**نام مقاله: استخراج خودكار عبارتهاي كليدي از متون مقاله هاي فارسي**

**نام نشريه: فصلنامه كتابداري و اطلاع رساني (اين نشريه در www.isc.gov.ir نمايه مي شود)**

**شماره نشريه: 35 \_ شماره سوم، جلد 9**

**پديدآور: علي گزني**

**مترجم:**

**چكيده**

**در پژوهش حاضر، عبارتهاي كليدي از متون مقاله‌هاي فارسي به صورت خودكار جداسازي گرديده است. استخراج عبارتها مبتني بر روشهاي آماري، نحوة توزيع واژگان، مجاورت و ... صورت پذيرفته است. سيستمي كه بر پايه پژوهش حاضر طراحي گرديده، با توجه به بازخوردهاي كاربر از قابليت يادگيري برخوردار است، با توجه به بازخوردهاي كاربر از قابليت يادگيري برخوردار است، به گونه‌اي كه در طول زمان مرتباً به كارايي آن افزوده مي‌شود. استخراج عبارتهاي كليدي مي‌تواند در پهنه گسترده‌اي از مسائل از جمله در طراحي سيستمهاي بازيابي اطلاعات، كاربر داشته باشد.**

**كليدواژه‌ها: كليدواژه، عبارتهاي كليدي، مقاله‌هاي فني**

**هدف از استخراج عبارتهاي كليدي[2] متون مقاله‌ها[3]، تسريع در تعيين حوزه موضوعي مقاله‌هاي چاپ شده مي‌باشد. اين كار فكري است و به آشنايي كلي با موضوع مورد نظر، مهارت و تجربه نياز دارد. بنابراين، انرژي‌اي كه مي‌تواند در راههاي ديگري صرف گردد، بايد در راه تسهيل دسترسي ديگران به اطلاعات مصرف گردد. افزايش روزافزون تعداد مقاله‌هاي فني و حجم انبوه اطلاعات از جمله مواردي هستند كه بر اهميت اين مسئله مي‌افزايند. به علاوه، استخراج عبارتهاي كليدي در اغلب اوقات تحت تأثير پيش‌زمينه‌هاي قبلي، عقايد شخصي يا تمايل فرد به انجام سريعتر كار قرار مي‌گيرد. بنابراين، كيفيت كار در بين افراد مختلف و حتي افراد يكسان در زمانهاي مختلف، متفاوت است (مهراد، 1373).**

**مقاله حاضر به تشريح پژوهش انجام شده در زمينه استخراج عبارتهاي كليدي متون مقاله‌هاي فني مي‌پردازد. در روش مورد نظر، متن مقاله‌ در قالب ماشين‌خوان به برنامه رايانه‌اي طراحي شده تحويل مي‌گردد. برنامه در روشي قابل مقايسه با آنچه توسط انسان خوانده مي‌شود، اطلاعات را مورد پردازش قرار مي‌دهد و از ميان كليه واژگان موجود در مقاله، عبارتهاي كليدي را كه نشان‌دهنده مرتبط‌ترين عبارتها و اطلاعات مقاله مي‌باشند، استخراج مي‌كند. اين عبارتها مي‌توانند به عنوان راهنمايي براي قضاوت در مورد متن مقاله مورد استفاده قرار گيرند. بنابراين، عبارتهاي كليدي مستقيماً از ميان نوشته‌هاي نويسنده انتخاب مي‌گردند.**

**عبارتهاي كليدي**

**عبارتهاي كليدي متن نشان‌دهندة مفاهيم و موضوع مقاله بوده، مي‌توانند در موارد زير مورد استفاده قرار گيرند:**

**1. استخراج خودكار عبارتهاي كليدي، يك متن بلند را به خلاصه‌اي كوتاه تبديل مي‌كند. به عنوان مثال، مي‌توان از اين ويژگي در مرورگرهاي وب[4] استفاده كرد؛ بدين ترتيب كه كاربر با فشار دادن يك دكمه، عبارتهاي كليدي متن را مشاهده و در نتيجه به حوزة موضوعي متن مورد نظر پي مي‌برد. براي مثال، شكل1 نتيجه يك جستجو از موتور كاوش Google را نشان مي‌دهد. پيوند Key Phrases در اين صفحه اضافه شده است. با كليك كردن بر روي اين گزينه عبارات كليدي متن نمايش داده مي‌شوند.**

**Information Technology Association of America Website**

**Information Technology Association of America Headlines, … April 24, 2003-Organizing for Results: Information Technology Structures and Staffing WEBCAST …**

**Description: Trade association representing the broad spectrum of the world-leading US IT industry.**

**That's why …**

**Category: Business> Information Technology > Associations**

**www.itaa.org/-55k-Cached-Similar Pages Key Phrases**

**شكل1. عبارتهاي كليدي در نمايش جستجو**

**2. عبارتهاي كليدي مي‌توانند به عنوان قسمتي از نتايج جستجو همراه با ساير مشخصه‌هاي متن بازيابي شده (همانند عنوان، قسمتهايي از متن، URL و ...) يا به جاي آنها نمايش داده شوند. در شكل1 مي‌توان تصور كرد كه به جاي پيوند Key Phrases عبارتهاي كليدي همراه با ساير قسمتهاي جستجو شده، نمايش داده شوند.**

**3. در مواردي كه به مشخصه‌هايي بيش از نامگذاري صرف به منظور درك سريعتر متن نياز داريم، عبارتهاي كليدي مي‌توانند مفيد باشند. به عنوان مثال، اگر نام يك فايل يا نامه الكترونيكي[5] به عنوان برچسب[6] با عبارتهاي كليدي ادغام گردند، حالت بهتري را ايجاد مي‌كنند. در اين حالت، مشاهده عبارتهاي كليدي همراه با عنوان، به فهم محتواي نامه كمك بيشتري مي‌كند.**

**4. برجسته‌كردن[7] عبارتهاي كليدي در متون الكترونيكي مي‌تواند به مرور سريع و اجمالي متن كمك كند.**

**5. كمك به نويسنده يا ويراستار در تخصيص عبارتهاي كليدي به متن. انجام اين كار به صورت خودكار مي‌تواند به عنوان يك استاندارد، نوعي يكدستي و مطابقت نوشته با كاركرد سيستم بازيابي اطلاعات و در نتيجه اطلاع‌رساني صحيح‌تر را به همراه داشته باشد.**

**6. در مواردي كه با مشكل پهناي خط يا مطابق با اصول نمايش گرافيكي اطلاعات با محدوديت فضاي نمايشي[8] مواجه هستيم، نمايش عبارتهاي كليدي بسيار مفيد است. اصولاً در كشورهاي جهان سوم كه خطوط از سرعت و پهناي خط پاييني برخوردارند و در مكانهايي كه محدوديت فيزيكي وجود دارد، همانند صفحات نمايش رايانه (اندازه ثابت)، حالت مطلوبتري را ايجاد مي‌كند.**

**7. استخراج خودكار عبارتهاي نمايه‌اي متون نشريات و صفحات وب، خواندن و جستجوي اطلاعات نشريات را براي خوانندگان تسهيل مي‌كند.**

**8 . حضور عبارتهاي كليدي در نتايج جستجو مي‌تواند به اصلاح و تعريف مجدد فرمول جستجو و حتي تغيير ديدگاه كاربران از ساختار موجود در يك زمينه خاص كمك كند؛ يعني كاربران مي‌توانند با افزودن، حذف واژگان دامنه جستجو را محدودتر كرده، ضريب دقت را بالاتر ببرند. در نتيجه، بالابردن ضريب دقت[9] يا با گسترده‌تركردن دامنة جستجو و در نتيجه به بالابردن ضريب بازيابي[10] كمك مي‌كند. بنابراين مي‌توان عبارتهاي كليدي را به عنوان جزئي لازم براي سيستمهاي بازيابي اطلاعات معرفي كرد.**

**9. در مفاهيم سازماندهي اطلاعات در سيستم‌هاي بازيابي اطلاعات[11] (1) مي‌توان به گونه‌اي مؤثر از عبارتهاي كليدي در خوشه‌بندي[12] و طبقه‌بندي مدارك استفاده كرد.**

**تعيين اهميت واژگان**

**براي تعيين عبارتهايي كه مي‌توانند به عنوان عبارتهاي كليدي متن مورد استفاده قرار گيرند، به يك معيار براي مقايسه و نمره‌گذاري محتواي اطلاعاتي مقاله نياز داريم. عاملي كه رتبه اهميت هر عبارت به وسيله آن تعيين مي‌گردد، تجزيه و تحليل كلمات موجود در جملات مي‌باشد. آنچه در مقاله حاضر براي اندازه‌گيري رتبه اهميت يك كلمه مناسب تشخيص داده شده و پيشنهاد مي‌شود تعداد رخداد كلمه، مجاورت مكاني اين واژگان با يكديگر و موقعيت مكاني آنها در مقاله نسبت به هم مي‌باشد. نكته‌اي كه در اينجا بايد به آن اشاره شود اينكه براي رايانه، واژگان موجود در متون همانند يك سري اشياي فيزيكي مي‌باشند. ماشين مي‌تواند تشخيص دهد كه آيا بعضي اشيا با هم مشابه اند يا نه، ماشين مي‌تواند اين قبيل يافته‌ها را به خاطر داشته باشد و مي‌تواند بر روي آنهايي كه قابل شمارش هستند، محاسبه انجام دهد. ماشين تمام اين كارها را با استفاده از يك رويه از قبل برنامه‌ريزي شده انجام مي‌دهد. در اينجا از هوش انسان فقط براي تهيه اين برنامه‌ها استفاده مي‌گردد.**

**دليل استفاده از تعداد رخداد براي اندازه‌گيري رتبه اهميت، بر اين باور استوار است كه نويسنده معمولاً از واژگان معيني براي پيشبرد، بحث يا تشريح دقيق جنبه‌هاي مختلف موضوع موردنظر استفاده و آنها را تكرار مي‌كند. تعداد رخداد هر واژه مي‌تواند به عنوان عامل تعيين درجه اهميت واژگان مورد استفاده قرار گيرد. در غالب اوقات، واژه‌هاي معيني وجود دارند كه با يكديگر يك گروه را تشكيل مي‌دهند، بايد به اين واژه‌ها رتبه اهميت بالاتري اختصاص داد. در اين ميان، بعضي كلمات براي نشان دادن ميزان ارتباط واژه‌ها با يكديگر و گروه‌بندي آنها به كار مي‌روند. به اين واژه‌هاي رابط، نمره‌اي اختصاص داده نمي‌شود. اين قبيل كلمات عمومي را مي‌توان با تعريف يك سياهه بازدارنده و تكميل اين سياهه در طول زمان، حذف كرد (دياني، 1381). به همين منظور، علاوه بر اينكه مي‌توان حد بالايي[13] را براي رخداد واژگان در نظر گرفت، يك سياهه بازدارنده كه قابليت افزايش و كاهش آن توسط كاربر وجود دارد، در سيستم گنجانده مي‌شود كه از اين طريق اثر بعضي از واژگان را خنثي و آنها را ناديده گرفت. تعيين حد بالا بر اين حقيقت استوار است كه واژگان عمومي نظير و، به، با و ... از رخداد بسيار بالايي در مقايسه با ساير كلمات موجود در مقاله برخوردارند.**

**در تعيين رتبه اهميت، از مسائل زبان‌شناختي همانند گرامر استفاده نمي‌شود. در يك نگاه كلي بايد گفت، در روش حاضر حتي بين شكلهاي مختلف كلمات نيز تفاوتي گذاشته نشده است. بنابراين، حالتهاي مختلف كلمات جستجو، جستجوهاي، جستجوها با يكديگر يكسان است. در روش حاضر، به ارتباطهاي منطقي و معنايي مورد نظر نويسنده توجه‌اي نشده است. به بيان ديگر، پس از بررسي متن، فهرستي از كلمات متن با يك نظم نزولي، بر حسب تعداد رخدادشان ايجاد و مرتب‌سازي مي‌شوند.**

**رويّه‌اي كه در روش حاضر مورد استفاده قرار مي‌گيرد، بسيار ساده بوده و از نظر اقتصادي مقرون به صرفه است. اين در حالي است كه هر چه روش پيچيده‌تر باشد، ماشين بايد سلسله عمليات بيشتري را متحمل گردد، كه اين خود باعث افزايش هزينه پردازشها خواهد شد. دليل انتخاب يك روش ساده براي كار بر روي مقاله‌هاي فني اين است كه با توجه به ماهيت مقاله‌هاي فني، احتمال بسيار كمي وجود دارد كه يك واژه براي نشان دادن بيش از يك مفهوم به كار رفته باشد يا نويسنده از كلمات متفاوتي براي نشان دادن بيش از يك مفهوم استفاده كند. حتي اگر يك نويسنده به دلايل نگارشي به انتخاب واژه‌هاي مترادف بپردازد، به زودي از اين كار خسته شده و دوباره به استفاده از كلمه‌اي كه اولين بار براي بيان مفهوم خود از آن استفاده كرده است، مي‌پردازد. فهرستي از واژه‌هاي به دست آمده مطابق روش حاضر در نمودار شكل (1) قابل مشاهده مي‌باشد. چنانكه قبلاً نيز به آن اشاره كرديم، كلمات عمومي رخداد بالايي دارند كه اين خود موجب اختلال در سيستم مي‌گردد. امكان كاهش تأثير اين اختلال، با ذخيره يك سياهه از واژگان عمومي به صورت جداگانه، مقايسه اين واژگان با واژگان متن و حذف واژگان عمومي از متن، وجود دارد. يك روش ساده‌تر اين است كه براي حصول اطمينان با استفاده از روشهاي آماري، حدي را براي بالاترين رخداد تعيين كنيم. اگر خط A در شكل (1) نشان‌دهنده اين حد باشد، آنگاه تنها واژگاني كه در سمت راست اين خط مي‌باشند، با اهميت در نظر گرفته مي‌شوند. به دليل اينكه ميزان رخداد به عنوان يك معيار براي تعيين اهميت واژگان تعيين شده است، بايد حدّ پاييني نيز در اين رابطه در نظر گرفته شود. در اين قسمت، خط C نشان‌دهنده اين حد باشد. تعيين يك محل مناسب براي اين دو خط تجربي بوده و با توجه به بررسي نمونه‌هاي مقاله‌هاي چاپ شده در سطح وسيع، قابل تعيين مي‌باشد. اين امكان در سيستم حاضر وجود دارد كه اين محل (ميزان حد بالا و پايين رخداد كلمات) براي تغيير خصوصيات خروجي حاصل، تغيير داده شود.**

**گهگاه ديده شده واژگاني غير عمومي نيز، در سمت چپ خط A ظاهر شده‌اند. اگر برنامه به خوبي فرمول‌بندي شده باشد، محل واژه‌ها در نمودار مي‌تواند بر از دست رفتن حد تمايز دلالت كند. تعيين يك مقدار براي حد، در يك رشته مي‌تواند براي شاخه‌هاي خاصّ آن رشته يا حتي در رشته‌هاي ديگر نيز كاربرد داشته باشد. مي‌توان به منظور افزايش كيفيت، حد تمايز را افزايش داد. در بعضي حالات تعدادي از واژه‌هاي عمومي در سمت راست خط A قرار مي‌گيرند. در اين حالت، تعداد اين واژگان كم بوده و به علاوه با كمك سياهه بازدارنده، تأثير آنها به حداقل خواهد رسيد.**

**A C**

**رخداد**

**واژه‌ها**

**شكل1. نمودار رخداد ـ كلمات بر روي محور افقي تك تك واژه‌ها**

**به ترتيب تعداد رخداد آنها نمايش داده شده‌اند.**

**استخراج عبارتهاي كليدي**

**در روش حاضر با معاني واژه‌ها كاري نداريم و پيوند و تركيب كلمات با يكديگر به صورت قوي مورد بحث قرار نمي‌گيرد. البته، از اين خصوصيت كه فاصلة كمتر واژه‌ها از يكديگر بر جنبه خاصي از يك موضوع دلالت مي‌كند، استفاده شده است. بنابراين، رخداد بالاي كلمات مختلف در مجاورت يكديگر، نشان دهندة احتمال مرتبط بودن اين واژه‌ها با محتواي مقاله‌ است. رتبه اهميت مجاورت مي‌تواند بر اساس خصوصيات زباني نوشته‌ها متفاوت باشد. به صورت فيزيكي، واژه‌هايي كه براي بيان تصورهاي متجانس ذهني به كار مي‌روند از لحاظ مكاني در موقعيت نزديكتري نسبت به هم قرار دارند. تقسيم متن نوشته به جملات، پاراگرافها، فصلها و ... از راههاي ديگري است كه در آن درجه همبستگي تصورات با يكديگر مشخص‌تر مي‌شود.**

**مرحله بعدي تركيب واژگاني است كه حروف آغازين آنها با يكديگر مشابه است؛ همانند جستجو، جستجوها، جستجوهاي. اين كار با يك تحليل ساده آماري با مقايسه هر جفت واژه به صورت حرف به حرف در فهرست الفبايي واژه‌ها صورت مي‌گيرد. اگر تعداد حروف مشابه آغازين برابر عدد چهار بود، دو واژه با يكديگر مشابه فرض مي‌شوند. متناسب با ماهيت زبان فارسي، اين عدد مي‌تواند بين 3 تا 5 قابل تغيير باشد. هرچند در اين روش تطبيق، امكان خطا وجود دارد، اما به نظر نمي‌رسد خطاها بيش از 5% باشد. بنابراين، در نتيجه نهايي تأثيري نخواهد داشت. در اينجا بايد اين نكته را خاطر نشان كرد كه اين مقدار در برنامه قابل تنظيم و تغيير مي‌باشد. در مرحله بعدي ماشين، تعداد رخداد واژه‌هاي مشابه را مورد محاسبه قرار مي‌دهد. مطابق با حد پاييني تعيين شده براي رخداد كلمات، واژه‌هايي كه رخداد آنها از اين ميزان كمتر باشد، حذف مي‌شوند و واژه‌هاي باقيمانده دوباره مرتب‌سازي مي‌شوند. واژه‌هاي باقيمانده، وضعيت كلمات با اهميت را نشان مي‌دهند.**

**در نهايت، عبارتهاي كليدي از ميان كلمات با رخداد بالايي كه در كنار يكديگر در سطح مقاله تكرار شده‌اند مشخص و با توجه به تكرارشان به هر كدام نمره‌اي اختصاص داده مي‌شود و كاربر مي‌تواند سياهه اين عبارتها همراه با نمره‌هاي آنها را مشاهده كند. مي‌توان مجموعه عبارتهايي را كه تعداد تكرار آنها كمتر از حدّ معيني مي‌باشد، حذف كرد. انجام اين كار با توجه به سياست سيستم در افزايش ضريب دقت يا بازيابي صورت مي‌گيرد. تعيين تعداد واژه‌هايي كه مي‌توانند در ميان واژگان موجود در عبارتها تكرار شوند و همچنين تعريف سياهه‌اي از واژه‌هاي مجاز بين عبارت نيز در سيستم امكان‌پذير مي‌باشد. موقعيت مكاني واژگان در كلّ مقاله (مثلاً در عنوان) نيز در تعيين عبارتهاي كليدي مورد توجه قرار مي‌گيرد كه در مقاله ديگري توسط نويسنده به صورت جزئي مورد بررسي قرار مي‌گيرد.**

**فرمول حاضر بر روي 50 مقاله 290 تا 5000 واژه‌اي مورد آزمايش قرار گرفت و مبتني بر نتايج اين جرأت حاصل گرديد كه عبارتهاي استخراج شده به منظور ارزيابي در اختيار 8 نفر از متخصصان نمايه‌سازي قرار گيرد.**

**يك مثال براي استخراج عبارتهاي كليدي در شكل (3) از نشريه فصلنامه كتابخانه مركزي آستان قدس رضوي وجود دارد و عبارتهاي استخراج شده توسط سيستم در جدول (1) قابل مشاهده است.**

**سازماندهي اطلاعات در نظامهاي بازيابي اطلاعات**

**علي گزني1**

**چكيده**

**هر نظام بازيابي اطلاعات (نرم‌افزار) داراي يك مبناي خاص براي تجزيه و تحليل اطلاعات است، كه نظام بر اساس آن به تفسير اطلاعات و مطابقت بين اقلام و درخواستهاي اطلاعاتي پرداخته و بدين ترتيب بازيابي اطلاعات صورت مي‌گيرد. اين تجزيه و تحليل «سازماندهي اطلاعات» ناميده مي‌شود. بدون يك سازماندهي بهينه اطلاعات، بازيابي اطلاعات به صورت كامل و دقيق صورت نخواهد گرفت. با توجه به متفاوت بودن سياستهاي بازيابي اطلاعات بايد به صورت همزمان امكان استفاده از روشهاي خودكار و نيمه‌خودكار فراهم آورده شود. پيش‌بيني سياهه بازدارنده، ايجاد انواع واژه‌نامه‌ها مانند واژه‌نامه ريشه لغات، سياهه پسوندها، واژه‌نامه عبارات، واژه‌نامه مفاهيم، برقراري روابط سلسله مراتبي مفاهيم، ريشه‌يابي واژگان، محاسبه همبستگي و خوشه‌بندي اطلاعات همگي از امكاناتي هستند كه بايد در يك نظام بازيابي اطلاعات بهينه وجود داشته باشد. مقاله حاضر، به بررسي اين مفاهيم پرداخته است.**

**واژه‌هاي كليدي: سازماندهي اطلاعات، نظامهاي بازيابي اطلاعات، فايل واژه‌نامه، ريشه‌يابي واژگان، خوشه‌بندي اطلاعات.**

**مقدمه**

**بدون سازماندهي بهينه اطلاعات، بازيابي اطلاعات به صورت كامل و دقيق صورت نخواهد گرفت. با توجه به متفاوت بودن سياستهاي بازيابي اطلاعات بايد به صورت همزمان امكان استفاده از روشهاي خودكار و نيمه‌خودكار فراهم آورده شود. پيش‌بيني سيـاهه بـازدارنده، ايجاد انواع واژه‌‌نامه‌ها مانند واژه‌نامه ريشه لغات، سيـاهه پسـوندها،**

**1. عضو هيئت علمي كتابخانه منطقه‌اي علوم و تكنولوژي شيراز**

**شكل3. مقاله سازماندهي اطلاعات در سيستمهاي بازيابي اطلاعات**

**جدول1. عبارتهاي استخراج شده از مقاله سازماندهي اطلاعات در نظامهاي بازيابي اطلاعات**

**عبارتهاي استخراج شده**

**واژه‌نامه**

**واژه‌نامه ريشه‌يابي**

**بازيابي اطلاعات**

**سازماندهي خودكار اطلاعات**

**نظامهاي بازيابي اطلاعات**

**روش خودكار**

**سازماندهي اطلاعات**

**سازماندهي واژه‌ها**

**واژه‌نامه ريشه**

**واژه‌نامه مفاهيم**

**رخدادهاي بالاتر**

**سازماندهي واژه‌ها**

**ركورد اطلاعاتي**

**ريشه واژه**

**بازيابي مدارك**

**خوشه‌بندي اطلاعات**

**واژه‌نامه‌ عبارتهاي**

**نظام خودكار**

**نظام اطلاعاتي**

**آموزش سيستم**

**همان‌گونه كه در شكل (2) قابل مشاهده است، سيستم استخراج عبارتهاي كليدي مي‌تواند يك سير تكاملي را طي كند؛ بدين نحو كه در طول زمان با توجه به آموزشهاي كاربر، سيستم عملكرد خود را مطابق با نياز و آموزشها تغيير مي‌دهد. در سيستم حاضر، اين آموزش مي‌تواند در 4 مقوله مورد توجه قرار گيرد كه عبارتند از:**

**الف) سياهه بازدارنده عمومي**

**اين بانك شامل واژگان عمومي مشترك ميان كلية حوزه‌هاي موضوعي مي‌گردد. واژگان عمومي نظير: به، با، و، كه و ... در اين مقوله قرار مي‌گيرند. اين قبيل واژگان كه معمولاً تعداد رخداد بالايي دارند به حوزه خاصي وابسته نبوده و مي‌توانند به عنوان يك بانك عمومي در كليه شاخه‌ها تعريف شوند.**

**ب) سياهه بازدارنده حوزه‌هاي موضوعي**

**با نمايش نتايج عبارتهاي استخراج شده، كاربر با علامتگذاري، عبارتهاي مرتبط را به خروجي ارسال مي‌كند. در اين مرحله، سيستم سياهه‌اي از واژگان تأييد نشده را در بانكي تحت عنوان سياهه بازدارنده حوزه‌هاي موضوعي نگهداري مي‌كند. اطلاعات اين بانك در پردازشهاي بعدي مورد توجه قرار گرفته و عبارتهاي موجود در اين بانك به صورت خودكار از فهرست نتايج حذف مي‌گردند.**

**ج) عبارتهاي مورد تأييد عمومي يا حوزه‌هاي خاص**

**هماننـد مـورد (ب)، سيـستم سيـاهه‌اي از عبـارتهـاي تأييد شـده كاربر در حوزه‌هاي**

**مختلف را نگهداري و بر اساس آن در پردازشهاي بعدي، اين عبارتهاي به عنوان عبارتهاي تأييد شده مورد توجه قرار مي‌گيرند.**

**د) سياهه‌اي از واژگان مترادف در حوزه‌هاي موضوعي**

**كاربر با معرفي واژگان مترادف، سيستم را قادر مي‌سازد تا در پردازشها آنها را تشخيص و به عنوان يك واحد مورد محاسبه و پردازش قرار دهد.**

**نتيجه‌گيري**

**نتايج حاصل بر روي مقاله‌هاي فني نشان داد كه انتخاب خودكار عبارتهاي كليدي به نحوي كه بيانگر موضوع كلي مقاله باشند عملي است و اين عبارتها تا حدود زيادي شبيه عبارتهايي هستند كه توسط انسان از ميان نوشته انتخاب مي‌شوند. مزيت اصلي اين روش، يكدستي و يكنواختي آنهاست. به دليل دخالت نداشتن تواناييها و تمايلات انسانها و استخراج عبارتهاي با تحليل آماري كلمات به كار رفته توسط نويسنده، عبارتهاي استخراج شده از شايستگي، يكدستي و پويايي برخوردارند. هنگامي كه عبارتهاي كليدي در سطح وسيع در اختيار كاربران قرار گرفت، كاربران ياد خواهند گرفت كه چگونه آنها را درك كنند و چگونه مفاهيم مورد نظر خود را تشخيص دهند. البته، اين احتمال نيز وجود دارد كه با توجه به سبك نگارش نويسنده در گسترش مطالب، عبارتهاي درجه دو انتخاب شوند يا نتايج حاصل نامناسب باشند، اما مي‌توان حالتهاي استثنايي را به سيستم آموزش داد؛ بدين صورت كه سيستم را به شاخه‌هاي مختلف در حوزه‌هاي مختلف دانش تقسيم كرد و آنگاه سيستم رفتار خود را مطابق با آموزشهاي كاربر و همچنين كارهاي قبلي تغيير داد.**

**منابع**

**دياني، محمدحسين (1381)، سازماندهي اطلاعات، جي‌ تيلور، انتشارات كتابخانه رايانه‌اي، ص 212ـ163.**

**مهراد، جعفر (1373)، سيستمها و خدمات اطلاع‌رساني كامپيوتر، هرمان و ويسمان، انتشارات نويد، ص124ـ97.**

**S. Rose (1999) The Sunflower Visual Metaphor: A New**

**--------------------------------------------------------------------------------**

**1. عضو هيئت علمي كتابخانه منطقه‌اي علوم و تكنولوژي شيراز**

**2. Key Phrase**

**3. در رشته‌هاي فني (علوم و تكنولوژي) نويسندگان اغلب مجبورند از مجموعه واژگان خاصي براي گسترش مقاله خود استفاده كنند. استخراج عبارتهاي كليدي با استفاده از روش حاضر براي اين‌گونه مقاله‌ها، پاسخ بهتري را در بر خواهد داشت.**

**1. Web Browser**

**1. Email**

**2. Label**

**3. Highligth**

**1. تعريف علمي نمايش گرافيكي اطلاعات عبارت است از محاسبه و انتقال علايم به اشكال هندسي به صورتي قابل درك و مشاهده توسط انسان، به منظور فهم و كشف روابط پنهان موجود بين عناصر مختلف داده‌ها (2). نمايش گرافيكي اطلاعات به عنوان يك شيوه علمي و زيرشاخه‌اي از مبحث تعامل انسان و رايانه و با استفاده از تواناييهاي گرافيكي رايانه‌ها اهداف زير را دنبال مي‌كند:**

**1) بالابردن سرعت فهم و پردازش اطلاعات توسط انسان در طي فرايند ادراك و كم كردن درگيريهاي ذهني او**

**2) ايجاد و برقراري ارتباط بين اجزاي مختلف اطلاعات**

**3) انجام عملياتهاي پيچيده با اعمالي بسيار ساده**

**2. Precision**

**3. Recall**

**4. هر نظام بازاريابي اطلاعات (نرم‌افزار) داراي يك مبناي خاص براي تجزيه و تحليل اطلاعات است، كه نظام بر اساس آن به تفسير اطلاعات و مطابقت بين اقلام و درخواستهاي اطلاعاتي پرداخته و بدين ترتيب بازيابي اطلاعات صورت مي‌گيرد. اين تجزيه و تحليل «ساماندهي اطلاعات» ناميده مي‌شود.**

**5. دسته‌بندي ركوردهاي اطلاعاتي در گروههاي مختلف در كل سيستم بازيابي اطلاعات با توجه به مشخصه‌هاي مشابه و نحوه توزيع آنها به منظور طبقه‌بندي آنها را «خوشه‌بندي اطلاعات» (Clustring) مي‌گويند.**