

# طراحی معماری یک عامل هوشمند تطبیقی برای جستجوی اطلاعات تجاری با استفاده از تئوری

## ترکیب اطلاعات

کاوه کاووسی

دانشجوی دکتری، قطب کنترل و پردازش هوشمند، دانشکده برق و کامپیوتر دانشگاه تهران، [kkavousi@ut.ac.ir](mailto:kkavousi@ut.ac.ir)

بهزاد مشیری

استاد قطب کنترل و پردازش هوشمند، دانشکده برق و کامپیوتر دانشگاه تهران، [moshiri@ut.ac.ir](mailto:moshiri@ut.ac.ir)

**چکیده:** بی شک مسئله بازیابی مناسب اطلاعات تجاری از اینترنت و سایر پایگاه های داده با مقیاس بزرگ و بسیار بزرگ از مهمترین مشکلات موجود بر سر راه استفاده ی بهینه از این منابع اطلاعاتی می باشد. دانشمندان علوم ارتباطی توسعه تجارت الکترونیکی را از یکسو مرتبط به گسترش قابلیت های فناوری اطلاعات و از سوی دیگر وابسته به کاهش هزینه های مترتب به استفاده از این فناوری می دانند. در حال حاضر سرویس های زیادی برای جستجوی وب موجود است. ولی هیچکدام از این ابزار ها آنقدر که مورد انتظار است مفید نیستند و عملا در بسیاری از موارد نتایج رضایت بخش نیستند. یکی از مهمترین عوامل این امر عدم برخورد تخصصی این ابزارها با اطلاعات است. بنابراین گرایش به سمت ابزارهای جستجوی تخصصی راه حل طبیعی این معضل می باشد. یکی از مهمترین عوامل در ایجاد اینگونه ابزارها وجود قابلیت یادگیری و اصلاح مدل رفتاری آنها در قبال موضوعات مختلف و به صورت تخصصی می باشد. در این مقاله راجع به یک عامل هوشمند بحث شده است که با استفاده از تکنیک های ترکیب اطلاعات نقش یک فراجویشگر سفارشی در زمینه های تجاری را برای کاربر ایفاء می کند. در معماری این عامل از تکنیک های ترکیب اطلاعات و مکانیزم های اصلاح تطبیقی بهره گرفته شده است.

**کلید واژه ها:** ترکیب اطلاعات، عامل هوشمند، ترکیب مجموعه ها، فراجویشگر.

## مقدمه

استفاده از فناوری های الکترونیکی در زمینه امور بازرگانی پیشینه ای نسبتاً طولانی دارد. اولین تلاش ها در این زمینه از سال ۱۹۶۵ م. آغاز گردیده است. امکان دریافت پول از دستگاه های خود پرداز (ATM) و خرید از طریق کارت های اعتباری از ابتدایی ترین اقدامات انجام گرفته در این زمینه می باشند (Molla and Licker, ۲۰۰۱). این روند با ایجاد سیستم های میان سازمانی که امکان تبادل اطلاعات بین سازمان ها و هدایت الکترونیکی کسب و کار را فراهم نمودند دنبال گردید. (Senn, ۲۰۰۰) در سال های نخستین دهه ۹۰ نیاز به مبادله اطلاعات تجاری میان شرکت های بزرگ به حدی جدی شد که پیش از توسعه تکنولوژی های مبتنی بر اینترنت، خود این شرکت ها اقدام به ایجاد شبکه های کامپیوتری با سطوح دسترسی محدود نمودند. در آن زمان این روش مبادله اطلاعات "مبادله الکترونیکی داده ها" یا EDI نامیده شد. ایجاد و توسعه اینترنت بستر مناسبی برای انتقال اطلاعات با حجم زیاد میان کشورها، سازمانها و افراد با یکدیگر ایجاد نمود. این بستر همان چیزی بود که امکان بازرگانی الکترونیکی را فراهم آورد (Vadapalli, ۱۹۹۸). دانشمندان علوم ارتباطات توسعه تجارت الکترونیکی را از یکسو مرتبط به گسترش قابلیت های فناوری اطلاعات و از سوی دیگر وابسته به کاهش هزینه های مترتب به استفاده از این فناوری می دانند. از این دیدگاه فناوری اطلاعات در دوره موسوم به عصر اینترنت قرار دارد. این عصر سومین دوره از دوره های جهش فناوری اطلاعات قلمداد گردیده است (Seddon, ۱۹۹۷).

. تقسیم بندی مورد نظر برای این فرآیند سه دوره بیست ساله به شرح ذیل قائل می باشد:

• ۱۹۷۴-۱۹۵۵: عصر پردازش الکترونیکی داده ها (EDP)

• ۱۹۹۴-۱۹۷۵: عصر سیستم های اطلاعاتی مدیریت (MIS)

• ۲۰۱۴-۱۹۹۵: عصر اینترنت

هر یک از دوره های مورد اشاره متناسب با توانایی های تکنولوژیکی همان دوره امکانات تجارت الکترونیکی را فراهم آورده اند. در عصر نخست ماشین های خود پرداز و کارت های اعتباری بوجود آمدند و در عصر دوم امکان

استفاده از مبادله الکترونیکی داده‌ها (EDI) ، سیستم بانکی بین‌المللی SWIFT، و انتقال وجه الکترونیکی (EFT) فراهم آمد. اما بزرگترین تحولات در عصر سوم و در دوره توسعه اینترنت و کاربرد های تجاری آن شکل گرفت. این تحولات به قدری شگرف بوده اند که باعث تمایز بین تجارت الکترونیکی از نوع سنتی و اینترنتی گردیده اند (Molla and Licker, ۲۰۰۱). از ویژگی های بارز دوران سوم تجارت الکترونیکی حجم عظیم داده های در دسترس از طریق اینترنت می باشد. دسترسی به اطلاعات تجاری لازم از میان این داده ها که شامل اطلاعات گوناگون و در تمام حوزه های متصور می باشند کاری بسیار دشوار می باشد. دشواری این فرآیند زمانی بیشتر محسوس می گردد که جستجوی اطلاعات مورد نیاز از طریق ابزار های متداول جستجو در اینترنت مانند موتور های جستجو منجر به ارجاع لیست هایی بسیار طولانی می گردند که غالباً فقط حوصله و امکان بازدید تعداد محدودی از آنها برای کاربر وجود دارد و این موجب می گردد که در بسیاری از موارد بهترین و کاملترین گزینه ها اصلاً مورد بازدید کاربر قرار نمی گیرند. گرچه این یک معضل عمومی برای کاربران اینترنت محسوب می گردد اما برای کاربرانی که به دنبال اطلاعات تجاری در زمینه هایی چون خرید، فروش، بازاریابی، سرمایه گذاری و ... می گردند از ابعاد پیچیده تری برخوردار است. بی شک مسئله بازاریابی مناسب اطلاعات تجاری از اینترنت و سایر پایگاه های داده با مقیاس بزرگ و بسیار بزرگ از مهمترین مشکلات موجود بر سر راه استفاده بهینه از این منابع اطلاعاتی می باشد. امروزه وب بزرگترین پایگاه داده از مستندات مختلف می باشد و هرچه وب گسترده تر می شود نیاز به ابزار های قوی جستجو در اینترنت بیشتر احساس می شود. در حال حاضر سرویس های زیادی برای جستجوی وب موجود است. ولی هیچکدام از این ابزار ها آنقدر که مورد انتظار است مفید نیستند و عملاً در بسیاری از موارد نتایج رضایت بخش نیستند. یکی از مهمترین عوامل این امر عدم برخورد تخصصی این ابزارها با اطلاعات است. به عنوان نمونه چون این ابزار ها برای جستجوی عام طراحی شده اند بنابراین نمی توانند نیاز های یک کاربر به اطلاعات تجاری را به صورت دقیق برآورده سازند. بنابراین گرایش به سمت ابزارهای جستجوی تخصصی راه حل طبیعی این معضل می باشد. یکی از مهمترین عوامل در ایجاد اینگونه ابزارها وجود قابلیت یادگیری و اصلاح مدل رفتاری آنها در قبال موضوعات مختلف، و به صورت تخصصی

می باشد. در این مقاله راجع به یک فراجویشگر\_هوشمند بحث شده است که با استفاده از تکنیک های ترکیب اطلاعات\_نقش یک فراجویشگر سفارشی\_در زمینه های تجاری را برای کاربر ایفاء می کند.

### خصوصیات عامل\_فراجویشگر هوشمند ترکیب اطلاعات

این عامل کلمات و عبارات تخصصی تجاری مورد علاقه کاربر را که مایل است مطالب مرتبط با آن کلمات را مطالعه نماید دریافت می کند و سپس از کاربر می خواهد تا برای اهمیت حضور و یا عدم حضور آن کلمه یا عبارت در متن، وزنی در نظر بگیرد. این وزن به صورت مفهومی زبانی\_می باشد. به این صورت که کاربر می تواند تعیین کند که حضور و یا عدم حضور یک لغت یا عبارت در متن با اهمیت کم، با اهمیت زیاد، و یا با اهمیت بسیار زیاد می باشد. سپس عامل توسط واحدی به نام Query Generator به تعداد سرویس دهنده های اطلاعاتی (مثلا جویشگر های اینترنتی و یا پایگاه های داده) که قالب پرس و جو\_ی آنها را می شناسد، پرس و جو آماده می کند. (شکل های ۱ و ۲ و ۳) پس از ارسال پرس و جو ها هر سرویس دهنده لیستی از مستندات را باز می گرداند که این لیست با توجه به الگوریتم کاری همان سرویس دهنده بر اساس نزدیکی موضوع به پرس و جو رتبه بندی شده است. سپس فراجویشگر این لیست ها را بررسی می کند و موارد تکراری را حذف می کند و آنها را بر اساس الگوریتم های مناسب ترکیب مجموعه ها به صورتی با یکدیگر ترکیب می کند که یک لیست مرتب شده از مستندات ایجاد شود. از این بابت به هر سند بر اساس رتبه اش در این لیست یک امتیاز تعلق می گیرد (شکل ۲). (Yager and Rybalov, ۱۹۹۷). سپس تک تک مستندات موجود در این لیست مورد پردازش قرار می گیرند و وضعیت آنها از لحاظ حضور و یا عدم حضور لغات و عباراتی که کاربر مشخص کرده است تعیین می گردد و بر اساس تعداد و پراکندگی حضور آنها دو امتیاز به هر سند تعلق می گیرد. در این مکانیزم، عملگر قدرتمند OWA\_ (Yager and Kacprzyk, ۱۹۹۷) نقش اساسی بر عهده دارد. جنبه های فوق از عملکرد فراجویشگر مورد بحث به تفصیل در (Kavousi and Moshiri, ۲۰۰۷) مورد بررسی قرار گرفته اند (شکل ۱).

تمرکز اصلی این مقاله بر نحوه ترکیب لیست های حاصل از موتورهای جستجو و مدل سازی از موتورهای جستجوی مورد استفاده توسط فراجویشگر قرار دارد. هر بار که کاربر قصد استفاده از فراجویشگر را دارد مشخص می کند که موضوع مورد علاقه اش در کدام خوشه\_موضوعی قرار می گیرد. منظور از یک خوشه موضوعی یک دسته بندی منطقی از موضوعات مورد علاقه است. هر بار که کاربر جستجوی جدیدی را آغاز می کند می تواند یکی از خوشه های موجود را برگزیند و یا یک خوشه جدید ایجاد نماید. برخی از دسته بندی هایی که در حال حاضر در این فراجویشگر پیش بینی شده اند در جدول ۱ آمده اند. عامل با توجه به تاریخچه عملکرد هر سرویس دهنده ی اطلاعاتی در هر دسته بندی موضوعی، به هر سرویس دهنده یک امتیاز اختصاص داده است. زمانی که کاربر خوشه خاصی را انتخاب نمود، به هر سند بر اساس سرویس دهنده ای که آنرا بازیابی نموده است امتیاز همان سرویس دهنده را اختصاص می دهد. این امتیاز ها پس از هر بار استفاده از فراجویشگر بر اساس یک الگوریتم که در ادامه به صورت کامل توضیح داده خواهد شد به روز رسانی می گردند و بنابر این به تدریج مدل رفتاری هر موتور جستجو و نحوه عملکرد آن در قبال یک موضوع خاص در ذهن فراجویشگر هوشمند شکل می گیرد

موضوع خوشه	شماره خوشه
MEMS marketing	۱
Portland Cement Production	۲
Sony laptop market in European Union	۳
...	...

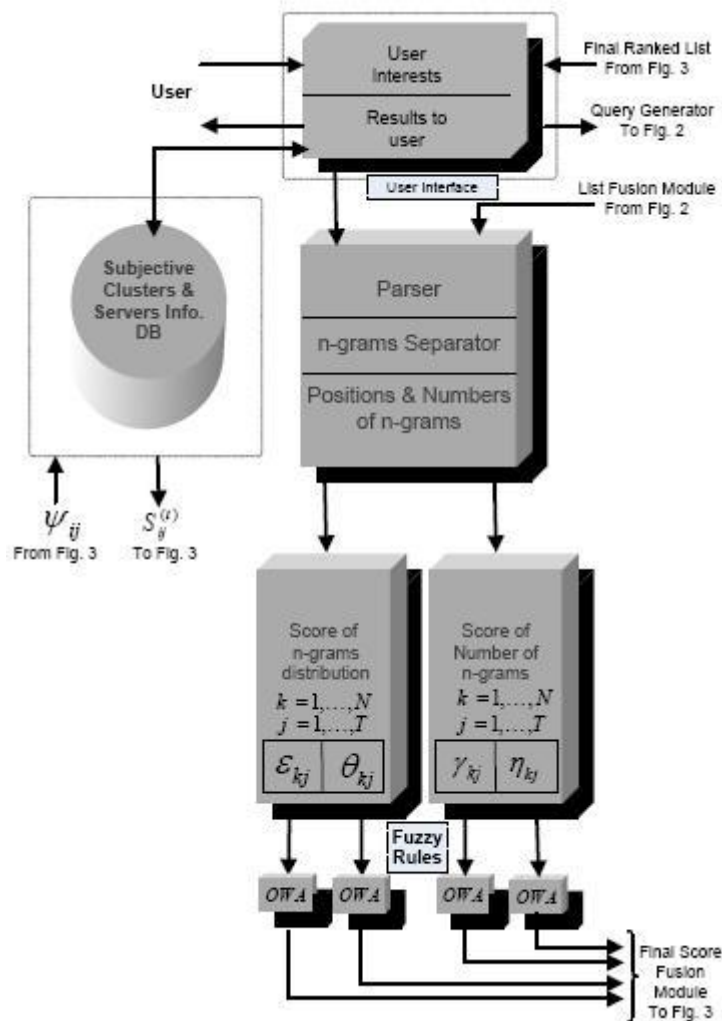
جدول ۱: لیست برخی از خوشه های اطلاعاتی که عامل هوشمند در حال حاضر سوابق آنها را برای شناسایی مدل رفتاری سرویس دهنده های اطلاعاتی نگهداری می کند

با توجه به توضیحات فوق برای هر سند ۴ امتیاز مختلف یدست می آید که در نهایت عامل باید امتیاز نهایی هر یک از مستندات را بر اساس آنها محاسبه نماید و در اختیار کاربر قرار دهد. این کار از طریق ترکیب این امتیازات با روش های ترکیب اطلاعات انجام می گیرد و در آن برای هر امتیاز وزنی در نظر گرفته می شود که این وزن

نشاندنده اهمیت آن معیار در تصمیم گیری نهایی می باشد. همچنین در نهایت امتیاز هر سرویس دهنده در آن دسته بندی خاص با/بدون بازخوردی که از کاربر دریافت می شود تصحیح می گردد (شکل ۳). به خاطر گستردگی بحث در این مقاله روش ها و الگوریتم های مورد استفاده در زمینه چگونگی امتیازات تخصیصی در ارتباط با حضور و یا عدم حضور لغات و عباراتی که کاربر مشخص کرده است و نیز بر اساس تعداد و پراکندگی حضور آنها مورد بحث نیستند و در (Kavousi and Moshiri, ۲۰۰۴) مورد بررسی کامل قرار گرفته اند.

### مسئله ترکیب مجموعه ها

در این قسمت به نحوه ی استفاده از عامل مورد بحث از مکانیزم ترکیب لیست ها می پردازیم. ابزار های مختلف جستجوی وب و سایر بانک های عظیم اطلاعاتی مرتبط با امور بازرگانی اغلب مکمل یکدیگر هستند. بنابر این پوشش دادن هرچه بیشتر منابع اطلاعاتی و بدست آوردن نتایج خالص تر یک استراتژی معقول آن است که از ابزار های جستجوی مختلف استفاده شود و در نهایت نتایج حاصل از آنها پالایش و سپس با هم ترکیب شوند. حال سوال اساسی آن است که بهترین روش برای ترکیب این لیست ها چیست؟ دلی اهمیت این سوال آن است که لیست هایی که ابزار های جستجو ارائه می کنند عمدتاً مرتب شده هستند. این ترتیب بر اساس مدل رفتاری هر ابزار، منابع اطلاعاتی در دسترس، و اولویت هایی که در نظر می گیرد تنظیم می شود و بنابر این این لیست ها برای ابزار های گوناگون جستجو متفاوت می باشند.



شکل ۱: دریافت علائق کاربر، استخراج ویژگی های مورد نظر از مستندات بازپایی شده توسط سرویس دهنده های اطلاعاتی، و ترکیب نتایج نهایی

فرض کنید گروهی از سرویس دهنده های اطلاعاتی (مانند موتورهای جستجوی اینترنتی) در اختیار داریم. این

سرویس دهنده ها را با  $\{S_i$   
 $i = 1, \dots, M$  نشان می دهیم که در آن  $M$  تعداد سرویس دهنده ها می باشد. همچنین فرض می کنیم هر سرویس دهنده مجموعه ی منحصر به فردی از اسناد را در اختیار دارد (یعنی موارد تکراری از لیست ها حذف شده اند و هر سند فقط در لیست یکی از سرویس دهنده ها قرار دارد). حال

برای یک کوئری مانند  $Q$ ، هر سرویس دهنده یک امتیاز به هر سند می دهد و در نهایت در پاسخ هر کوئری یک لیست مرتب شده از اسناد مرتبط با آن کوئری را باز می گرداند.

حال ما میخواهیم همه این لیست ها را با هم ادغام کنیم و یک لیست واحد بوجود آوریم که اقلام آن به ترتیب اهمیت از میان همه قلم های موجود در تمام لیست ها انتخاب شده اند (شکل ۲). اما نحوه ادغام آنها و تولید یک لیست نهایی بحث بسیار مهمی است که پارامترهای متعددی در آن نقش دارند. فرض کنید که  $M$  لیستی که باید با هم ادغام شوند را با  $L_1, L_2, \dots, L_M$  نشان می دهیم. همچنین تعداد اقلام موجود در لیست  $j$  ام را با  $N_j$  نشان می دهیم. منطقی ترین راه برای ادغام این  $M$  لیست و تولید یک لیست مرتب شده آن است که به هر قلم از لیست ها یک مقدار  $v$  نسبت داده شود و سپس همه قلم از روی این مقادیر مرتب شوند. بنابر این سوال اساسی این خواهد بود که چگونه می توان این مقدار  $v$  را برای هر قلم اطلاعاتی تعیین نمود؟ برای این منظور ما تابعی از دو متغیر نیاز داریم. یعنی رتبه نهایی هر قلم از لیست ها بوسیله دو پارامتر مشخص می شود.

- عدد  $j$  که نشان می دهد یک قلم از لیست  $L_j$  انتخاب شده است.
- رتبه  $i$  که مکان این صفحه در لیست  $L_j$  را نشان می دهد (اولین قلم هر لیست را با ۱ نشان می دهیم).

بنابر این مقدار  $v$  باید با استفاده از مقادیر  $i$  و  $j$  به صورت منحصر به فرد محاسبه شود. بر اساس اصل فقدان دلیل کافی که در تئوری آمار بیزین و در تئوری تصمیم سازی مطرح می شود، برای دو لیست که طول مساوی دارند مقدار امتیاز  $v$  تخصیصی به قلم  $i$  ام در هر دو لیست برابر خواهد بود. بنا براین برای لیست های با طول مساوی اگر فرض کنیم که هیچ اطلاعاتی در مورد کیفیت اقلام اطلاعاتی لیست ها در دست نداریم، به مقادیر



هم مکان در این لیست ها مقادیر  $v$  یکسان اختصاص داده خواهد شد. بنا براین وابستگی  $v$  به  $j$  به معنی وابستگی آن به  $N_j$  خواهد بود.

یعنی:

$$v = v(i, N_j) \quad (1)$$

حال باید سعی کنیم تابع دو متغیری مناسبی برای جایگزینی با  $v$  پیدا کنیم. در (Yager, ۹۷) نشان داده شده است که هر تابع پیوسته ی دو متغیری حقیقی مانند  $v(i, N)$  در صورتیکه دو شرط زیر را ارضا نماید می تواند جهت ادغام لیست های رتبه بندی شده مورد استفاده قرار گیرد:

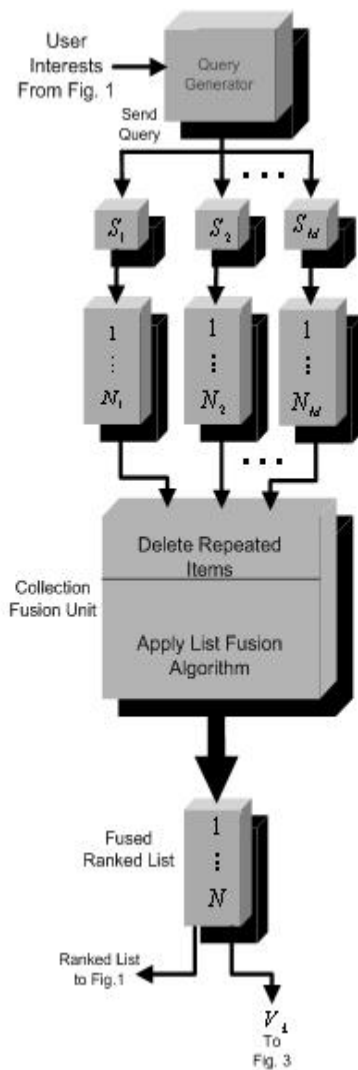
• برای هر  $i$  و  $k$  و  $i'$  و  $k'$  و  $N$ :

$$v(k, N) - v(i, N) = v(k', N) - v(i', N) \quad \text{اگر } k - i = k' - i'$$

• برای هر  $i$  و  $i'$  و  $N$  و  $N'$  و  $i$  داشته باشیم:

$$v(i, N) < v(i', N') \Leftrightarrow v(Li, LN) < v(Li', LN')$$

شرط اول بیان می دارد که عناصر با فاصله مساوی در لیست نهایی نیز همان فاصله را حفظ می کنند. شرط دوم نیز بیان می کند که با تغییر سطح غربال ترتیب رتبه برای دو عنصر رخ نمی دهد.



شکل ۲: ارسال کوئری به سرویس دهنده های اطلاعاتی و ایجاد یک لیست رتبه بندی شده با استفاده از نتایج بازگشت داده شده

بر اساس الگوریتم های ترکیب لیست

در همان مرجع قضیه زیر اثبات شده است:

قضیه: تابع  $v(i, N)$  یک ترکیب کننده ی لیست ها را توصیف می کند اگر برای مقادیر حقیقی  $\alpha$  و  $\beta$  داشته

باشیم:

$$v(i, N) = N^c (i + cN) \quad (2)$$

در (Yager, ۹۷) نشان داده شده است که رابطه فوق می تواند رفتار یک فرد خبره را شبیه سازی نماید. در نهایت با اعمال الگوریتم های مناسب برای تعیین بهینه ی مقادیر  $\alpha$  و  $c$  در رابطه ی فوق، می توان یک لیست مرتب شده بوجود آورد. در این لیست مرتب شده هر سند بر اساس موقیتش در این لیست مرتب شده یک امتیاز دریافت خواهد کرد که در محاسبه نهایی امتیاز مستندات یکی از معیار های چهارگانه خواهد بود. این امتیاز برای هر سند به صورت زیر محاسبه می گردد:

$$\left\{ \begin{array}{l} V_k = \frac{H_k}{\max_{i=1}^N(H_i)} \\ k = 1, \dots, N \end{array} \right. \quad (3)$$

که در آن  $H_k$  امتیاز تخصیصی به هر سند بر اساس رابطه ی (۲) و  $N$  تعداد مستندات در لیست نهایی می باشند.

امتیاز تخصیصی به اسناد بر اساس تعداد n-gram های مورد علاقه ی کاربر در ابتدا تعریفی کوتاه از n-gram ارائه می نمائیم. یک n-gram یک زیر رشته به طول n از یک رشته از کلمات می باشد. از مفهوم n-gram در حیطه های وسیعی از جمله پردازش زبان های طبیعی، و آنالیز توالی های ژنتیکی استفاده می شود. می توان نشان داد که یک n-gram، یک مدل مارکوف از مرتبه ی n-۱ می باشد. فرکانس تکرار یک n-gram در یک متن تا حدود زیادی می تواند ارتباط آن متن با عبارت مورد نظر را نشان دهد. البته این بحث در حیطه ی همبافت\_شناسی متون مطرح می گردد و دارای ظرافت های بسیاری است. مثلا عدم حضور n-gram هایی که کاربر حضور آنها را در متن نمی پسندد گاهی به اندازه ی حضور عبارات مورد علاقه اهمیت دارد. در این میان حجم سند مورد نظر نیز از اهمیت زیادی برخوردار است. به

این معنی که تعداد n-gram ها در واحد حجم سند پارامتر مناسبی محسوب می شود و تعداد آنها به تنهایی ملاک مطلوبی نیست.

اگر فرض کنیم که تعداد n-gram های مورد علاقه ی کاربر T عدد می باشد، برای هر سند از بابت تعداد n-gram ها T امتیاز مختلف به دست می آید که این امتیاز ها بوسیله ی یک عملگر ترکیب اطلاعات ترکیب می

گردند و به این ترتیب به هر سند از این بابت یک امتیاز تعلق می گیرد (شکل ۱) که این سند را  $\{U_k\}_{k=1, \dots, N}$  می نامیم و عامل از آن در تصمیم گیری نهایی استفاده خواهد کرد. در اینجا از ذکر جزئیات امر خودداری می گردد. برای آشنایی با جزئیات این کار، خواننده می تواند به (Kavousi and Moshiri, ۲۰۰۷) مراجعه نماید. در آن مرجع از عملگر OWA برای ترکیب اطلاعات استفاده شده و جزئیات به تفصیل شرح داده شده است.

امتیاز تخصیصی به اسناد بر اساس پراکندگی n-gram های مورد علاقه ی کاربر (میزان دوری و نزدیکی یک n-gram به ابتدای متن) میزان دوری و نزدیکی n-gram ها به ابتدای متن در یک سند و یا در بیانی جامع تر، نحوه ی توزیع n-gram ها در متن یک سند پارامتر مهمی در سنجش میزان ارتباط یک متن با n-gram مورد علاقه ی کاربر می باشد. البته خود ایم موضوع جای بحث فراوانی دارد. زیرا این امکان وجود دارد که قسمت های پایانی یک سند طولانی ارتباط عمیقی با موضوع مورد علاقه ی کاربر داشته باشد. در این حالت اگر به تنهایی از این پارامتر برای رتبه بندی مستندات استفاده شود نتایج ضعیفی به دست خواهد آمد. یکی از فوائد ترکیب اطلاعات همین امر است که حتی اگر یکی از منابع اطلاعاتی به دلایلی اطلاعات غیر دقیق و توأم با عدم قطعیت ارائه نماید، باز هم این امکان وجود دارد تا با استفاده از سایر منابع اطلاعاتی نتایجی قابل قبول بدست آید. رای محاسبه این امتیاز برای هر سند ابتدا تمام اسناد باید هم حجم شوند. یعنی برای آنکه متریک مناسبی برای محاسبات داشته باشیم فرض می کنیم که همه ی مستندات تعداد کلمات برابر

دارند. سپس با این فرض مکان جدید هر  $n$ -gram را در هر سند محاسبه می کنیم. در این حالت نیز از ذکر جزئیات عمل محاسبه پرهیز می گردد. برای آگاهی از الگوریتم این کار می توان به (Kavousi and Moshiri ۲۰۰۷) مراجعه نمود. به این ترتیب در این مرحله نیز چنانچه تعداد  $n$ -gram ها برابر  $T$  باشد، برای هر سند  $T$  امتیاز بدست می آید که عامل باید این امتیازات را با یک عمگر ترکیب اطلاعات با هم ترکیب نماید. در همان مرجع نحوه ی انجام این کار با استفاده از عملگر OWA تشریح گردیده است. امتیاز نهایی حاصل از اعمال عملگر OWA سومین امتیاز دخیل در رتبه بندی نهایی را شکل می دهد و آن را با

$$\left\{ \begin{array}{l} P_k \\ k=1, \dots, N \end{array} \right.$$

نشان می دهیم (شکل ۱).

امتیاز تخصیصی به هر سند بر اساس سرویس دهنده ای که آن سند را بازیابی نموده است ایده اصلی در نظر گرفتن این پارامتر به عنوان یکی از امتیاز های مورد استفاده در رتبه بندی نهایی مستندات آن است که سرویس دهنده های اطلاعاتی و موتورهای جستجوی اینترنتی هریک با فلسفه های گوناگونی طراحی شده اند و نقاط قوت و ضعف یکسانی ندارند. بنابر این آگاهی از نقاط قوت و ضعف این سرویس دهنده ها در بازیابی اطلاعات مرتبط با هر خوشه ی موضوعی می تواند عامل را در تصمیم گیری بهتر یاری نماید. به عنوان مثال این امکان وجود دارد که یک منبع اطلاعاتی، اطلاعات مربوط به نوسانات قیمت طلا را بسیار دقیق در اختیار بگذارد و سرویس دهنده ای دیگر اطلاعات مربوط به بازار جهانی سیمان را بسیار دقیق منعکس نماید. طبیعی است در این شرایط مستندات که از منبع اول استخراج شده اند احتمالاً برای فردی که به بازار طلا علاقه مند است جالب تر از اطلاعات مرتبط مستخرج از منبع دوم می باشد. عامل طراحی شده باید قادر باشد تا به مرور زمان مدل رفتاری هر سرویس دهنده اطلاعاتی را در قبال خوشه های موضوعی مختلف شناسایی نماید و به هر سند بر اساس پیشینه ی رفتار سرویس دهنده ای که آن سند را استخراج نموده است یک امتیاز تخصیص دهد. در اینجا مدلی ریاضی برای این کار ارائه می گردد. در ابتدا عامل در هر خوشه ی موضوعی امتیاز

۰،۵ (نصف امتیاز بیشینه) را به هر سرویس دهنده تخصیص می دهد. همچنین هر بار که خوشه ی موضوعی جدیدی به عامل معرفی گردد، عامل به تمام سرویس دهنده ها در آن خوشه ی موضوعی امتیاز ۰،۵ را نسبت می دهد. هر بار که عامل هوشمند نتایج نهایی را به کاربر نشان می دهد، در صورت رضایت کاربر از نتایج، الگوریتمی توسط عامل اجرا می شود که طی آن امتیاز های تخصیصی به سرویس دهنده ها در آن خوشه ی موضوعی اصلاح می شود. یعنی چنانچه نتایج بازیابی شده توسط یک منبع اطلاعاتی در لیست نهایی رتبه های بهتری را اشغال کرده باشد، امتیاز آن سرویس دهنده در آن دسته بندی موضوعی متناسب با وضعیت جدید افزایش خواهد یافت و چنانچه نتایج یک سرویس دهنده ضعیف باشد امتیاز آن کاهش خواهد یافت. بدین ترتیب پس از گذشت مدتی، عامل برای هر دسته بندی موضوعی، مدلی رفتاری از سرویس دهنده های اطلاعاتی بدست می آورد که با ارائه ی آن به کاربر او را در انتخاب موتورهای جستجو و سرویس دهنده های اطلاعاتی کارآمدتر در هر زمینه یاری می نماید. در ادامه نحوه ی انجام این عملیات توسط عامل را شرح خواهیم داد. فرض کنیم  $s_i^{(t)}$  امتیاز سرویس دهنده ی  $i$  ام در مرحله جاری باشد. ما به دنبال تابعی هستیم که با دریافت مقادیر  $s_i^{(t)}$  و  $w_i$  نیز امتیاز تخصیص داده شده به هر سرویس دهنده در مرحله ی جاری ( $w_i$ ) که با توجه به تعداد و رتبه ی مستندات بازیابی شده توسط این سرویس دهنده در لیست نهایی محاسبه می گردد، بتواند  $s_i^{(t+1)}$  (امتیاز سرویس دهنده ی  $i$  ام در مرحله ی بعد) را محاسبه نماید. یعنی:

$$s_i^{(t+1)} = f(s_i^{(t)} \cdot w_i) \quad (4)$$

با تعیین مناسب تابع  $f$  می توان مقادیر امتیاز هر سرویس دهنده را به روز نمود. اما باید در نظر داشت که پارامتر زمان نیز به طور غیر مستقیم در این تابع نقش دارد. در واقع اهمیت این تابع از آن جهت است که بتواند متوسط عملکرد رفتار هر سرویس دهنده ی اطلاعاتی را بازنمایی نماید. ساده ترین راه حل در این مورد که به

نحو مطلوبی نیز عمل کرده است، محاسبه ی میانگیت امتیازات نسبی هر سرویس دهنده در هر خوشه می باشد. برای این منظور رابطه ی زیر مناسب می باشد:

$$s_i^{(t+1)} = s_i^{(t)} - \frac{t}{t+1} + \frac{\psi_i}{t+1} \quad (5)$$

که در آن  $t$  نشاندهنده ی تعداد دفعاتی است که در خوشه ی  $i$ ام به سرویس دهنده ی  $i$ ام کوئری ارسال شده است و نتایج حاصل از آن مورد استفاده قرار گرفته است. در این صورت اگر امتیاز ناشی از سرویس دهنده ی بازیابی کننده ی سند  $k$ ام را با  $Q_k$  نمایش دهیم خواهیم داشت:

$$Q_k = s_i^{(t)} \quad (6)$$

از این امتیاز به عنوان چهارمین امتیاز دخیل در رتبه بندی نهایی مستندات استفاده خواهد شد.

### محاسبه ی امتیاز نهایی

همانطور که در قسمت های قبلی شرح داده شد برای هر سند چهار امتیاز به شرح زیر محاسبه می گردد:

- امتیاز ناشی از ترکیب لیست های رتبه بندی شده برای سند  $k$ ام  $V_k =$
- امتیاز تعداد  $n$ -gram های موجود در سند  $k$ ام  $U_k =$
- امتیاز پراکندگی  $n$ -gram ها در سند  $k$ ام  $P_k =$
- امتیاز سرویس دهنده ی بازیابی کننده ی سند  $k$ ام در خوشه ی موضوعی  $z$ ام  $Q_k =$

در اینجا نیز برای محاسبه ی امتیاز نهایی تخصیصی به هر سند، استفاده از یک اپراتور ترکیب اطلاعات مانند OWA ضروری می باشد (شکل ۳).

مکانیزم انجام این کار نیز در (Kavousi and Moshiri, ۲۰۰۷) شرح داده شده است.

واضح است که اهمیت همه ی معیار های چهار گانه ی فوق در تصمیم گیری یکسان نیست.  $\{A_i\}_{i=1, \dots, 4}$  ها ضرایب اهمیتی هستند که به هر یک از معیارهای چهارگانه ی فوق تعلق می گیرند، و در طول زمان تغییر می کنند. به عنوان مثال امتیاز ناشی از سرویس دهنده های اطلاعاتی در ابتدا که هنوز مدل رفتاری سرویس دهنده ها ناقص است، زیاد معتبر نیست و بنابر این اهمیت کمتری خواهد داشت. ولی به مرور این اهمیت بیشتر خواهد شد. در اینجا جلوه ی دیگری از عملکرد تطبیقی عامل مورد نظر دیده می شود.

برای آموزش  $A_i$  ها می توان استراتژی های مختلفی در پیش گرفت. در ادامه یکی از روش های ممکن که توسط نگارندگان پیشنهاد گردیده است، توضیح داده خواهد شد. مکانیزم این کار به این صورت است که پس از اینکه در یک دوره ی کامل، لیست نهایی مستندات بر اساس امتیاز ها آماده شد، امتیاز نهایی با هر یک از مقادیر  $Q_i, U_i, V_i$  و مقایسه می شود و به هر کدام از این امتیاز ها که نزدیکتر باشد، ضریب اهمیت آن معیار باید افزایش بیشتری داشته باشد. برای این منظور میانگین خطای امتیاز نهایی از امتیاز هر یک از معیارهای چهارگانه



ملاک عمل قرار می گیرد. هربار این مقدار خطا با توجه به داده های جدید میانگین گیری می شود. مطلب فوق

$$\left\{ \begin{aligned} e_1^{(t+1)} &= e_1^{(t)} \frac{t}{t+1} + \frac{\frac{1}{N} \sum_{k=1}^N |D_k - V_k|}{t+1} \\ e_2^{(t+1)} &= e_2^{(t)} \frac{t}{t+1} + \frac{\frac{1}{N} \sum_{k=1}^N |D_k - U_k|}{t+1} \\ e_3^{(t+1)} &= e_3^{(t)} \frac{t}{t+1} + \frac{\frac{1}{N} \sum_{k=1}^N |D_k - P_k|}{t+1} \\ e_4^{(t+1)} &= e_4^{(t)} \frac{t}{t+1} + \frac{\frac{1}{N} \sum_{k=1}^N |D_k - Q_k|}{t+1} \end{aligned} \right.$$

را می توان به صورت زیر بیان نمود: (۷)

که در آن  $\beta_i$  نشاندهنده ی تعداد دفعات استفاده از عامل از ابتدا (زمان ریست کردن امتیازات) تا کنون می باشد.

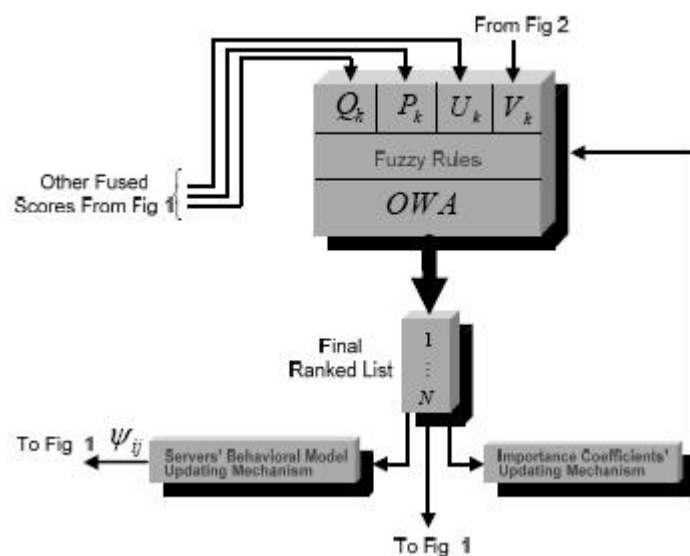
بنابر این هرچه مقدار خطای  $e_i^{(t+1)}$  برای معیار  $i$  ام کوچکتر باشد آن معیار، معیاری مناسب تر می باشد. بنابر این

مقادیر  $(0 \leq \beta_i \leq 1; i=1,2,3,4)$  به صورت زیر اصلاح می شوند:

$$\left\{ \begin{aligned} \beta_i^{(t+1)} &= \frac{1 - e_i^{(t+1)}}{\sum_{j=1}^4 (1 - e_j^{(t+1)})} \\ i &= 1, \dots, 4 \end{aligned} \right.$$

(۸)

همانطور که مشخص است شرط  $\sum_{i=1}^4 \beta_i = 1$  نیز برقرار می باشد.



شکل ۳: ترکیب امتیازاتی که نحوه تشکیل آنها در شکل های ۱ و ۲ نشان داده شده اند و ایجاد لیست رتبه بندی شده ی نهایی و نیز به روز رسانی ضرایب اهمیت و همچنین مدل رفتاری سرویس دهنده های اطلاعاتی

- Automatic Teller Machine
- Inter-organizational Systems
- Electronic Data Interchange
- Electronic Data Processing Era
- Management Information System Era
- Society for Worldwide Inter- bank Financial Telecommunication
- Electronic Funds Transfer
- Behavioral Model
- Meta Search Engine
- Data/Information Fusion
- Customized
- Agent
- Linguistic Concept
- Query
- Ordered Weighted Averaging Operator
- Cluster
- Micro Electro Mechanical Systems
- Principle of Insufficient Reason
- Context
- Ordered Weighted Averaging Operator

## منابع

### منابع فارسی

کاوه کاوسی، بهزاد مشیری (۱۳۸۳)، بهبود روش های هوشمند عامل ترکیب اطلاعات، نشریه دانشکده فنی دانشگاه تهران، جلد ۳۸، شماره ۱، ص ۸۵-۹۸.

### منابع انگلیسی

- Design of an Intelligent Information Fusion Agent Architecture , (Kavousi K., Moshiri B.(۲۰۰۷) Dynamic Environments , Proc. of Fusion ۲۰۰۷, Québec city, for Information Retrieval From .Canada
- Molla. A., and Licker P.S.(۲۰۰۱), E-commerce Systems Success: An Attempt to Extend and Re-Specify the Delone and McLean Model of IS Success, Journal of Electronic Commerce Research, Vol. ۲, No. ۴, pp. ۴۸-۵۸.
- Seddon, P.A. (۱۹۹۷), “respecification and extension of the Delone and McLean model of IS success”, Information Systems Research, Vol. ۸ No. ۳, pp. ۲۴۰-۵۳.
- Senn, James A(۲۰۰۰). “Business-To-Business E-Commerce”, Information SystemsManagement, Spring.
- Vadapalli A., Ramamurthy K.(۱۹۹۸), Business use of the Internet: an analytical framework and exploratory case study, International Journal of Electronic Commerce ۲ (۲), pp. ۷۱-۹۴.
- Yager, A. Rybalov(۱۹۹۷), On the Fusion of Documents from Multiple Collection Information Retrieval Systems, Journal of the American Society for Information Science.
- Yager RR, Kacprzyk J.(۱۹۹۷), The Ordered Weighted Averaging Operators: Theory and Applications. Norwell: Kluwer.