه شود. برای این منظور، به ازای هر پرس و جو، نتایج بازگشتی از جویش گرهای متفاوت با یکدیگر مقایسه شده و نتیجه‌ای که توسط بیشتر از m جویش گر ($m < n$) بازیابی شده است، به عنوان نتیجه مرتبط در نظر گرفته



(شکل-1): شمای کلی سامانه پیشنهادی ارزیابی خودکار جویش گرهای ویدئویی مبتنی بر تجمیع آرا

(Figure-1): Overview of the proposed system of automated evaluation of video search engines based on aggregation of votes

خواهند شد. از جمله این عناصر می توان به <https://www.index.htm> اشاره کرد. پس از این گام، نشانی نرمال شده را به سه قسمت دامنه، نشانی و پارامترها تقسیم کرده و برای مقایسه هر قسمت همانند دو رشته با آن‌ها برخورد می‌شود. لازم به ذکر است که در قسمت پارامترها امکان جابه‌جا بودن پارامترها نیز لحاظ شده است؛ بدین صورت که نشانی‌های مشابه ممکن است، در ترتیب پارامتر ارسالی به یک وب سرویس متفاوت باشند؛ در این حالت این نشانی‌ها یکسان تشخیص داده خواهند شد.

در سطح دوم، نتایج از نگاه محتوا مورد شباهت‌سنجی قرار خواهند گرفت. بدین صورت که پس از محاسبه شباهت مبتنی بر نشانی، تمامی نتایج به مؤلفه استخراج بردار حرکت ارسال شده تا بردار حرکت آن استخراج و ذخیره شود. بردار حرکت، برداری است شامل، مناطق یا بلاک‌های متحرک که این مناطق یا بلاک‌ها در عمل جبران‌سازی حرکت¹ مورد استفاده قرار می‌گیرند. برای پایداری و بهبود عملکرد مؤلفه

¹ Motion compensation

واضح است که بخش اصلی هر سامانه ارزیابی با هدف ارزیابی دقت جویش گرها، گام دوم است. در این مقاله، سعی شده است راه‌کاری نوین مبتنی بر تجمیع آرا برای مشخص کردن مرتبط بودن یا نبودن یک نتیجه و همچنین سطح مرتبط بودن آن ارائه شود. برای این منظور، به ازای هر پرس و جو، نتایج بازگشتی از جویش گرهای متفاوت با یکدیگر مقایسه شده و نتیجه‌ای که توسط بیشتر از m جویش گر ($m < n$) بازیابی شده است، به عنوان نتیجه مرتبط در نظر گرفته خواهد شد. لازم به ذکر است که مقایسه نتایج از طریق شباهت سنجی در دو سطح نشانی صفحه وب و همچنین محتوای آن صفحه صورت خواهد گرفت.

در مؤلفه شباهت‌سنجی، نتایج در سطح نخست از نگاه نشانی URL مورد مقایسه قرار گرفته و نتیجه‌ای که URLهای آن بعد از هنجارسازی شدن مشابه با URL نرمال شده از $m-1$ جویش گر دیگر باشد، به عنوان نتیجه مرتبط در استخر نتایج مرتبط به ازای پرس و جوی مربوطه در نظر گرفته خواهد شد. منظور از نرمال کردن نشانی یک صفحه وب، حذف عناصری است که به طور متداول در بیش تر نشانی‌ها تکرار

شباهت‌سنج نیاز است تا پیش‌پردازش‌هایی انجام گیرد. پیش‌پردازش‌ها به این قرار است که ابتدا فایل‌های ویدئویی را از نگاه نرخ تعداد فریم در ثانیه و تعداد فریم‌های مرجع (I) یکپارچه‌سازی می‌شوند؛ پس از آن هر فایل ویدئویی را به فایل‌هایی ویدئویی با طول سه ثانیه تقسیم می‌کنیم؛ سپس هیستوگرام بردار حرکت هر کدام از فایل‌های ویدئویی استخراج می‌شود.

مجموعه بردارهای حرکت هر دو فایل ویدئویی که در مؤلفه شباهت‌سنج مبتنی بر نشانی URL مشابه تشخیص داده به‌عنوان ورودی‌های الگوریتم شباهت‌سنج در نظر گرفته خواهند شد. در این الگوریتم، ابتدا، مجموعه بردارهای حرکت به‌عنوان دنباله‌ای از بردار حرکت در نظر گرفته می‌شود. همان‌طور که مشخص است، ویدئوی کوچک‌تر دنباله‌ای با طول کوچکتر (۳ ثانیه) خواهد بود؛ سپس سعی می‌کنیم تا بیشترین میزان شباهت دنباله کوچک‌تر را با دنباله بزرگ‌تر پیدا کنیم. پس از این مرحله نتیجه به‌دست آمده را به‌عنوان نتیجه میزان شباهت دو ویدئو گزارش خواهیم کرد. روند پیدا کردن بیشترین میزان شباهت بدین صورت است که پنجره‌ای به طول ویدئو با دنباله کوچکتر در نظر می‌گیریم. در این پنجره میزان شباهت دو دنباله را محاسبه و نگه می‌داریم. این عمل را به میزان اختلاف طول دو دنباله تکرار می‌کنیم و در انتها بیشترین میزان شباهت، میزان شباهت دو ویدئو در نظر گرفته خواهد شد. در شکل (۲) الگوریتم کلی ارزیابی نشان داده شده است.

۳-۱-۳. در صورت وجود تشابه بین نشانی URLها، به مقدار آرای نتایج مربوطه یکی اضافه خواهد شد. و روند الگوریتم پس از استخراج بردار حرکت به گام شماره پنج می‌رود.

۳-۱-۲. در مؤلفه استخراج بردار حرکت، فایل‌های ویدئویی، یکپارچه شده، سپس به ویدئوهای سه‌ثانیه‌ای شکسته می‌شوند؛ پس از آن استخراج بردار حرکت توسط نرم‌افزار تغییر یافته JM-H264 به صورت موازی برای هر یک از فایل‌های شکسته شده انجام می‌شود. از این رو، به‌ازای هر نتیجه (ویدئو)، چندین فایل متنی حاوی بردار حرکت هر یک از ویدئوهای سه‌ثانیه‌ای ذخیره خواهد شد.

۴-۱. در مؤلفه شباهت‌سنج مبتنی بر محتوا، بردارهای حرکت مربوط به یک فایل ویدئو بازخوانی و عملیات زیر برای آن‌ها تکرار می‌شود.

۴-۱. بردارهای حرکت بازخوانی شده به‌صورت دنباله‌ای در نظر گرفته شده و عمل مقایسه به‌ازای هر دو نتیجه از جویش‌گرهای متفاوت که در مؤلفه شباهت‌سنج مبتنی بر نشانی URL مشابه تشخیص داده نشده‌اند، انجام می‌شود. این عمل تنها برای هر دو نتیجه‌ای که در مؤلفه شباهت‌سنج مبتنی بر نشانی URL مشابه تشخیص داده نشده‌اند، انجام می‌گیرد.

۴-۲. در صورتی که شباهت یک نتیجه با نتیجه دیگر از حد آستانه تعیین شده T بالاتر باشد، به مقدار آرای نتایج مربوطه اضافه خواهد شد.

۵. مرتبط بودن هر یک از نتایج محاسبه می‌شود.

۵-۱. در صورتی که مقدار آرای یک نتیجه بزرگتر یا مساوی ۲ باشد، نتیجه مورد نظر به‌عنوان یک نتیجه مرتبط با پرس‌وجو علامت‌گذاری می‌شود.

۶. معیارهای ارزیابی نظیر فراخوانی، دقت و nDCG برای هر جویش‌گر به‌ازای یک پرس‌وجو محاسبه می‌شود.

(شکل-۲): الگوریتم سامانه ارزیابی خودکار جویش‌گرهای

ویدئویی مبتنی بر تجمیع آرا

(Figure-2): Algorithm for the automated evaluation of video search engines based on aggregation of votes

خاطر نشان می‌شود ماژول زیرساخت دقت، وظیفه دارد تا به‌عنوان یک پارسر عمل کرده و نتایج بازگشتی به‌ازای یک پرس‌وجو را از هر جویش‌گر استخراج و برای بارگزاری آماده‌سازی می‌کند. همچنین پس از انجام ارزیابی‌های ذکر شده نتیجه ارزیابی‌ها به این ماژول ارسال می‌شود تا معیارهای دقت محاسبه و گزارش شوند.

برای درک بهتر روند صورت گرفته، فرض کنیم V_1 ویدئو با دنباله کوچک‌تر و V_2 ویدئو با دنباله بزرگ‌تر هستند که تفاوت طول آن‌ها n ثانیه است. میزان شباهت این دو ویدئو به‌صورت زیر محاسبه می‌شود:

$$Sim(V_1, V_2) = \max_i^n Sim(V_{1,i}, V_2) \quad (1)$$

۱. ابتدا فایل پیکربندی که درقبل توسط کاربر سامانه ارزیابی خودکار ایجاد شده است خوانده و تنظیمات لازم برای پارامترهای موجود در سامانه صورت می‌گیرد.

۲. به‌ازای هر پرس‌وجو از پایگاه داده پرس‌وجوها، نتایج جویش‌گرهای مختلف توسط زیرساخت دقت جمع‌آوری شده، فایل ویدئوی آن داندلود و یک مخزن که حاوی تمامی نتایج است، ایجاد می‌شود.

۳. به‌ازای نتایج هر جویش‌گر مراحل زیر دنبال می‌شود:

۳-۱. به‌ازای هر نتیجه مراحل زیر دنبال می‌شود:

۳-۱-۱. در مؤلفه شباهت‌سنج، هر نتیجه با نتایج بازگشتی از سایر جویش‌گرها براساس نشانی URL آن‌ها مورد مقایسه و شباهت‌سنجی قرار می‌گیرد.

۳-۱-۱-۱. نشانی URLها نرمال شده و موارد معمول آن حذف می‌شود.

۳-۱-۱-۲. نشانی URLها به دو قسمت دامنه نشانی و پارامترها تقسیم می‌شود. و هر قسمت به‌صورت جداگانه مورد مقایسه قرار می‌گیرد.

که روش پیشنهادی قادر است در تمامی موارد به غیر از هنگامی که دو ویدئوی یکسان نسبت طول به عرض متفاوت دارند، شباهت‌سنجی را به‌درستی انجام دهد. چون در ویدئوهای یکسان با نسبت طول به عرض متفاوت و حتی با وضوح متفاوت هر چند که حرکات به‌طور کامل مشابه هستند، ولی بردارهای حرکت دارای ابعاد متفاوتی خواهند بود برای پیدا کردن درک بهتر از این موضوع فرض کنید هر کدام از ابعاد یک ویدئو نصف ویدئوی به‌طور کامل مشابه آن باشد، در این صورت با تقریب خوبی می‌توانیم بگوییم که تمامی بردارهای حرکت این ویدئو دارای ابعادی نصف بردارهای حرکت ویدئوی مشابه هستند و بنابراین تمامی روش‌های مقایسه بر اساس بردارهای حرکت در مورد این‌گونه ویدئوها با مشکل روبه‌رو می‌شوند. به‌منظور ارزیابی روش ارائه‌شده، در این بخش، آزمایشی ترتیب داده شده که در آن نتایج حاصل از ارزیابی خودکار دقت جویش گره‌های ویدئویی با نتایج حاصل از ارزیابی انسانی همین جویش‌گرها مورد مقایسه قرار گرفته است.

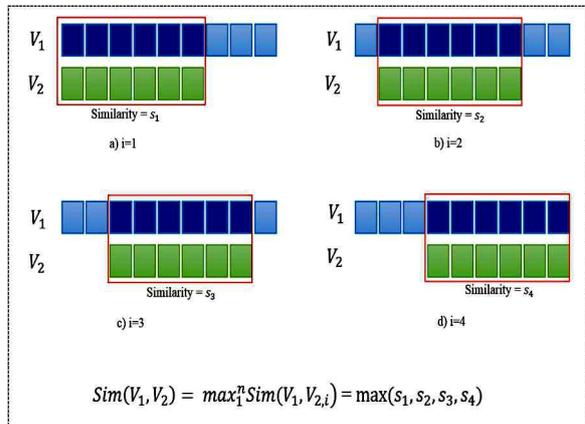
بدین منظور، مجموعه‌ای متشکل از ده کاربر انسانی در نظر گرفته شده و از هر یک از آنها دو الی سه پرس‌وجوی مختلف در ده موضوع متفاوت مانند علوم، رایانه، سرگرمی و هنر، کودکان، تجارت، تاریخ، ورزش، سلامت، فرهنگ و اجتماع و اخبار دریافت شده است؛ سپس این پرس‌وجوها در اختیار ده کاربر انسانی قرار گرفت تا توسط آنها، سه جویش‌گر گوگل، بینگ و پارس‌جو را مورد ارزیابی قرار دهند. برای هر پرس‌وجو، پنج نتیجه نخست هر موتور جستجو دریافت شده و برای ارزیابی در اختیار کاربران قرار گرفته است. به‌منظور فراهم کردن امکان مقایسه نتایج ارزیابی خودکار و انسانی جویش‌گرها به‌کمک مجموعه واحدی از پرس‌وجوها، پرس‌وجوهای انتخابی نیز به‌عنوان ورودی سامانه ارزیابی خودکار جویش‌گرهای ویدئویی مبتنی بر تجمیع آرا مورد استفاده قرار گرفته شده است.

به‌منظور تعیین میزان دقت نتایج بازگشتی از جویش‌گرها، در مقایسه‌ها جویش‌گرها از معیارهای اصلی Precision، Recall، MAP و همچنین nDCG محاسبه شده طبق فرمول (۲) استفاده شده است.

$$nDCG_p = \frac{DCG_p}{IDCG_p} \quad (2)$$

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2^{i+1}}$$

که در آن $V_{1,i}$ بخشی از V_1 است که از ثانیه i -ام شروع شده و دارای طولی برابر V_2 است. (شکل-۳) ۳- میزان جابه‌جایی ویدئوی V_2 از ابتدای ویدئوی V_1 را بر حسب ثانیه نشان می‌دهد. در این شکل اختلاف دو ویدئو چهار واحد است در نتیجه V_2 را چهار مرتبه با قسمت‌های مختلف V_1 مقایسه می‌کنیم.



(شکل-۳): نحوه محاسبه شباهت‌سنجی مابین دو دنباله از دو

فایل ویدئویی

(Figure-3): How to calculate the similarity between two sequences of two video files

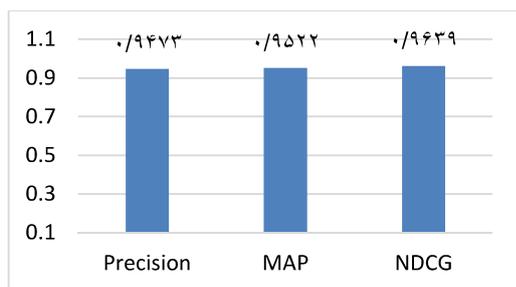
در روش پیشنهادی، پس از مشخص شدن شباهت بین نتایج بازگشتی از جویش‌گرهای متفاوت، سطح ارتباط آن‌ها در سه سطح مختلف: «نامرتب» (۰)، «کمی مرتب» (۱) و «مرتب» (۲) سطح‌بندی می‌شود. از آنجایی که جویش‌گر گوگل در حال حاضر بزرگ‌ترین و بهترین جویش‌گر موجود در دنیا است و بیش‌تر جویش‌گرها با آن مقایسه شده و همچنین سعی در رسیدن به کارکرد مشابه به آن را دارند، پنج نتیجه نخست این جویش‌گر به‌صورت پیش‌فرض «کمی مرتب» در نظر می‌گیریم و باقی نتایج به همراه نتایج سایر جویش‌گرهای دیگر، اگر طبق مؤلفه شباهت‌سنجی روش پیشنهادی شبیه بودند به‌عنوان «مرتب» و اگر شبیه نبودند به‌عنوان «نامرتب» در نظر گرفته خواهند شد.

۴- نتایج

برای مشخص کردن نقاط قوت و ضعف روش پیشنهادی در پیدا کردن ویدئوهای مشابه مجموعه‌ای از سی ویدئو را که در یکی از موارد سیاه و سفیدبودن، داشتن زیرنویس، داشتن لوگو، برش زمانی، نقطه آغازین و نسبت طول به عرض با یکدیگر متفاوت بودند، جمع‌آوری کرده و عمل شباهت‌سنجی را بر روی آن‌ها انجام دادیم. از نتایج به‌دست‌آمده مشخص شد

برای محاسبه معیار همبستگی مورد استفاده قرار گرفته است تا بتوان با بهره‌بردن از معیار پیرسن اطمینان لازم جهت قابل اعتماد بودن نتایج را به دست آورد.

این نتایج نشان می‌دهند که میزان همبستگی در سطح کلی مابین دو روش ارزیابی خودکار و انسانی برای هر چهار معیار میزان قابل قبولی است. از این رو، به دلیل وجود همبستگی بالا مابین دو روش ارزیابی، می‌توان ادعا کرد که نتایج به دست آمده از روش ارزیابی خودکار پیشنهاد شده در این مقاله مناسب و قابل اتکا است.



(شکل-۵): میزان همبستگی نتایج حاصل از سامانه ارزیابی انسانی و ارزیابی خودکار ارائه شده

(Figure-5): The correlation between the results of the human assessment system and the automated evaluation

۵- جمع‌بندی

در این پژوهش ابتدا به مسئله لزوم وجود جویش‌گرها به خصوص جویش‌گرهای ویدئویی پرداخته شد. همچنین بیان شد که برای ارزیابی سطح کیفی و کمی این جویش‌گر به روش‌هایی برای ارزیابی آن‌ها نیاز است. از مطالعات انجام شده مشخص شد که روش ارزیابی خاصی که مختص جویش‌گرهای ویدئویی باشد، وجود ندارد؛ از این رو روش‌های ارزیابی دیگر جویش‌گر مورد مطالعه قرار گرفت تا الهام‌بخش روش پیشنهادی ما برای ارزیابی جویش‌گرهای ویدئویی باشد. روش پیشنهادی در این مقاله براساس نشانی ویدئوها و بردارهای حرکت فریم‌های آن‌ها عمل شباهت‌سنجی مابین دو ویدئو را انجام می‌دهد. و در گام بعد با مقایسه ویدئوهای هر دو جویش‌گر ویدئوهای مرتبط را از نامرتبط تشخیص داده و در انتها معیارهای متداول را برای هر کدام از جویش‌گرها محاسبه می‌کند. نتایج به دست آمده از این آزمایش‌ها بیان داشت که روش خودکار ارائه شده از همبستگی بالایی با روش انسانی دارد که این مورد بیان می‌دارد که می‌توان از روش خودکار ارزیابی جویش‌گر در روند ارزیابی جویش‌گرهای ویدئویی بهره برد.

² IDCG

در جدول (۱) نتایج حاصل از این مقایسه به ترتیب بر اساس معیارهای ذکر شده در بالا نمایش داده شده است. همچنین، در شکل (۵) نتایج به دست آمده از مقایسه جویش‌گرهای ویدئویی توسط سامانه ارزیابی خودکار در مقایسه با ارزیابی انسانی بر مبنای معیار nDCG به نمایش در آمده است.

(جدول-۱): نتایج حاصل از ارزیابی خودکار جویش‌گرهای ویدئویی گوگل، بینگ و پارس‌جو

(Table-1): The results of the automatic evaluation of Google Video Search Engines, Bing and Parsijoo

	Google	Bing	Parsijoo
Recall	0.24	0.29	0.17
Precision	0.79	0.59	0.38
MAP	0.70	0.67	0.48
nDCG	0.85	0.65	0.45



(شکل-۴): نتایج حاصل از ارزیابی خودکار و ارزیابی انسانی جویش‌گرهای ویدئویی بر اساس معیار nDCG

(Figure-4): The results of automated evaluation and human evaluation of video search engines based on nDCG

نکته جالب توجه آن است که رتبه‌بندی جویش‌گرها در هر دو ارزیابی خودکار و انسانی یکسان است. به منظور تعیین میزان همبستگی مابین نتایج به دست آمده از ارزیابی خودکار و ارزیابی انسانی، از معیار همبستگی پیرسن^۱ در سطح کلی استفاده شده است؛ به صورتی که، میزان همبستگی مابین مقادیر به دست آمده برای معیار nDCG به عنوان مهم‌ترین معیار دقت در ارزیابی دقت جویش‌گر، در نظر گرفته شده است. نتایج به دست آمده برای همبستگی در شکل (۵) نشان داده شده است. برای ارزیابی خودکار پنج نتیجه نخست تمامی جویش‌گرها با یکدیگر مقایسه شده‌اند. این نکته را باید در نظر گرفت که در محاسبه nDCG دو روش ارزیابی خودکار و انسانی هر کدام مرجع ایده‌آل^۲ مخصوص به خود را دارند؛ پس نمی‌توان این دو را به صورت مستقیم با یکدیگر مقایسه کرد. باید به این نکته توجه داشت که ریزنتایج nDCG به ازای تک تک پرس‌وجوها

¹ Pearson

symposium on Applied computing, 2003, pp. 811-816.

- [9] J. Callan, M. Connell, and A. Du, "Automatic discovery of language models for text databases," in ACM SIGMOD Record, 1999, pp. 479-490.
- [10] A. Chowdhury and I. Soboroff, "Automatic evaluation of world wide web search services," in Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, 2002, pp. 421-422.
- [11] S. M. Beitzel, E. C. Jensen, A. Chowdhury, and D. Grossman, "Using titles and category names from editor-driven taxonomies for automatic evaluation," in Proceedings of the twelfth international conference on Information and knowledge management, 2003, pp. 17-23.
- [12] F. Can, R. Nuray, and A. B. Sevdik, "Automatic performance evaluation of Web search engines," Information processing & management, vol. 40, pp. 495-514, 2004.
- [13] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005, pp. 154-161.
- [14] Y. Liu, Y. Fu, M. Zhang, S. Ma, and L. Ru, "Automatic search engine performance evaluation with click-through data analysis," in Proceedings of the 16th international conference on World Wide Web, 2007, pp. 1133-1134.
- [15] Y. Liu, M. Zhang, L. Ru, and S. Ma, "Automatic query type identification based on click through information," in Information Retrieval Technology, ed: Springer, 2006, pp. 593-600.
- [16] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," ed: Citeseer, 2003.
- [17] G. Mood, "Boes, Introduction to the theory of statistics," McCraw-Hill Statistics Series, 1974.
- [18] H. Sharma and B. J. Jansen, "Automated evaluation of search engine performance via implicit user feedback," in Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005, pp. 649-650.
- [19] R. Ali and M. S. Beg, "Automatic performance evaluation of web search systems using rough set based rank aggregation," in Proceedings of the First International Conference on Intelligent

6- References

۶- مراجع

- [۱] موسوی سبحان، عظیم‌زاده معصومه، محمودی مریم، یاری علیرضا، ارائه چارچوبی جامع و کارا برای ارزیابی موتورهای جستجوی فارسی، هجدهمین کنفرانس ملی سالیانه انجمن کامپیوتر، تهران، اسفند ۱۳۹۱.
- [1] Moosavi Sobhan, Azimzadeh Masoumeh, Mahmoodi Maryam, Yari Alireza, A Comprehensive and Effective Framework for the Assessment of Persian Search Engines, The 18th Annual National Conference of the Computer Society, Tehran, Esfand 1391.
- [۲] عظیم‌زاده معصومه، سموری شهریار، یاری علیرضا، بررسی و مقایسه کیفی موتورهای جستجو در حوزه وب فارسی، هجدهمین کنفرانس ملی سالیانه انجمن کامپیوتر، تهران، اسفند ۱۳۹۱.
- [2] Azimzadeh Masoumeh, Somoori Shahriar, Yari Alireza, Qualitative and qualitative comparison of search engines in the Persian Web domain, 18th National Computer Society Conference, Tehran, Esfand 1391.
- [3] R. Badie, M. Azimzadeh, M.S. Zahedi, S. Samuri, "Automatic evaluation of search engines: Using webpages' content, web graph link structure and websites' popularity" Seventh International Symposium on Telecommunications (IST2014), September 09-11, 2014.
- [4] Maryam Mahmoudy, Mohammad Sadegh zahedi, Masomeh Azimzadeh, "Evaluating the retrieval effectiveness of search engines using Persian navigational queries", Seventh International Symposium on Telecommunications (IST2014), September 09-11, 2014.
- [5] I. Soboroff, C. Nicholas, and P. Cahan, "Ranking retrieval systems without relevance judgments," in Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, 2001, pp. 66-73.
- [6] S. P. Harter, "Variations in relevance assessments and the measurement of retrieval effectiveness," JASIS, vol. 47, pp. 37-49, 1996.
- [7] A. Spink and H. Greisdorf, "Regions and levels: measuring and mapping users' relevance judgments," Journal of the American Society for Information science and Technology, vol. 52, pp. 161-173, 2001.
- [8] S. Wu and F. Crestani, "Methods for ranking information retrieval systems without relevance judgments," in Proceedings of the 2003 ACM

فناوری اطلاعات، موسوم به مرکز تحقیقات مخابرات ایران است. زمینه‌های پژوهشی وی شامل سامانه‌های چندرسانه‌ای، پردازش و بازیابی تصویر و ویدئو و پیاده‌سازی سخت‌افزاری استانداردهای کدگذاری تصویر و ویدئو است.

نشانی رایانامه ایشان عبارت است از:
zargari@itrc.ac.ir



مژگان فرهودی دانش‌آموخته رشته مهندسی فناوری اطلاعات از دانشگاه صنعتی امیرکبیر است. وی از سال ۱۳۸۰ در زمینه‌های بازیابی اطلاعات با تمرکز بر جویس‌گرها و سامانه‌های پرسش و پاسخ، وب‌کاوی و پردازش هوشمند زبان طبیعی فعالیت دارد. ایشان مدیر قبلی گروه سکوه‌های فناوری اطلاعات بوده و در حال حاضر به‌عنوان هیئت علمی و مجری آزمایشگاه ارزیابی خدمات وب (وب آزما) در پژوهشگاه ارتباطات و فناوری اطلاعات در حال فعالیت است.

نشانی رایانامه ایشان عبارت است از:
farhoodi@itrc.ac.ir

Human Computer Interaction, 2009, pp. 344-358.

- [20] R. Nuray and F. Can, "Automatic ranking of information retrieval systems using data fusion," Information processing & management, vol. 42, pp. 595-614, 2006.
- [21] H. Sadeghi. "Automatic Performance Evaluation of Web search Engines using judgements of Meta search Engines", Online Information Review, ISSN:1468-4527, Emerald Publishing Limited, pp.957-971. (2011).
- [22] Tawileh W, Griesbaum J, Mandl T. Evaluation of five web search engines in Arabic language. Proceedings of LWA. (2010).
- [23] Lewandowski, Dirk. Evaluating the retrieval effectiveness of Web search engines using a representative query sample. Journal of the Association for Information Science and Technology (2015).
- [24] Bar-Ilan J, Levene M. A method to assess search engine results. Online Information Review 35(6), 854-868. (2011).
- [25] Keyvanpour M, alamdar F. Effective browsing of image search results via diversified visual summarization by Clustering. JSDP. 2012; 8 (2):57-74



محمد مهدی یداللهی دانش‌آموخته کارشناسی ارشد مهندسی کامپیوتر گرایش نرم‌افزار از دانشگاه تهران است. وی در حال حاضر در زمینه داده‌کاوی، بازیابی اطلاعات و مهندسی داده فعالیت دارد. مهم‌ترین فعالیت حرفه‌ای ایشان در زمینه راه‌اندازی سامانه‌های داده‌کاوی در شبکه‌های اجتماعی و همچنین تجزیه و تحلیل آن‌ها است. از فعالیت‌های فرعی ایشان نیز می‌توان به بازیابی اطلاعات در جویس‌گرها و ارزیابی آن‌ها اشاره کرد.

نشانی رایانامه ایشان عبارت است از:

mm.yadollahi@itrc.ac.ir



فرزاد زرگری مدرک کارشناسی خود را در رشته مهندسی برق از دانشگاه صنعتی شریف، کارشناسی ارشد و دکترای تخصصی خود را از دانشگاه تهران دریافت کرده است. او اینک دانشیار دپارتمان فناوری اطلاعات در پژوهشگاه ارتباطات و