**نام مقاله: بررسي مدل فضا برداري در بازيابي اطلاعات**

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

**شماره نشريه: 38 \_ شماره دوم، جلد 10**

**پديدآور: دكتر جعفر مهراد، مهندس سارا كليني**

**چكيده**

**بازيابي اطّلاعات، فرايند يافتن اطّلاعات (مدارك) مربوط به جستجوي كاربر در مجموعة مدارك است. با پياده سازي الگوريتمهاي متفاوت، استراتژيهاي مختلفي در بازيابي اطّلاعات وجود دارد. وجه مشترك استراتژيهاي بازيابي، يافتن مدارك مشابه با موضوع جستجوي كاربر است. يكي از الگوريتمهاي مهمي كه در بازيابي اطّلاعات، كاربرد بسيار دارد، الگوريتم فضاي برداري است كه مي كوشد تمام مدارك را در مجموعه و جستجوهاي كاربر به صورت بردارها نشان دهد و ضريب تشابه ميان بردارهاي مدرك و بردار جستجو را جهت بازيابي مدرك مربوط، محاسبه نمايد.**

**كليدواژه‌ها: بازيابي اطّلاعات، مدل فضاي برداري، فراواني اصطلاح، وزن اصطلاح، رتبه‌بندي مدارك.**

**مقدمه**

پيشرفتهاي اخير در علم الكترونيك، به توليد ابزارهاي مدرن براي ذخيره‌سازي انبوهي از اطّلاعات منجر گرديده است. انفجار اطّلاعات باعث شده است تا جامعة پژوهشگران در حوزة بازيابي اطّلاعات، امكان و شيوة فراخواني اطّلاعات درخواستي را در پنجاه سال اخير بسيار بهبود ببخشند. ‍‍[2 و 1] با سيستمهاي بازيابي اطّلاعات امروزي، امكان جستجو در چند ترابايت اطّلاعات، فقط در چند ثانيه وجود دارد [3].

نظام بازيابي اطّلاعات به سازماندهي، ذخيره سازي، بازيابي و نمايش اطّلاعات كتابشناختي مربوط است. سيستمهاي بازيابي اطّلاعات با هدف فراهم آوردن زمينة لازم براي پاسخگويي به جستجوهاي كاربر از طريق ارجاع به مدارك مربوط، طراحي مي‌گردد. در چنين محيطي، مجموعه‌اي از مدارك مانند كتابها، مقاله‌ها، گزارشهاي تحقيقاتي و... وجود دارد، به اضافة گروهي از كاربران. نياز اطّلاعاتي كاربر در يك زمان خاص مي‌تواند شامل يك يا چند مدرك باشد. مفهوم «ربط»، عامل مورد توجّهي در مسئلة بازيابي است.

   يك مدرك با توجه به ويژگيهايي كه دارد (نحوة نگارش، موضوع و...) و يا با در نظر گرفتن مشخصة كاربر (سابقة دانش فني وي) ممكن است براي يك كاربر خاص، مربوط و يا نامربوط تلقّي شود. در تمام سيستمهاي بازيابي اطّلاعات، چنانچه مدرك بازيابي شده در قضاوت كاربر، مورد توجه وي واقع گردد، آن مدرك به عنوان مدرك مربوط و در غير اين‌صورت، مدرك نامربوط شناخته مي شود. عوامل بسياري در قضاوت درباره عنصر «ربط» مؤثرند. از آنجا كه عوامل بسياري، قضاوت درباره ربط را با استفاده از روشهاي پيچيده تعيين مي كنند، يك سيستم بازيابي اطّلاعات نمي‌تواند به طور دقيق تمام مدارك مربوط را انتخاب نمايد. بنابراين، سيستم بايد روشهايي را بپذيرد كه رتبه‌بندي مدارك را به ترتيب احتمال استفادة كاربر از آنها آسان كند.

   يكي از روشهاي مناسب، محاسبة همبستگي اصطلاحات، بر اساس فراواني اصطلاحات هم آيند است. در صورت فرض تعامد بردارهاي مدارك و اصطلاح، استفاده از ماتريس هم آيند مي‌تواند يك عامل تنظيم‌كننده باشد. پژوهشگران اين حوزه، در فرايند بازيابي، روشهاي متفاوتي براي تشخيص همبستگي اصطلاحات ارائه كرده‌اند، از جمله مي‌توان به تحليل آماري جستجوها در مدارك مربوط و نامربوط به ترتيب همبستگيهاي مثبت و منفي اصطلاحات، اشاره كرد [5]. در پژوهش ديگري كه از ماتريس اصطلاح هم آيند استفاده شد، مجموعة اصلي بردارهاي اصطلاح، از روشهاي تحليل عامل يا مقياس چند بعدي بدست آمد [7 و6].

«كال» در پژوهش خود، طرحي ارائه كرد كه با استفاده از آن مي‌توان همبستگيهاي ميان اصطلاحات را بدون ماتريس اصطلاح هم آيند، ادغام نمود [8].

در يك سيستم بازيابي اطّلاعات، معمول است كه يك مدرك به وسيلة كليد واژه‌ها يا واژه‌هاي موضوعي نمايانده شود. كليد واژه‌ها معمولاً در فرايند نمايه‌سازي، از متن يا چكيدة مدرك استخراج مي شوند.

   علاوه بر گزينش اصطلاحات براي بازنمون مدارك، معمولاً به هر اصطلاح وزني مي‌دهند تا اهميّت آن اصطلاح خاص را در مدرك نشان دهد. بنابراين، در طراحي استراتژيهاي جستجو مي‌توان ماتريس مدرك ـ اصطلاح را به وجود آورد، به نحوي كه عنصر (i , r) اين ماتريس، متناظر با وزن اصطلاح i در مدرك r باشد [9]. در اين ماتريس، عنصر را به عنوان مؤلفة i بردار متناظر با مدرك r در نظر مي‌گيرند. هنگام جستجو، سيستم، بردار جستجو را به دست مي‌آورد و با بردارهاي مدارك بر اساس روش بيان تشابه ميان بردارها منطبق مي سازد [4]. با در اختيار داشتن اين ماتريس و با توجه به هدف رتبه‌بندي مدارك، روشهاي گوناگوني براي مدلسازي بازيابي به كار مي‌رود. يك روش كه در سالهاي اخير به طور گسترده استفاده شده، مدلسازي مدارك و جستجوها بر اساس بردار است [10 و 9] و به آن مدل «فضاي برداري»[[3]](http://192.168.0.110/editor/main.htm" \l "_ftn3" \o ") گفته مي شود. هر جستجو به صورت بردار نمايانده مي شود و تشابه آن را با بردارهاي مدارك در نظر مي‌گيرند. هر چه بردار جستجو به بردار مدرك نزديكتر باشد، به موضوع جستجو مربوط تر خواهد بود. مدل فضاي برداري را «سالتون» پيشنهاد كرده است [11 و 10 و 9].

**مدل فضاي برداري**

   مدل فضاي برداري يكي از مدلهاي بازيابي اطّلاعات است كه در سطح وسيعي به كار مي‌رود [13 و 12]. در اين مدل، هر مقولة اطّلاعاتي ـ شامل متون ذخيره شده و هر تقاضاي اطّلاعاتي زبان طبيعي ـ به صورت مجموعة بردارهايي از اصطلاحات نگهداري مي‌شوند. به طور نظري، اين اصطلاحات مي‌توانند از واژگان كنترل شده‌ انتخاب شوند. به خاطر وجود مشكلاتي در تهية اين واژگان، اصطلاحات از متون استخراج مي‌شوند. معمولاً براي كاهش اندازة واژگان از ريشة واژه‌ها استفاده مي شود. همچنين معمولاً از واژه‌هاي بازدارنده نظير…., an ,of ,the صرف نظر مي گردد. از تمام واژه‌هاي موجود در مدارك، يك مجموعه واژگان به وجود مي آيد. هر مدرك به صورت برداري از تمام واژگان نمايانده مي شود. بعيد است واژه‌هايي كه فاقد بار معنايي هستند و به طور معمول در مدارك يافت مي‌شوند، اطّلاعات مهّمي ارائه دهند، بنابراين مي‌توان اين واژه‌ها را براي سرعت دادن به پردازش، حذف كرد. واژه‌هاي تكراري كه مي‌توان از آنها چشم پوشيد فهرست واژه‌هاي غيرمجاز را مي‌سازند. در حذف واژه‌هاي غير مجاز، بايد دقّت زياد به كار برده شود. براي مثال:

چنانچه واژه‌هاي غير مجاز در جملة: «to be or not to be» حذف شوند، اين جمله غير قابل بازيابي خواهد بود.

مدل فضاي برداري، شيوه اي استبراي نمايش مدارك از طريق واژه‌هاي موجود در آنها. اين مدل، يك تكنيك استاندارد در بازيابي اطّلاعات است. بر اساس مدل فضاي برداري، مي‌توان تصميم گرفت كه كدام مدارك شبيه به يكديگر و يا به كليدواژه‌هاي جستجو شبيه هستند [12].

   بردار مربوط به هر مدرك (يا هر جستجو) داراي n مؤلّفه است. n برابر با تعداد اصطلاحات موجود در مجموعه مدارك است.

به هر يك از اصطلاحات هر مدرك، به طور خودكار وزني اختصاص مي‌يابد كه بر فراواني رخداد اصطلاح در كلّ مجموعة مدرك و تعداد دفعات حضور يك اصطلاح در مدرك خاص مبتني است. با افزايش فراواني اصطلاح در يك مدرك، وزن آن اصطلاح درمدرك افزايش مي يابد. برعكس، وقتي فراواني اصطلاح در مجموعة مدارك بيشتر باشد، اين وزن كاهش مي يابد.

به طور كلي، مي‌توان مزيّتهاي اصلي مدل فضايي برداري را چنين بيان نمود [13]:

1. طرح وزن دهي به اصطلاح در اين مدل، عملكرد بازيابي را بهبود مي بخشد.

2. استراتژي تطبيق جزئي اين مدل، بازيابي مداركي را مجاز مي شمارد كه به شرايط جستجو نزديك هستند.

3. فرمول رتبه بندي كسينوسي آن، مدارك را بر طبق درجة تشابهي كه به موضوع جستجو دارند، مرتّب مي‌كند.

**1-2. وزن دهي به اصطلاح**

چون اصطلاحات متفاوت، داراي اهميت مختلفي در متن هستند، از يك نشانگر مهم، يعني «وزن اصطلاح» استفاده مي‌شود كه همراه هر اصطلاح است [15و 14]. به اصطلاحات مهمتر، وزن بيشتر اختصاص مي يابد. براي وزن دادن به يك اصطلاح، از تعداد رويداد يك اصطلاح (فراواني آن يا tf استفاده مي‌شود). اهميت اصطلاح، مستقل از بستر جستجو نيست. براي مثال، اصطلاح "درياچه" در مجموعة مقالات درياچة خزر اهميت زيادي ندارد، اما در زمان جستجو در مقالات پيرامون كوير و صحرا، اين اصطلاح مي‌تواند بسيار مهم باشد. اين امر نشان دهندة اين موضوع است كه مداركي كه يك اصطلاح در آن مكرراً وجود دارد، ممكن است اهميت كمتري داشته باشد. از اين رو از «فراواني مدرك معكوس» يا «عامل idf» همراه با وزن اصطلاح استفاده مي‌شود. فراواني اصطلاح معكوسِ براي محاسبة اهميت واژه‌هاي نادر نسبت به واژه‌هاي معمولي به وجود آمده است. فراواني اصطلاح معكوس واژة i به وسيلة فرمول زير محاسبه مي شود.

در اينجا N تعداد كلّ مدارك و  تعداد مدارك حاوي واژة i است.

معمولاً در يك مدرك طولاني به طور مكرر از اصطلاح خاصي استفاده مي شود. عامل فراواني اصطلاح ممكن است براي مدرك طولاني، بزرگ باشد. همچنين مدرك طولاني، شامل بسياري از اصطلاحات گوناگون و متمايز است. اين امر باعث افزايش تعداد تطبيق واژه‌هاي يك جستجو و مدرك طولاني مي شود و به طور ناعادلانه‌اي احتمال بازيابي اين مدرك را نسبت به مدارك كوتاه تر افزايش مي‌دهد. براي جبران اين اثر، معمولاً وزنهاي اصطلاح را نرمال‌سازي مي‌كنند. نرمال‌سازي فراواني اصطلاح، يكي از اصلي ترين مباحث در بازيابي اطّلاعات طي سالهاي اخير است.

فراواني اصطلاح به طول مدرك بستگي دارد. بنابراين، در يك سيستم بازيابي اطّلاعات، نيازمند يكنواخت كردن طول مدرك با تكنيك نرمال‌سازي فراواني اصطلاح هستيم. بيشتر روشهاي نرمال‌سازي فراواني اصطلاح، در بازيابي اطّلاعات از پارامترها استفاده مي‌كنند. تنظيم اين پارامترها، مي‌تواند به تغييرات مهمي در ميزان دقت و بازيافت منجر گردد. يكي از پارامترهاي مهم، وزن است.

 نرمال‌سازي كسينوسي يكي از روشهاي موثر نرمال‌سازي است. هر بردار مدرك به طول اقليدسي آن تقسيم مي شود،  در اينجا،  وزن  اصطلاح i در مدرك است. وزن نهايي براي يك اصطلاح به صورت زير محاسبه مي گردد.

|  |
| --- |
| طول اقليدسي بردار مدرك |

وزن اصطلاحي كه در يك مدرك موجود نباشد را صفر در نظر مي‌گيرند [1].

باتوجه به نكات فوق مي‌توان چنين گفت كه تمام واژه‌هاي موجود در مدرك اهميت يكساني ندارند. يك واژه اگر داراي شرايط زير باشد، به احتمال زياد به مدرك بسيار مربوط خواهد بود:

الف) تكرار آن در ساير مدارك كم باشد.

ب) تكرار آن در مدركبالا باشد.

**2-2. رتبه‌بندي مدارك**

تابع حاصل‌‌ضرب داخلي برداري مي‌تواند براي يافتن همپوشاني واژگان ميان هر دو بردار متن استفاده گردد. جستجوي زبان طبيعي كه كاربر انجام مي‌دهد، به بردار وزن داري تبديل مي‌شود و با استفاده از تابع حاصل‌ضرب داخلي، تشابه عددي ميان بردار جستجو و بردار هر مدرك در مجموعه محاسبه مي گردد. با در نظر گرفتن بردار جستجوي Q و نمايش برداري مدركi به صورت، تشابه ميان جستجو و مدرك به صورت زير محاسبه مي شود [1]:

|  |
| --- |
| اصطلاحات مشترك |

در اينجا tj اصطلاحي است كه در جستجو و مدرك، ظاهر شده و qj وزن اصطلاح tj در جستجو و  tijوزن آن در مدركi   است. تمام اصطلاحات  tjكه هم در جستجو و هم در مدرك وجود دارند با هم جمع مي‌شوند. تشابه حاصل‌ضرب داخلي فهرستي از مدارك رتبه‌بندي شده با توجه به ميزان استفاده آنها ارائه مي دهد.

به طور معمول، كاربر جستجويي را در پايگاههاي اطّلاعاتي وارد مي كند. جستجو با تمام مدارك با اندازه‌گيري تشابه مقايسه مي شود. مدارك به ترتيب نزولي ِتشابهي كه با اصطلاح جستجو دارند، به كاربر ارائه مي شود.

**3-2. محاسبة تشابه**

روشهاي مختلفي براي اندازه گيري تشابه ميان دو مدرك، يا تشابه يك مدرك با يك جستجو وجود دارد. اندازه‌گيري كسينوسي، يك روش بسيار معمول اندازه‌گيري تشابه است، كه در اين روش كسينوس زاوية بين مدرك و جستجو اندازه گيري مي شود. با اندازه‌گيري تشابه، مجموعه اي از مدارك را مي‌توان با جستجو مقايسه كرد و آنگاه مربوط ترين مدرك را بازيابي نمود [12].

**اندازه گيري كسينوسي:**

براي دو بردار d و q، تشابه كسينوسي بين d و q به صورت زير محاسبه مي شود:

(1)

در اينجا،   حاصل‌ضرب برداريd  و q است كه با ضرب كردن فرايندهاي متناظر در هم، محاسبه مي شود.

اندازه گيري كسينوسي، زاوية بين بردارها را در فضاي چند بعدي محاسبه مي كند.

براي نمونه، شكل يك را در نظر بگيريد.

با توجه به اين شكل، تشابه ميان مدرك  و جستجوي Q، برابر با كسينوس زاوية بين دو بردار، يعني  است

و همين‌طور داريم:

**4-2. پياده سازي مدل فضاي برداري**

  براي مدل‌برداري، وزن wi,q مربوط به زوج (djوki) مثبت و غير دودويي است. علاوه بر آن، واژه‌هاي موضوعي در جستجو نيز وزن دار مي باشند. فرض كنيد wi,q وزن مربوط به زوج [ki,q]است كه در آن wi,q ? 0 است. سپس، بردار جستجوي به صورت   تعريف مي شود كه در آن t تعداد كلّ واژه‌هاي موضوعي در سيستم است. همانطور كه قبلاً گفته شد، براي مدرك  dj، بردار آن به صورت  نمايش داده مي شود.

**بنابراين، مدرك djو جستجوي كاربر، مثلq ، به صورت بردارهاي t بعدي نمايش داده مي‌شود. در مدل برداري پيشنهاد مي‌شود براي ارزيابي (سنجش) درجة تشابه مدرك dj با ملاحظة جستجويq ، از همبستگي ميان بردارهاي و  استفاده شود. اين همبستگي را مي‌توان تعيين كميّت كرد. براي مثال، مي‌توان از كسينوس زاوية ميان اين دو بردار به صورت زير استفاده كرد [2].**

Sim(dj,q)     =

 در اين جا و نرم بردارهاي مدرك و جستجو هستند. عامل (فاكتور)  تأثيري بر رتبه بندي (يعني ترتيب مدارك) ندارد، زيرا اين عامل براي تمام مدارك، يكسان است. عامل امكان نرمال‌سازي مدارك را فراهم مي‌كند. در اينجا وزن از فرمول زير به دست آمده است:

   از آن جا كه wi,j ? 0 و wi,q ? 0 است، مقدار Sim(q,dj) از 0 تا 1 متغيّر است. بنابراين، به جاي پيش‌بيني مربوط يا نامربوط بودن يك مدرك، مدل‌ برداري، مدارك را بر اساس درجة تشابه آنها نسبت به جستجو رتبه‌بندي مي‌نمايد. ممكن است يك مدرك، حتّي در حالت تطبيق جزئي با جستجو بازيابي گردد. براي مثال، مي‌توان آستانه اي را براي Sim(dj,q) تعيين كرد و مداركي را كه درجة تشابه آنها بيش از آستانه است،           بازيابي نمود [13].

**نحوة محاسبة رتبه بندي مدارك**

براي تبيين مسئله، از مثال ساده‌اي براي نشان دادن نحوة ساخت بردار استفاده مي‌كنيم:

مثال: مدارك و جستجوي زير را در نظر بگيريد:

: speech recognition and image processing and signal processing

: speech            models and image processing

q: image models

براي جستجو و هر يك از مدارك، جدولهايي شامل اصطلاحات موجود در هر ركورد ساخته مي شود.

براي هر يك از اصطلاحات مستقل، يك مؤلفة در جدول در نظر گرفته شده و فراواني اصطلاح نيز مشخص گرديده است.

: speech recognition and image processing and signal processing

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| speech | recognition | and | Image | processing | Signal |
| 1 | 1 | 2 | 1 | 2 | 1 |

: speech models and image processing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| speech | Models | and | image | Processing |
| 1 | 1 | 1 | 1 | 1 |

q: image models

|  |  |
| --- | --- |
| image | Models |
| 1 | 1 |

در اين مثال، تمام واژه‌ها براي تهية مجموعه واژگان به كار رفته‌اند.

واژگان ايجاد شده شامل تمام واژه‌هايي است كه در مدارك به كار رفته‌‌اند:

speech, recognition, and, image, processing, signal, models.

درمدل فضاي برداري، واژه‌هاي موجود در واژگان، مرتب سازي مي شود:

and, image, models, processing, recognition, signal, speech.

بنابراين در اين مثال بردارهاي مدارك و جستجو داراي 7 مؤلفه (به تعداد واژگان) بوده و بردار حاصل، هفت بُعدي خواهد بود.

بردارمدرك باتوجه به فراواني اصطلاحات موجود در آن به صورت زير نشان داده مي شود:

                : speech recognition and image processing and signal processing

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| and | image | Models | processing | recognition | signal | Speech |
| 2 | 1 | 0 | 2 | 1 | 1 | 1 |

و بردار مدرك به صورت زير به دست مي آيد:

: speech models and image processing

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| speech | signal | recognition | processing | Models | image | and |
| 1 | 0 | 0 | 1 | 1 | 1 | 1 |

بردار جستجو را مي‌توان مانند بردارهاي مدارك به وجود آورد:

q: image models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| and | image | Models | processing | recognition | signal | speech |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 |

رتبه بندي مدارك موجود نسبت به جستجويq  به صورت زير محاسبه مي شود:

ابتدا ضريب تشابه هر يك از بردارهاي مدارك و با بردار جستجوي q به شيوه زير محاسبه مي‌شود:

ضريب تشابه مدرك با جستجويq  به ترتيب زير به دست مي آيد.

با توجه به مقادير به دست آمده براي ضريب تشابه، از آنجا كه ضريب تشابه مدرك  با جستجويq    0.632 و ضريب تشابه مدرك    با جستجوي q برابر با 0.204  است، نتيجه مي‌شود كه مدرك **، مدرك مربوط تري نسبت به مدرك  براي جستجوي**، مدرك مربوط تري نسبت به مدرك  براي جستجويq  است. بنابراين، رتبه‌بندي به صورت و   است.

نتيجة اين محاسبات نيز با نگاه اجمالي به مدارك و جستجو قابل تأييد است.

**نتيجه گيري**

در اين مقاله، از ميان استراتژيهاي مختلف بازيابي اطّلاعات، مدل فضاي برداري به عنوان يكي از معتبرترين تكنيكهاي بازيابي به طور ساده بيان و نشان داده شد كه مدل‌برداري يك استراتژي رتبه‌بندي است كه با مجموعه‌هاي عمومي بهبودپذير است. اين استراتژي مجموعه جوابهاي رتبه‌بندي شده‌اي توليد مي‌كند كه بهبود آنها بدون بسط جستجو يا بازخورد ميزان ربط در چارچوب مدل برداري، مشكل است. در رتبه‌بندي، روشهاي بسيار مختلفي با مدل برداري مقايسه شده است، امّا به طور كلّي به نظر مي‌رسد مدل برداري، يا برتر بوده و يا تقريباً به خوبي ساير روشهاي موجود عمل مي نمايد. به علاوه، مدل برداري، آسان و سريع است. با توجّه به اين دلايل، مدل برداري يك مدل بازيابي معتبر است.

**منابع**

[1] Salton, G. (1989) Automatic Text Processing – The Transformation, Analysis and Retrieval of      Information by Computer, Addison – Wesley Publishing Co., Reading, MA, 1989.

[2] Salton, G. (1991) Developments in Automatic Text Retrieval, Science, 253, 974-980, August.

[3] Tai, X., Ren, F. Kita, K. (2001) An Information Retrieval Model based on Vector Space Method by Supervised Learning, Information Processing & Management.

[4] Raghavan,V.V., Wony, S.K.M. (1986) Critical Analysis of Vector Space Model for   Information Retrieval; Journal of the American Society for information Science.

[5] Raghavan, V.V., Yu, C.T. (1979) Experiments on the Determination of the Relationships Between Terms. ACM Transactions on Database Systems no. 4. pp.240 – 260.

 [6] Katter, R.v. (1967) A Study of Document Representations: Multidimension Scaling of Index Terms. SDC – Final Report.

[7] Switzer, P. (1964) Vector Images in Information Retrieval. Proceedings of the Symposium on Statistical Association Methods for Mechanical Documentation, Wash. D.C. (NBS Misc. Publ. 269, 1965) Stevens, M.E., Heilprin, L., Guiliano, V.E (eds.). pp. 163 – 171.

[8] Koll, M. (1979) Weird, An Approach to Concept – based Information Retrieval. ACM – SIGIR Forum, vol XIII, no. 4, (spring 1979), pp. 32- 50.

[9] Salton, G., McGill, M.J. (1983) Introduction to Modern Information Retrieval. McGraw hill, New York.

[10] Salton, G. (1971) The SMART Retrieval System – Experiments in Automatic Document Processing. Prentice – Hall, Englewood Cliffs, New Jersey.

 [11] Salton, G. (1983) Dynamic Information and Library Processing. Prentice – Hall, Englewood Cliffs, New Jersey.

[12]. گراسمن، ديويد و افير فريدر. «بازيابي اطّلاعات، الگوريتمها و روشهاي اكتشافي» ترجمه جعفرمهراد و سارا كليني، انتشارات كتابخانه رايانه‌اي، كتابخانه منطقه اي علوم و تكنولوژي، 1384.

[13] Baeza - Yates, R. Ribeiro- Neto, B,  Modern information Retrieval, Addison  Wesley, 1999.

[14] Salton, G., Yang, C.G., Yu, C.T. (1975) A Theory of Term Importance in Automatic Text Analysis, Journal of the ASIS, 26:1, 33-44.

 [15] Salton, G. (1988) Buckley, C., Term weighting Approaches in Automatic Text Retrieval, Information Processing and Management, 24:5, 513-523.

[16]. داورپناه، محمدرضا (1384). «ضرورتهاي نوين بازنگري در ذخيره و بازيابي اطّلاعات». **كتابداري و اطّلاع رساني**، جلد 8 ، شمارة 3 ، پاييز 1384، ص 67-88.

[17]. چاودري، جي جي، (1379). «پژوهش دربارة اينترنت و بازيابي اطّلاعات». ترجمة مهدي خادميان، **كتابداري و اطّلاع رساني**، جلد 3 ، شماره 3 ، پاييز 1379،          ص 133-162.

1. استاد بخش علوم كتابداري و اطلاع رساني دانشگاه شيراز و رئيس كتابخانه منطقه اي علوم و تكنولوژي

2. كارشناس ارشد مهندسي كامپيوتر و رئيس اداره فناوريهاي اطّلاعاتي كتابخانه منطقه اي علوم و تكنولوژي

1. Vector space model.