**نام مقاله: مروري بر نمايه‌سازي معاني پنهان: نظريه و كاربردها**

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

**شماره نشريه: 25 \_ شماره اول ، جلد 7**

**پديدآور: محمدباقر دستغيب**

**چكيده**

**نمايه‌سازي معاني پنهان روشي است كه اطلاعات را در قالب معاني ذخيره مي‌كند، و از رابطة پنهان ميان اصطلاحات و متن استفاده مي‌نمايد. در اين روش معايب توجه به اصطلاح در يك سند به صورت منفرد، برطرف مي‌گردد. در نظام بازيابي كه به اين روش فراهم مي‌آيد، به‌جاي تطبيق لغوي اصطلاحات كليدي، رابطة معنايي ميان اصطلاحات نيز مورد توجه قرار مي‌گيرد. در اين مقاله، نظريه و كاربردهاي نمايه‌سازي معاني پنهان در نظام‌هاي بازيابي مورد بررسي قرار مي‌گيرند.**

**كليدواژه‌ها: بازيابي اطلاعات، نمايه‌سازي معاني پنهان، تجزية مقدار ويژه**

**مقدمه**

در ميان گونه‌هاي مختلف اطلاعات موجود در اينترنت، بيشتر اطلاعات، بخصوص اسناد و مدارك علمي، داراي قالب‌بندي متني مي‌باشند و بنابراين بازيابي اطلاعات متني از اهميت بسياري برخوردار است (Kowalski, 1997).

براي آنكه بازيابي اسناد در اينترنت امكان‌پذير باشد، بايد روشي مناسب براي پياده‌سازي، ذخيرة اسناد و نمايه‌سازي انتخاب گردد. در ميان روش‌هاي گوناگون پياده‌سازي سند و درخواست، غالباً روش «فضاي بُرداري»[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn1" \o ") مورد استفاده قرار مي‌گيرد. در اين روش، سند و درخواست به صورت بُردارهايي از فركانس يا وزن اصطلاحات نمايه، پياده‌سازي مي‌گردند. در ميان فرمول‌هاي كلاسيك وزن‌دهي اصطلاحات نمايه، ساده‌ترين فرمول، استفاده از فراواني، و بسامد معكوس مي‌باشد: IDF\*W=TF

در اين روش، وزن هر اصطلاح از ضرب دفعات تكرار اصطلاح در سند (TF)[[2]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn2" \o ") در دفعات تكرار اصطلاح در كل اسناد (IDF) به‌دست مي‌آيد. براي آن‌كه طول اسناد در وزن اصطلاحات، اثرگذار نباشد مي‌توان وزن نهايي را با استفاده از طول سند، «نرمال» كرد. درنهايت هر سند به يك بُردار تبديل خواهد شد؛ با مقايسة بين‌بُردار درخواست و بردار سند، اسناد مرتبط[[3]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn3" \o ") بازيابي مي‌گردند. يكي از روش‌هاي معمول براي محاسبة شباهت ميان‌ بردارها، محاسبة زاوية ميان بردار درخواست و بردار سند است. هرچه زاوية ميان اين دو بردار كمتر باشد، سند و درخواست، شبيه‌ترند (شكل شماره 1)(Salton, 1983). درنهايت پس از نمايه‌سازي و محاسبة وزن براي تمامي اصطلاحات كليدي سند، يك ماتريس به نام ماتريس اصطلاح ـ سند[[4]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn4" \o ") به دست خواهد آمد. هر سطر از اين ماتريس، بردار مشخصة يكي از اسناد مي‌باشد و هر مدخل از ماتريس، وزن‌هاي محاسبه‌شده در اسناد است (شكل شماره 2)(Salton, 1983).

عموماً اطلاعات با تطبيق لفظي كلمات درخواست با اسناد، بازيابي مي‌شود. اگرچه تطبيق لغوي ممكن است روشي نادقيق در تطبيق درخواست و سند باشد، ولي از اين شيوه استفاده مي‌شود. معمولاً روش‌هاي زيادي براي بيان مفاهيم با واژه‌هاي مترادف وجود دارد، از اين رو در روش لغوي، كلمات موجود در درخواست ممكن است با كلمات اسناد مشابه، تطبيق نيابند. از سوي ديگر، اغلب كلمات داراي معاني گوناگون هستند؛ بنابراين كلمات موجود در درخواست كاربر ممكن است به صورت لغوي، با كلمات موجود در اسناد غيرمرتبط تطبيق داده شوند. روش بهتر آن است كه در زمان بازيابي اطلاعات، مفهوم و معناي پاية اسناد مورد استفاده قرار گيرد(Bery, Dumais & Shippy, 1995; Rosario, 2000).

نمايه‌سازي معاني پنهان[[5]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn5" \o ") بر آن است تا مشكل مقايسة لغوي را با بهره‌گيري از شاخص‌هاي ادراكي آماري، و نه تكيه بر اصطلاحات به صورت منفرد، برطرف كند. نمايه‌سازي معاني پنهان بيان مي‌كند كه در به‌كارگيري اصطلاحات در متون، يك ساختار پنهان وجود دارد كه با كاربرد و معناي اصطلاح در پاراگراف مرتبط است. براي آشكارسازي ساختار اصطلاح در سند، تجزية مقدار ويژه به كار برده مي‌شود. سپس با استفاده از پايگاه دادة مقادير ويژه و بُردارهايي كه از تجزية مقدار ويژه[[6]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn6" \o ") براي اسناد به دست آمده، بازيابي انجام مي‌شود. كارآيي نهايي سيستم بازيابي نشان مي‌دهد بردارهايي كه به صورت آماري از مقادير ويژه به دست آمده‌اند، ارتباط معنايي اصطلاحات و اسناد را به شكلي نيرومندتر آشكار كرده‌اند (Rosario, 2000).

در ادامه، ابتدا مفاهيم پاية مورد نياز براي درك نمايه‌سازي معاني پنهان، و سپس مزايا و معايب نمايه‌سازي معاني پنهان را در كاربردهاي مختلف بررسي مي‌كنيم.



**مفاهيم پايه**

نمايه‌سازي معاني پنهان، روشي است كه اسناد و درخواست را به فضايي با ابعاد معنايي پنهان منتقل مي‌كند. در فضاي معاني پنهان، سند و درخواست مي‌توانند شباهت كسينوسي (كسينوس زاوية ميان دو ‌بردار، كه هرچه به 1 نزديك‌تر باشد زاوية ميان دو بردار كمتر است) زيادي داشته باشند، حتي اگر از نظر لغوي، اصطلاحات مشترك نداشته باشند؛ درواقع شباهت ميان معناي اصطلاحات، سنجيده مي‌شود. مي‌توان به روش نمايه‌سازي معاني پنهان، به عنوان يك استاندارد شباهت، براي رفع مشكل يكسان‌بودن لغوي اصطلاحات در فرمول IDF\*TF توجه كرد (Rosario, 2000).

فضاي معاني پنهان، كه درخواست و سند را درونش پياده‌سازي مي‌كنيم، داراي ابعاد كمتري نسبت به فضاي اولية اسناد است. ابعاد را مي‌توان اصطلاحات كليدي درنظر گرفت. در نتيجه مي‌توان نمايه‌سازي معاني پنهان را روشي براي كاهش ابعاد (پيچيدگي) دانست. روش كاهش ابعاد، داراي يك مجموعه از اشياء است كه در يك فضا با بُعد بالاتر، موجودند و ما اعضاي مجموعه را در فضايي با ابعاد پايين‌تر پياده‌سازي مي‌كنيم. براي درك بهتر مي‌توان مثال فضاي سه‌بُعدي و دوبُعدي را درنظر گرفت(Rosario, 2000).

نمايه‌سازي معاني پنهان، كاربردي از يك تابع محاسباتي رياضي مي‌باشد، كه «تجزية مقدار ويژه براي ماتريس كلمه ـ سند» ناميده مي‌شود. بنابراين، پاية روش نمايه‌سازي معاني پنهان، محاسبة كوچكترين مربعات از نظر رياضي است. پياده‌سازي در محيط معنايي پنهان، نسبت به فضاي اوليه كوچكتر است، زيرا با درنظرگرفتن تفاضل كوچكترين مربعات، حداقل ابعاد براي پياده‌سازي به‌دست آمده است. روش تجزية مقدار ويژه از ماتريس A، ماتريس Ā را محاسبه مي‌كند و تفاضل دو ماتريس بوسيلة نُرم[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn1) دوم، كمينه مي‌گردد: (Ming, 1994)∆=║A-Ā║

نُرم دوم براي ماتريس، برابر با فاصلة اقليدسي ميان بردارها است. تجزية مقدار ويژه، فضايي با N بُعد را به فضايي با K بُعد تبديل مي‌كند، و K از N بسيار كوچكتر است. در كاربرد بازيابي اطلاعات و ماتريس اصطلاح ـ سند، N تعداد اصطلاحات موجود در مجموعه است. مقدار K معمولاً بين 100 تا 150 انتخاب مي‌گردد. نمايه‌سازي معاني پنهان، بردارها را از يك فضاي N بُعدي به يك فضاي K بُعدي منتقل مي‌كند.

روش‌هاي گوناگوني براي انقياد از فضايي با ابعاد بيشتر، به فضايي با ابعاد كمتر وجود دارد. نمايه‌سازي معاني‌پنهان روشي را برمي‌گزيند كه مقداركمينه‌ براي ∆ حاصل‌ گردد. پس از محاسبة«تجزية مقدارويژه ماتريس اصطلاح‌ـ‌سند»، d\*At به سه ماتريس n\*Tt وn\*Sn وn\*Dd تجزيه‌مي‌گردد:



t تعداد اصطلاحات، d تعداد اسناد و n مقدار كمينة t و d است و ماتريس‌هاي T و D داراي ستون‌هاي نرمال[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn1) هستند. به عبارت ديگر، حاصل‌ضرب آن‌ها در ترانهاده[[2]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn2)اش ماتريس واحد I، و رتبة[[3]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn3) ماتريس A برابر r مي‌باشد.

انتخاب مقدار K براي Ā مسئلة جالبي است. كم‌كردن مقدار K در برطرف‌كردن نويز هر بردار موثر است؛ ولي از طرف ديگر، ذخيرة ابعاد پايين، اطلاعات مهمي را حذف مي‌كند. به‌دست آوردن مقدار بهينه براي K در كاربرد، روش آزمون و خطا است. مي‌توان ثابت كرد كه تجزية مقدار ويژه براي يك ماتريس، منحصربه‌فرد است. بردارها نيز مانند اسناد، در فضاي K بُعدي جديد پياده‌سازي مي‌شوند. براي انقياد بردار ماتريس از فضاي N بُعدي به فضاي K بُعدي از اين فرمول استفاده مي‌شود: k\*kS-1k\*q=QTTt

**بهنگام‌سازي اطلاعات**

براي به‌روز نگه‌داشتن اطلاعات در فضاي جديد، بايد روشي براي تزريق درخواست‌ها و اسناد تازه به فضاي جديد انتخاب گردد. به‌دليل پرهزينه‌بودن، مقرون‌به‌صرفه‌ نيست كه مجدداً براي ماتريس تغييريافته، تجزية مقدار ويژه محاسبه شود. بنابراين اصطلاحات و اسناد جديد به درون ماتريس موجود تزريق مي‌گردند. براي تزريق اسناد در ماتريس مي‌توان از فرمول زير استفاده كرد:

A=TSDT

TT A=TTTSDT

TT A=SDT

بنابراين فقط كافي است كه بردار درخواست يا سند، در ترانهادة ماتريس اصطلاح T ضرب شود. پس از آن، ابعادش كاهش مي‌يابد و به فضاي دلخواه منتقل مي‌گردد (Rosario, 2000; Hong, 2000).

**مزايا و معايب**

**مزايا**

در نمايه‌سازي معاني پنهان، فرض بر آن است كه فضاي جديد، محيطي مناسب‌تر از محيط اصلي براي پياده‌سازي اسناد و درخواست‌ها مي‌باشد. به‌كاربردن كلمة «پنهان» در اين روش استعاره از آن است كه ابعاد جديد، پياده‌سازي درستي مي‌باشد. اين پياده‌سازي از آن نظر درست مي‌باشد كه در اين ابعاد خاص، مجموعه‌اي از اصطلاحات در برخي اسناد و مجموعة متمايز ديگري در اسناد ديگر، پياده‌سازي را به عهده دارند. روش نمايه‌سازي معاني پنهان، ابعاد اوليه و ساختار معنايي آن‌را ترميم مي‌كند. بنابراين بزرگترين مزيت روش نمايه‌سازي معاني پنهان، در اصطلاح‌هايي است كه مترادف، داراي چند معنا، و وابسته به اصطلاحات ديگر هستند (Rosario, 2000; Hong, 2000).

مترادف‌بودن، به اين معنا است كه يك مفهوم را مي‌توان با اصطلاحات مختلف بيان كرد. انجام روش‌هاي كلاسيك بازيابي، با كلماتي كه در اسناد مختلف، با معاني يكسان و تلفظ متفاوت به‌كار مي‌روند. دشوار است. در روش نمايه‌سازي معاني پنهان، ارتباط معنايي ميان اين گونه اصطلاحات، در وزن جديدي كه شاخص محاسبه مي‌نمايد، درنظر گرفته مي‌شود(Rosario, 2000; Hong, 2000).

تعدد معاني يعني اين كه اصطلاحاتي هستند كه بيش از يك معنا دارند و اين، يك ويژگي عام در تمام زبان‌ها است. وجود اصطلاحاتي كه تعدد معنا دارند، دقت نظام بازيابي را كاهش مي‌دهد. روش نمايه‌سازي معاني پنهان در فضاي جديد، نويز را كمينه مي‌كند. بردار اصطلاحات نمايه‌سازي معاني پنهان، ميانگين وزني[[4]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn4) معاني مختلف اصطلاح مي‌باشد. وقتي معناي واقعي با ميانگين معاني متفاوت باشد، نمايه‌سازي معاني پنهان در عمل، كيفيت جستجو را كم مي‌كند (Rosario, 2000; Hong, 2000).

مدل برداري كلاسيك فرض مي‌كند كه اصطلاحات مستقل، به صورت بردارهاي پاية عمود برهم هستند. به دليل پيوستگي ميان اصطلاحات در زبان، اين فرض هيچ‌گاه صحيح نيست. استقلال يا پيوستگي اصطلاحات، اگر به درستي مورد توجه قرار گيرد، كارآيي جستجو را افزايش مي‌دهد. اضافه‌كردن عبارت[[5]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn5) به اجزاي جستجو، از كاربردهاي اين قسمت است.

**معايب**

به نظر مي‌رسد كه تجزية مقدار ويژة ماتريس، داراي حجم كمتري نسبت به مدل اوليه باشد. مثلاً اگر عدد 150 را براي پارامتر K درنظر بگيريم، طول بردارها 150 مي‌باشد، ولي در عمل مشكل ديگري وجود دارد. در حالتي كه بردارها به روش معمول در فضاي اوليه پياده‌سازي مي‌گردند، ماتريس و بردارها، به صورت ماتريس پراكنده[[6]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn6) ذخيره مي‌شوند، ولي پس از انتقال به ابعاد جديد، ديگر حالت پراكنده ندارند. مشكل ذخيره‌سازي بر روي ديسك، امروزه قابل حل است، ولي برخي از عمليات رياضي، روي ماتريس‌هاي پراكنده سريعتر انجام مي‌گيرند (Rosario, 2000; Hong, 2000).

يكي از مشكلات مهم در به‌كارگيري نمايه‌سازي معاني پنهان، كارآيي سامانه از نظر سرعت پاسخگويي است. سامانه‌هايي كه از فضاي برداري در حالت كلاسيك استفاده مي‌كنند، نماية مقلوب را در زمان بازيابي براي مقايسة بردارها به‌كار مي‌برند. بنابراين فقط اسنادي كه اصطلاحات كليدي درخواست را شامل مي‌شوند، مقايسه مي‌گردند. ولي در روش نمايه‌سازي معاني پنهان، بردار درخواست بايد با تمام بردارهاي موجود در مجموعه مقايسه گردد. بنابراين ضرب بردارها براي محاسبة شباهت و زاوية ميان بردارها، زمان بيشتري طلب مي‌كند (Rosario, 2000; Hong, 2000).

تصوير 3 مقايسه‌اي ميان روش نمايه‌سازي پنهان و ديگر روش‌هاي كلاسيك را نشان مي‌دهد. مشاهده مي‌گردد كه اين روش، كارآيي بيشتري از مدل‌هاي ديگر دارد. اين آزمايش به روي مجموعة «گرانفيلد» كه داراي 1400 سند و 225 درخواست مي‌باشد انجام شده است.

**كاربردهاي نمايه‌سازي معاني پنهان**

**بازيابي اطلاعات:** كاربرد نمايه‌سازي معاني پنهان در بازيابي اطلاعات، و در به‌دست‌آوردن ليست اسناد مرتبط با درخواست است. كارآيي اين روش در مجموعه‌هاي مختلف آزمايش شده است و نتايج به‌دست آمده نشان مي‌دهد كه در روش نمايه‌سازي معاني پنهان، ميانگين دقت حدود 30 درصد بهتر است(Foltz, 1998; Dumais, 1997).

**ارتباط بازخورد:** در اين روش، براي اعمال بازخورد از اسناد انتخاب‌شده توسط كاربر، بردار درخواست با مجموع بردارهاي درخواست‌هاي انتخاب‌شده به عنوان مرتبط در ليست پاسخ، جايگزين مي‌گردد. اگر بردار درخواست، با اولين سند شبيه، جايگزين گردد، كارآيي سيستم حدود 33 درصد اضافه مي‌شود، و جايگزيني بردار درخواست با ميانگين اولين سه سند شبيه، حدود 67 درصد كارآيي كل سيستم را اضافه مي‌كند. پس اين روش به عنوان كاربرد نمايه‌سازي معاني پنهان در بهينه‌سازي درخواست، مورد توجه است (Foltz, 1998; Dumais, 1997).

**پالايش‌كردن اطلاعات:** كاربرد نمايه‌سازي معاني پنهان در جداسازي اطلاعات، بسيار ساده است. در اين كاربرد، بردار سند انتخابي توسط كاربر با بردار اسناد ديگر مقايسه مي‌گردد و معيار شباهت اسناد با يك مقدار آستانه،[[7]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn7" \o ") اسناد شبيه را از ميان اسناد پالايش خواهد كرد. در اين ميان به‌كارگيري روش‌هاي هوشمند و بازخورد، نتايج خوبي را حاصل خواهد كرد (Foltz, 1998; Rosario, 2000; Dumais, 1997).

**بازيابي مستقل از زبان:** نكتة مهم در به‌كارگيري نمايه‌سازي معاني پنهان آن‌ است كه اين روش از قواعد لغوي يا معنايي مربوط به زبان انگليسي استفاده نمي‌كند، بنابراين براي بازيابي اطلاعات در هر زبان مي‌توان از اين روش به منظور بازيابي اسناد مشابه استفاده كرد. امكانات مورد نياز فقط فضاي مشترك براي پياده‌سازي اصطلاحات و بردارها است، بنابراين مي‌توان اسناد و ماتريس اصطلاح‌ـ سند را براي مداركي كه در بيش از يك زبان داراي نسخة متني هستند مورد استفاده قرار داد. مثلاً مي‌توان از زبان انگليسي و فرانسوي استفاده كرد و يك فضاي مشترك براي هر دو زبان ايجادكرد، و در اين حالت نياز به ترجمة درخواست نيست و مشكلي در بازيابي اسناد مرتبط به‌ وجود نخواهد آمد(Foltz, 1998; Rosario, 2000; Dumais, 1997).

**ورودي‌هاي نويزي:** به‌دليل آن‌كه روش نمايه‌سازي معاني پنهان به تلفظ اصطلاحات كليدي وابسته نيست، در مواردي كه همراه با نويز باشد، بسيار مفيد است. مثلاً زماني كه املاي اصطلاح غلط باشد يا در شناخت نويسه‌هاي تصويري،[[8]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftn8" \o ") مي‌توان اين روش را به‌كار برد. اگر غلط املايي فقط در يك مكان وجود داشته باشد در فضاي جديد اسناد، ترميم مي‌گردد و با نسخة صحيح از اصطلاح، در ميانگين مربعات جايگزين مي‌گردد (Foltz, 1998; Rosario, 2000; Dumais, 1997



**نتيجه**

نمايه‌سازي معاني پنهان روشي تازه و اميد‌بخش در بازيابي اطلاعات است، كه اسناد را در فضايي با ابعاد كمتر، شاخص‌گذاري و بازيابي مي‌كند. هدف از تغيير فضاي اسناد، استفاده از نمايه‌سازي و بازيابي معنايي اصطلاحات است. در اين روش براي كمينه‌كردن نُرم ماتريس، از تجزية مقدار ويژه استفاده مي‌شود و در مقايسه‌هاي انجام‌شده در TREC[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftn1) جايگاه مناسبي به‌دست آورده است. اين روش به‌خصوص زماني كه نمايه‌سازي و جستجوي مستقل از زبان، در اولويت باشد، كارآيي مناسبي خواهد داشت. مثلاً موتور جستجوگر «گوگل» از نمايه‌سازي معاني پنهان، در نمايه‌سازي‌ها و جستجوهاي مستقل از زبان استفاده مي‌كند.

**منابع**

Kowalski, G. “Information retrieval systems, Theory and Implementation”, Kluwer Publisher. 1997. [Online] Available: <http://gunther.smeal.psu.edu/context/48183/0>

Salton G., McGill M J. “An introduction to Information Retrieval”, McGraw Hill, 1983 [Online] Available: <http://citeseer.nj.nec.com/context/1241/0>

Berry M. W., Dumais S T., Shippy A. T., “A case study of latent semantic indexing”, Tech. Rep. CS-95-271, University of Tennessee, Knoxville, January 1995.

ROSARIO B., “Latent Semantic Indexing: An overview”, INFOSYS 240, spring 2000.

Ming Gu, Demmely James, Dhillonz Inderji, “Efficient Computation of the Singular Value Decomposition with Applications to Least Squares Problems”, Lawrence Berkeley National Laboratory Technical Report LBL-36201,September 29, 1994 [Online] Available: [www.cs.berkeley.edu/~inderjit/fac\_resume.ps](http://www.cs.berkeley.edu/~inderjit/fac_resume.ps)

Hong Jason I., “An Overview of Latent Semantic Indexing”, SIMS 240, spring 2000.

PETER W. Foltz, “Using Latent Semantic Indexing for Information Filtering” IEEF-CS\TCDA, University of Colorado, 1998.

Dumais Susan, Bellcore, “Using Latent Semantic Indexing (LSI) For Information Retrieval, information Filtering and Other Things”, Cognitive Technology Conference Cognitive Technology Conference April 4, 1997.

1. نشست بازيابي اطلاعات متني

[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref1" \o ") . Ortho normal

[[2]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref2" \o ") . Transpose

[[3]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref3" \o ") . Rank

[[4]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref4" \o ") . Weighted Average

[[5]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref5" \o ") . Phrase

[[6]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref6" \o ") . Sparse Matrix

[[7]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref7" \o ") . Threshold

[[8]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default" \l "_ftnref8" \o ") . OCR (Optical Character Recognition)

[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref1) . Norm

[[1]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref1) . Vector Space Model

[[2]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref2) . Term Frequency

[[3]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref3) . Relevant

[[4]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref4). Term-Document Matrix

[[5]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref5) . Latent Semantic Indexing (LSI)

[[6]](http://www.aqlibrary.org/modules/FCKEditor/pnincludes/editor/fckeditor.html?InstanceName=desc&Toolbar=Default#_ftnref6) . Singular Value Decomposition (SVD)