

طراحی سیستم‌های بازیابی اطلاعات بهینه در نرم‌افزارهای کتابخانه‌ای و اطلاع‌رسانی

فصلنامه اطلاع‌رسانی، دوره ۱۶، شماره ۱ و ۲

نوشته: علی گزنی

کارشناس ارشد کتابداری در کتابخانه منطقه‌ای علوم و تکنولوژی شیراز

کلیدواژه‌ها

سیستم‌های بازیابی اطلاعات، سازماندهی اطلاعات، بازیابی اطلاعات، نمایش اطلاعات

چکیده

بررسی فنی نرم‌افزارهای مختلف کتابخانه‌ای نشان می‌دهد که سرویس بازیابی اطلاعات در این نرم‌افزارها به صورت بهینه طراحی و پیاده‌سازی نشده‌اند. هدف پژوهش حاضر این بوده است که یک سیستم بازیابی اطلاعات بهینه را تعریف کند، راهنمایی برای طراحان سیستم‌های بازیابی اطلاعات، محققان و دانش‌پژوهان این سیستم‌ها را ارائه دهد و ارزیابی سیستم‌های بازیابی اطلاعات را ممکن سازد. در این مقاله به موضوعات زیر پرداخته شده است: امکانات و قابلیت‌های سیستم‌های بازیابی اطلاعات، بازخورد در سیستم‌های بازیابی اطلاعات، نحوه نمایش اطلاعات خروجی، ارزیابی سیستم بازیابی اطلاعات، و نمایش اطلاعات.

مقدمه

مسئله بازیابی اطلاعات از زمانی آغاز شد که بشر سعی کرد محیط پیرامون خود را کنترل کند یا حداقل از فشارهای خارجی که باعث نابودی او می‌شدند، جلوگیری کند. بشر برای ایجاد محیطی مطلوب برای ادامه بقا به اتخاذ تصمیمات سریع، صحیح و دقیق نیاز داشت. کیفیت این تصمیمات به توانایی تصمیم‌گیرنده در حل مسائل وابسته بود، ولی قبل از آن به میزان ارتباط و کیفیت اطلاعاتی وابسته بود که تصمیم‌گیرنده برای حل مشکل فراهم آورده بود. بتدریج و در طول تاریخ، جمع‌آوری، سازماندهی و نگهداری امری متداول و مرسوم شد. در چند دهه اخیر، بازیابی رایانه‌ای مورد توجه خاصی قرار گرفته که علت آن را می‌توان چهار عامل دانست: افزایش اهمیت زمان، تغییر در کمیت و کیفیت اطلاعات، تغییر در ماهیت نیازهای اطلاعاتی، تغییر در اهمیت منابع اطلاعاتی. این عوامل، تغییر و توسعه در سیستم‌های بازیابی اطلاعات را به دنبال داشتند و موجب مطرح شدن بحث‌هایی از جمله لزوم تغییر در ابزارهای اطلاعاتی، ایجاد و سیستم‌های ارتباطی جدید و استفاده از آنها، تغییر معانی و اشاعه اطلاعات شده‌اند.

امروزه رشد تصاعدي و اهمیت حیاتی اطلاعات، لزوم صرفه‌جویی در وقت و هزینه جستجوکنندگان، جوان بودن شبکه‌های گسترده بخصوص از لحاظ سرعت، لزوم دستیابی سریع، جامع و مانع به اطلاعات خاص مورد نیاز، از جمله مسائل مهمی هستند که اهمیت یک فرآیند بازیابی حساب شده، کنترل شده و کامل را آشکار می‌سازند.

تعریف مسئله

در حال حاضر در سطح کشور نرم‌افزارهای کتابخانه‌ای مختلفی طراحی شده‌اند، از جمله می‌توان از نرم‌افزارهای پارس آذرخش، نوسا، کاوش، بهمین برنا، گنجینه، ارغون و کتیبه نام برد. بررسی فنی این سیستم‌ها نشان می‌دهد که سرویس بازیابی اطلاعات در این نرم‌افزارها به صورت بهینه طراحی و پیاده‌سازی نشده و کارایی مطلوبی ندارد. با توجه به استفاده تعداد زیادی از مراکز کتابخانه‌ای و اطلاع‌رسانی از این نرم‌افزارها و نقشی که در بازیابی و نیز گردش اطلاعات در جامعه دارند، می‌توانند در مجموعه اطلاع‌رسانی کشور موثر باشند. بدین لحاظ ارائه مبنای صحیح طراحی و ارزیابی سیستم‌های بازیابی اطلاعات دارای اهمیت می‌باشد.

تعاریف عملیاتی

سیستم‌های بازیابی اطلاعات: سیستم‌هایی که به منظور بازیابی و پردازش داده‌های ساختار نیافته طراحی شده‌اند و به لحاظ نوع سازماندهی، ساختار پایگاه‌ها، راهبرد بازیابی، گروه‌های خدمت‌گیرنده، و فرآیندی که در طی آن به درخواست‌های اطلاعاتی کاربران پاسخ می‌دهند، از دیگر سیستم‌های موجود متمایز می‌شوند.

سازماندهی اطلاعات: تجزیه و تحلیل داده‌های موجود در یک سیستم اطلاعاتی به منظور آماده‌سازی آن‌ها برای انجام بازیابی اطلاعات.

بازیابی اطلاعات: چرخه‌ای که با فرمول‌بندی نیاز اطلاعاتی شروع می‌شود، با جستجو ادامه می‌یابد، و با بازیابی و انتخاب اطلاعات مناسب خاتمه می‌یابد.

نمایش اطلاعات: محاسبه و انتقال علائم به شکل هندسی به صورتی قابل درک و مشاهده توسط انسان، به منظور فهم و کشف روابط پنهان بین عناصر مختلف داده‌ها.

پرسش‌های اساسی پژوهش

نحوه سازماندهی اطلاعات، ساخت فایل‌های شاخص و مقلوب و سایر فایل‌های کمکی در سیستم‌های بازیابی

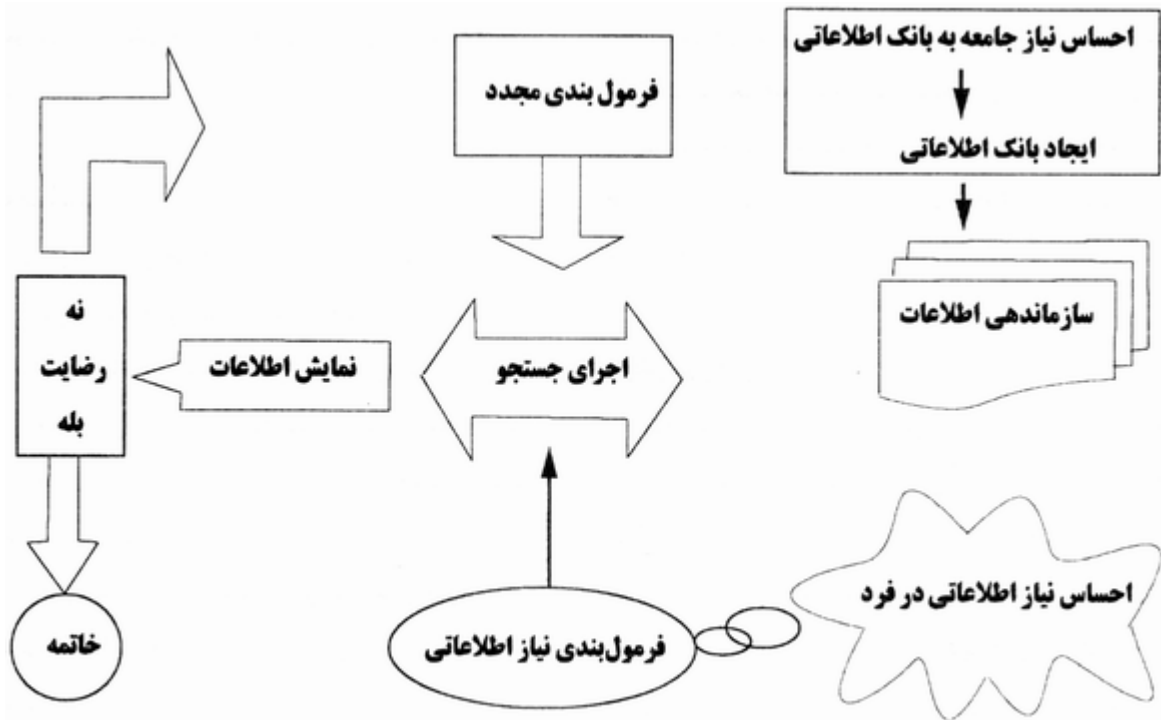
اطلاعات بهینه به چه صورت می‌باشد؟
چه امکانات و قابلیت‌هایی در سیستم‌های بازیابی اطلاعات بهینه لازم است تا کاربران بتوانند به بهترین شکل ممکن به بیان نیازهای اطلاعاتی خود پردازند؟ به بیان دیگر، چه قابلیت‌ها و امکاناتی باید در سیستم وجود داشته باشد تا کاربران بتوانند یک راهبرد بازیابی صحیح و دقیق را ایجاد کنند؟
بازخورد در سیستم‌های بازیابی اطلاعات بهینه باید به چه صورت باشد؟ یا به بیان دیگر، سیستم بازیابی اطلاعات چگونه می‌تواند کاربران خود را به مرتبط‌ترین اطلاعات، در رابطه با نیاز اطلاعاتیشان هدایت کند؟
نحوه نمایش اطلاعات در خروجی‌های حاصل از نتایج جستجو در سیستم‌های بازیابی اطلاعات بهینه چگونه کاربر را به انتخاب اطلاعات مرتبط هدایت می‌کند؟ یا به بیان دیگر، شیوه‌های بهینه نمایش اطلاعات در خروجی‌های حاصل از نتایج جستجو کدامند؟
چگونه می‌توان یک سیستم بازیابی اطلاعات را مورد ارزیابی قرار داد؟

روش پژوهش

پژوهش حاضر یک پژوهش از نوع تحلیلی کاربردی می‌باشد. بخشی از پژوهش به صورت بررسی کتابخانه‌ای، بر روی اجزای تشکیل دهنده سیستم‌های بازیابی اطلاعات صورت گرفته است. در این بخش به مراحل زیر توجه شده و اطلاعات لازم گردآوری شده است: ساختار داده‌ها، ساختار فایل‌های اطلاعاتی، ساختار فایل‌های نمایه، نمایه‌سازی خودکار، پردازش در سیستم‌های بازیابی اطلاعات، قابلیت‌های سیستم‌های بازیابی اطلاعات، دسته‌بندی واژه‌ها، راهبرد بازیابی، الگوریتم‌های بازیابی، نمایش اطلاعات، ارزیابی سیستم‌های بازیابی اطلاعات. در بخش دیگری از پژوهش به بررسی چند نرم‌افزار داخلی و خارجی پرداخته شده است. انجام این بررسی با توجه به اهداف پژوهش و به منظور تشریح وضعیت کنونی سیستم‌های بازیابی اطلاعات داخلی و خارجی می‌باشد که به همین منظور نمونه‌هایی انتخاب و تشریح می‌شوند. از میان نمونه‌های انتخاب شده در ایران نرم‌افزار کتابخانه‌ای پارس آذرخش - به لحاظ قدمت کارکرد، تعداد مراکز استفاده‌کننده، توانایی انجام جستجو در مقایسه با سایر هم‌تاهای خود - و نرم‌افزار کتابخانه‌ای پویا - به دلیل جدید بودن و توانمندی سیستم فوق - مورد بررسی قرار خواهند گرفت. در میان نرم‌افزارهای خارجی نرم‌افزار «دایالوگ» (۱) به لحاظ قدمت کارکرد و اهمیت آن مورد بررسی قرار خواهد گرفت و چند بسته نرم‌افزاری دیگر از جمله «اواک» (۲) و «داوکواست» (۳) نیز به لحاظ دارا بودن موتورهای جستجوی مبتنی بر پردازش به زبان طبیعی و محیط گرافیکی، به صورت ضمنی مورد بررسی قرار خواهند گرفت.
به لحاظ کاربردی و عملی بودن موضوع پژوهش، فهم بهتر مطالب و مفاهیم، کار عملی با زبان‌های برنامه‌نویسی ++C/C، فاکس پرو و زبان پرس و جوی ساختاری (Structure Query language) صورت گرفت و یک سیستم بازیابی اطلاعات نمونه طراحی شد. در نهایت براساس مطالعات نظری و عملی، عوامل موثر در یک نظام بازیابی بهینه شناسایی و معرفی شدند.

یافته‌های پژوهش

براساس یافته‌های پژوهش و با بهره‌گیری از مدل‌های موجود مانند مدل «بوک‌استین» (۱۹۸۲) و «مدل بویس» و «کرافت» (۱۹۸۵)، یک مدل که نشان‌دهنده ساختار یک سیستم بازیابی اطلاعات می‌باشد ارائه شد. این مدل در شکل ۱ مشاهده می‌شود.
با توجه به مدل ارائه شده، بحث بر سه محور سازماندهی، بازیابی و نمایش اطلاعات متمرکز شده که در ادامه به بررسی یافته‌ها در این زمینه‌ها می‌پردازیم:



شکل (۱): مدلی برای سیستم‌های بازیابی اطلاعات
 شکل (۱): مدلی برای سیستم‌های بازیابی اطلاعات

۱- سازماندهی اطلاعات

هر سیستم اطلاعاتی دارای یک مبناي خاص برای تجزیه و تحلیل اطلاعات می‌باشد که سیستم براساس آن به تفسیر اطلاعات و مطابقت بین اقلام و درخواستهای اطلاعاتی می‌پردازد و بدین ترتیب بازیابی صورت می‌گیرد. این تجزیه و تحلیل، سازماندهی نامیده می‌شود. به لحاظ اهمیت سازماندهی اطلاعات، سیستم‌های بازیابی اطلاعات به دو دسته هوشمند و غیرهوشمند تقسیم می‌شوند. سیستم‌های بازیابی اطلاعات هوشمند برای اولین بار بین سال‌های ۱۹۶۲ تا ۱۹۶۵ در دانشگاه هاروارد طراحی و به کار گرفته شدند («سالتون»، ۱۹۶۴؛ «سالتون»، ۱۹۶۵). سیستم‌های هوشمند سیستم‌هایی می‌باشند که در آنها تمام پردازش‌ها بر روی متن به صورت خودکار انجام می‌شود. هنگام پردازش متن‌هایی با زبان طبیعی، مسئله پیچیدگی زبان و بیقاعدگی‌هایی که در حوزه نحوی و معنایی وجود دارند خود را نشان خواهند داد که باید این مشکلات را مد نظر قرار داد. این مشکلات همان طوری که «سالتون» (۱۹۶۴) نیز به آن اشاره دارد هنگام اختصاص خودکار توصیفگرها به متن مشاهده می‌شوند. این مشکلات عبارت‌اند از واژه‌های مترادف، عملکردهای نحوی، ترکیبات نحوی، ارجاعات غیرمستقیم، متفاوت بودن معانی متناسب با هر متن، ارتباط معنایی واژه با متن‌های قبلی، و تغییر معنای واژه‌ها در طول زمان. در این میان واژه‌نامه‌ها قادر نیستند ابهامات و پیچیدگی‌های موجود در زبان را به صورت کامل برطرف کنند، اما این بیقاعدگی‌ها را کاهش می‌دهند. انواع مهمتر واژه‌نامه‌هایی که در سیستم‌های بازیابی به کار می‌روند عبارت‌اند از: (۱) سیاهه بازدارنده (۲)؛ شامل واژه‌هایی است که نباید مورد استفاده قرار گیرند. (۲) واژه‌نامه ریشه لغات و سیاهه پسوندها؛ ریشه لغات و پسوندهای آنها به صورت جداگانه در هر مدخل نگهداری می‌شوند. یکی از روش‌های متداول در ریشه‌یابی، استفاده از روش «لاوینز» می‌باشد. (۳) واژه‌نامه عبارات؛ واژه‌ها براساس رخداد همزمان و میزان این رخداد، به صورت یک عبارت در نظر گرفته می‌شوند. با واژه‌نامه کیفیت و کارایی یک تحلیل محتوایی بالا می‌رود. با تعیین عبارات، حالات مبهم معنایی در متن بمراتب کاهش خواهد یافت. به عنوان مثال واژه «برنامه‌نویسی» و «زبان» هر یک به تنهایی بیان‌کننده حالات مختلفی هستند، اما قطعاً «زبان برنامه‌نویسی» یک معنای مشخصی را القا می‌کند. (۴) واژه‌نامه مفاهیم (مترادف‌ها)؛ در هر مدخل تعدادی از واژه‌های مترادف یا مفاهیم هم‌طبقه آورده شده است. گسترش دامنه مفاهیم و اصطلاحات در فرمول جستجو از طریق جایگزین کردن واژه‌های مبهم با واژه‌های هم‌طبقه، از طریق این واژه‌نامه صورت می‌گیرد. (۵) سلسله مراتب مفاهیم؛ مشابه طرح‌های رده‌بندی کتابخانه‌ای می‌باشد؛ با حرکت به بالای رده به قسمت‌های اعم و با حرکت به قسمت‌های پایینی به قسمت‌های اخص خواهیم رسید. به منظور سازماندهی واژه‌ها باید دامنه شمول واژه‌ها و نوع دسته‌بندی مفاهیم را مدنظر قرار داد. واژه‌های موجود در متون به سه دسته واژگان عمومی، واژگان با رخداد بالا، و واژگان با رخداد پایین قابل تقسیم می‌باشند که باید در مورد شمول هر یک، تصمیم‌گیری صورت گیرد. نوع طبقه‌بندی واژه‌ها و مفاهیم نیز از مسائلی هستند که در ضرایب بازیابی و دقت، تاثیر بسزایی خواهند داشت. بنابراین باید در مورد اعم یا اخص بودن طبقات، تصمیم‌گیری صورت گیرد. پاسخگویی به مسائل فوق، وابسته به سیاستگذاری سیستم در رابطه با ضرایب بازیابی و دقت می‌باشد. اگر خواهیم ضریب بازیابی را بالا ببریم باید تمام واژه‌هایی که به نحوی در بازیابی موثر باشند، به واژه‌نامه اضافه شوند. در مقابل اگر خواهیم ضریب دقت بالاتری داشته باشیم و اقلام مرتبطتری مورد جستجو قرار گیرند، باید به سمت

واژه‌نامه‌های تخصصی‌تر حرکت کنیم.

ساخت واژه‌نامه‌ها به صورت خودکار با دو روش تمام خودکار و نیمه خودکار قابل اجرا می‌باشد. به دلیل گستردگی کار، این کار به صورت کمیته‌ای و گروهی انجام می‌شود. این کمیته تعدادی سؤال را تعیین و مطرح می‌نماید و در نهایت، استانداردهای لازم را برای تدوین واژه‌نامه پیشنهاد می‌کند (دوول، ۱۹۶۵). اغلب روش‌های مورد استفاده در ساخت واژه‌نامه‌ها با روش تمام خودکار، به واژه‌های موجود در مجموعه مدارک وابسته می‌باشند.

(دوول، ۱۹۶۵؛ دنیس، سالتون، ۱۹۶۶). عملی بودن روش خودکار بیشتر از آن که به یک شاخه موضوعی خاص وابسته باشد، به مجموعه مدارک مورد پردازش بستگی دارد. بنابراین، باید در رویه‌های پردازشی تغییراتی اعمال شود و از فضاوت انسان‌ها نیز استفاده کنیم. روش‌های مورد استفاده براساس اهداف سیستم‌ها متفاوت‌اند (اسپارک جونز، ۱۹۶۵؛ لوری، آبراهام، ۱۹۶۵؛ رایزرن، ۱۹۶۵).

به منظور بالا بردن کارایی سیستم‌های بازیابی اطلاعات از روش‌های محاسبه همبستگی و خوشه‌بندی اطلاعات نیز استفاده می‌گردد. محاسبه همبستگی براساس دفعات رخداد همزمان واژه‌ها، به عنوان معیاری برای سنجش میزان مشابهت و همبستگی بین دو متن به کار می‌رود. میانگین حاصل از این تعداد رخداد، بنحو شایسته‌تری بیان‌کننده این همبستگی خواهد بود. اگر بخواهیم این همبستگی مقدار بالاتر و معنادارتری را نشان دهد باید این ارزیابی براساس لغات مشترک و غیرمشترک بین دو متن صورت گیرد. به منظور انجام این محاسبه، فرمول‌های مختلفی پیشنهاد شده که در پژوهش حاضر به بررسی فرمول «سالتون» و «مک‌گیل» (۱۹۸۲) پرداخته شد که به قرار زیر است:

$$SIM(j,k) = \frac{\sum_{i=1}^n t_{ij} \cdot t_{ik}}{\sum_{i=1}^n (t_{ij})^2 + \sum_{i=1}^n (t_{ik})^2 - \sum_{i=1}^n t_{ij} \cdot t_{ik}}$$

در این فرمول t_{ij} نشان‌دهنده ضریب وزنی (دفعات تکرار) واژه i در متن j ، و n تعداد کل واژه‌های موجود در واژه‌نامه است که می‌تواند مجموع واژه‌های موجود در دو مدرک باشد. نکته مهم این است که اگر واژه i در مدرک j موجود نباشد، t_{ij} برابر صفر خواهد بود و اگر واژه i هیچ یک از دو مدرک وجود نداشته باشد در آن صورت نیز $t_{ij} = t_{ik} = 0$ خواهد بود. خوشه‌بندی برای دسته‌بندی رکوردهای اطلاعاتی مشابه با هم مورد استفاده قرار می‌گیرد. خوشه‌بندی از طریق فرمول‌های مختلفی قابل اجرا است. یکی از ساده‌ترین فرمول‌های موجود که توسط «بیر» ارائه شده در پژوهش حاضر مورد بررسی قرار گرفته که به قرار زیر است:

$$S_{ij} = \frac{T_{ij}}{T_{ii} + T_{jj} - T_{ij}}$$

در این فرمول T_{ij} بیانگر تعداد واژه‌های مشترک است که در دو رکورد j و i وجود دارند. T_{ii} و T_{jj} به ترتیب نشان‌دهنده کل واژه‌های موجود در رکوردهای i و j می‌باشند. اولین قدم در ایجاد یک خوشه، تشکیل یک ماتریس از مشخصه‌های رکوردها می‌باشد. به هر رکورد به صورت مجزا یک سطر اختصاص داده می‌شود و هر واژه یا مشخصه به صورت یک ستون نمایش داده می‌شود.

۲. بازیابی اطلاعات

مطالعات انجام شده توسط «پنی‌من» (۱۹۷۵) و «چایمن» (۱۹۸۱) در موضوع جستجوی اطلاعات نشان داد که جستجوگران بانک‌های متنی و کتابشناختی از یک چرخه دستورها به صورت مشترک استفاده می‌کنند. در این مطالعات تأکید بر دستورهایی پراستفاده‌تر بود تا این که بخواهند سلسله مراتب استفاده از این دستورها را مورد توجه قرار دهند. چرخه دستورها شامل ۷ مرحله زیر است:

۱) انتخاب یک بانک اطلاعاتی برای انجام جستجو

بر مبنای بانک‌های موجود و نیاز اطلاعاتی کاربر، یک بانک اطلاعاتی توسط او انتخاب می‌شود. آنگاه سیستم خلاصه‌ای از اطلاعات، مانند محدوده تاریخی رکوردها، تعداد رکوردها، قیمت و... را در اختیار کاربر قرار می‌دهد.

۲) جستجو برای واژه‌های مورد نظر در بانک واژگان

قبل از انجام جستجو و برای فرمول‌بندی جستجو، کاربر می‌تواند هر یک از واژه‌نامه‌های موجود در سیستم را که در بخش سازماندهی اطلاعات به آن اشاره شد مورد استفاده قرار دهد. واژه‌نامه‌ها نقش بسیار مهمی در بازیابی اطلاعات ایفا می‌کنند، حداقل استفاده‌ای که از بانک واژگان به عمل می‌آید، این است که کاربر خواهد فهمید که آیا واژه‌های مورد نظر او در بانک اطلاعاتی وجود دارند؟ شکل صحیح آن‌ها به چه صورت می‌باشد؟ واژه در چند رکورد، در کدام فیلدها و به چه میزان تکرار شده است؟ اگر واژه درخواستی کاربر، اشتباه تایپ شده باشد یا در بانک واژگان وجود نداشته باشد، مقدار عددی صفر برای آن نمایش داده می‌شود. کاربر با استفاده از این فایده‌ها می‌تواند به واژه‌های مرتبط با واژه‌های مورد نظر خود دست یابد و به اخص یا اعم‌تر کردن فرمول جستجو بپردازد.

۳) ایجاد فرمول جستجو و انجام جستجو

فرمول‌بندی صحیح و دقیق نیازهای اطلاعاتی کاربران وابسته به امکانات بازیابی اطلاعات برای فرمول‌بندی می‌باشد. هر چه قابلیت‌ها و امکانات بهتری فراهم آورده شود، کاربر راحت‌تر و آسان‌تر می‌تواند به بیان نیازهای خود بپردازد. امکانات لازم برای فرمول‌بندی جستجو در یک سیستم بهینه شامل امکان استفاده از عملگرهای بولی، عملگرهای مجاورت، جستجوی املائی، کوتاه‌سازی واژگان، تعیین محدوده در داده‌های کمی، گسترش مفاهیم و اصطلاحات،

استفاده از ضریب وزنی، سیاهه واژه‌ها می‌باشد. بعد از انجام جستجو، نمره‌گذاری مجموعه‌های بولی و اندازه‌گیری مشابهت، در بازیابی بهتر اطلاعات کاربران بسیار مفید است.

- ۴) نمایش و بازیابی رکوردها
بازخورد خوب سیستم در این قسمت نقش مهمی در هدایت کاربر برای رسیدن به اطلاعات مورد نظرش دارد. در این قسمت سیستم باید بتواند به سؤالات زیر پاسخ دهد:
- چه رکوردهایی با نیاز اطلاعاتی کاربر مطابقت دارند؟
 - آیا اطلاعاتی وجود دارد تا بتوان واژه‌هایی را که حوزه‌های موضوعی وسیعی را پوشش داده‌اند تشخیص داد؟
 - آیا اطلاعاتی وجود دارد که بتوان واژه‌هایی را که موجب اخص‌تر شدن دامنه جستجو می‌شوند، تشخیص داد؟
 - آیا مجموعه رکوردهای بازیابی شده با نیازهای اطلاعاتی کاربر مطابقت دارد؟
 - آیا اطلاعاتی وجود دارد که از طریق آن کاربر بتواند به تصحیح منطق بولی فرمول جستجو بپردازد؟

۵) سفارش مدرک

۶) درخواست برای اطلاعاتی درباره سیستم بازیابی اطلاعات

- ۷) برقراری پارامترهای نمایشی و ارتباطی
در مجموعه بازخورد سیستم باید چهار مقوله به عنوان مهم‌ترین اهداف، مد نظر طراحان سیستم‌های بازیابی اطلاعات قرار گیرند:
- ◇ بهینه‌سازی انتخاب واژگان جستجو توسط کاربر
 - ◇ بهینه‌سازی فرمول جستجوی کاربر
 - ◇ بهینه‌سازی و تنظیم تعداد رکوردهای بازیابی شده
 - ◇ بهینه‌سازی ضریب دقت و بازیابی یا بهینه‌سازی کل بازیابی اطلاعات
- در نهایت باید این نکته را خاطرنشان کرد که جستجوگران باتجربه و متخصصان بازیابی این نکته را می‌دانند که انجام بازیابی مستلزم صرف وقت و هزینه است و در این مسیر باید با اصلاح و بازنویسی و تکرار چرخه جستجو به نتایج دلخواه دست یافت.

۲. نمایش اطلاعات

امروزه در طراحی اغلب نرم‌افزارهای حرفه‌ای، «نمایش اطلاعات» (۵) از جایگاه مهمی برخوردار است. امکانات نرم‌افزاری و سخت‌افزاری موجود، تحول مهمی در نمایش اطلاعات به وجود آورده‌اند. نمایش اطلاعات راه‌های آسان‌تر و موثرتری را برای تعامل با سیستم‌های رایانه‌ای، بخصوص برای جستجوگران سیستم‌های اطلاعاتی فراهم آورده است. تعریف علمی نمایش گرافیکی اطلاعات عبارت است از «محاسبه و انتقال علائم به اشکال هندسی به صورتی قابل درک و مشاهده توسط انسان، به منظور فهم و کشف روابط پنهان موجود بین عناصر مختلف داده‌ها» (رز، ۱۹۹۶). نمایش گرافیکی اطلاعات به عنوان یک شیوه علمی و زیرشاخه‌ای از میحث «تعامل انسان و رایانه» و با استفاده از توانایی‌های گرافیکی رایانه‌ها، اهداف زیر را دنبال می‌کند:

- ۱- بالا بردن سرعت فهم و پردازش اطلاعات توسط انسان در طی فرآیند ادراک و کم کردن درگیری‌های ذهنی او ،
 - ۲- ایجاد و برقراری ارتباط بین اجزای مختلف اطلاعات ،
 - ۳- انجام عملیات‌های پیچیده با اعمالی بسیار ساده ،
- در نمایش گرافیکی اطلاعات جنبه‌های مختلفی مورد توجه قرار می‌گیرند، که مهم‌ترین آن‌ها عبارت‌اند از: جهت‌دار کردن اطلاعات نمایشی، برجسته کردن اطلاعات، رنگ‌ها، ابعاد و اندازه‌ها، پیکره‌بندی تصاویر، تکرار در یک فضا. در نمایش اطلاعات و ارائه فنون مختلف، همواره باید چهار عامل را مورد توجه قرار داد. این چهار عامل عبارت‌اند از:
- الف - فضا: در نمایش ساختارهای سلسله مراتبی، باید از روش‌های موثری برای نمایش اطلاعات استفاده گردد، زیرا با افزایش پیچیدگی، کار مشکل‌تر خواهد شد. به هر حال با هر حجمی از اطلاعات، باید بتوان به نحوی به بازیابی این ساختارها پرداخت که دید کلی و جزئی، با هم و به صورت همزمان، در اختیار کاربر قرار گیرند. مشکلاتی که همواره در این مورد وجود خواهند داشت عبارت‌اند از:
- ۱) محدودیت‌های موجود در توانایی ادراکی کاربران ،
 - ۲) محدودیت موجود در فضای نمایش رایانه‌ها .
- ب - زمان: در ارائه فنون و راهکارهای متفاوت برای نمایش اطلاعات، همیشه باید سخت‌افزار و زمان لازم برای بارگذاری و روزآمد کردن صفحات یا مناطق نمایشی را در نظر گرفت. گاهی اوقات با یک تغییر کوچک، باید کل منطقه نمایشی از نو ترسیم گردد. این خود نیازمند زمان است و بسته به توانایی‌های سخت‌افزاری، از یک سیستم به سیستم دیگر متفاوت می‌باشد.
- ج - ثبات: ثبات نیز از جنبه‌های مهم در نمایش اطلاعات می‌باشد که باید مورد توجه قرار گیرد. نبودن ثبات موجب بالا رفتن بار فکری و انحراف فکری کاربران می‌گردد.
- د - بازیابی: یکی از جنبه‌های مهم در نمایش اطلاعات، فراهم آوردن بازیابی بهینه اطلاعات می‌باشد. بازیابی بهینه اطلاعات نیازمند توجه به سه عامل فضا، زمان و ثبات می‌باشد.

۱- روش‌های متداول در نمایش اطلاعات

منظور از روش‌های متداول در نمایش اطلاعات، روش‌هایی هستند که به صورت مرسوم در بسیاری از سیستم‌های نرم‌افزاری موجود وجود دارند. این روش، سه فن عمده را در برمی‌گیرد که در ادامه به بررسی مختصر آن‌ها می‌پردازیم:

۱/۱- پیمایش و تنظیم اندازه (۶)

منظور از پیمایش، حرکت در سطح صفحات نمایشی اطلاعات به منظور بازبینی قسمت‌های مختلف صفحه می‌باشد. در این حالت گاهی برای نمایش بهتر اطلاعات، نیاز به تغییر اندازه صفحات اطلاعات خواهیم داشت.

۱/۲- روش چند پنجره‌ای

در این روش از چند پنجره به صورت همزمان برای نمایش اطلاعات استفاده می‌گردد. استفاده از چند پنجره با قابلیت‌های تنظیم و پیمایش، نتایج بهتری را به همراه خواهد داشت. هر پنجره دارای قابلیت تنظیم و پیمایش مخصوص به خود می‌باشد، بنابراین کاربر می‌تواند از چند پنجره برای مشاهده و مقایسه جنبه‌های مختلف یک سلسله مراتب استفاده کند.

۱/۳- راهبرد بازبینی نقشه‌ای (۷)

در این روش، از دو پنجره به صورت همزمان برای نمایش اطلاعات استفاده می‌شود (برد، ۱۹۹۰). در پنجره اول، کل سلسله مراتب در تصویری کلی با مقیاسی کوچک‌تر قابل مشاهده می‌باشد. همزمان در پنجره دوم، جزئیات سلسله مراتب در مقیاسی بزرگ‌تر مشاهده می‌شود.

۲- فنون تاکید بر محتوا (۸)

با افزایش تعداد و عمق گره‌ها در ساختارهای سلسله مراتبی، ضعف فنون متداول در مهار نمایشی کلی و جزئی اطلاعات و کاهش درگیری‌های ذهنی کاربر، بیش‌تر نمایان می‌شود. به دلایل فوق، تحقیقات مختلفی در این زمینه صورت گرفت. در این مسیر بتدریج راهبردهایی تدوین شدند که سعی در نمایش اطلاعات مبتنی بر محتوای اطلاعات داشتند. از بهترین فنون مورد استفاده در این زمینه، فنون تاکید بر محتوا می‌باشند. در این فنون با استفاده از روش‌های فشرده‌سازی، کلیات و جزئیات اطلاعات به صورت همزمان در اختیار کاربر قرار می‌گیرد. در این روش از تعدادی از توابع برای تعیین محل استقرار و اندازه هر یک از گره‌ها در کل ساختار استفاده می‌شود. و گسترش فضای اطلاعاتی از یک کانون مرکزی که به صورت پیش‌فرض در صفحه وجود دارد آغاز، متناسب با واکنش‌ها و تعاملات کاربر گسترش پیدا می‌کند. در ادامه به بررسی مختصر چهار فن مورد استفاده در این روش می‌پردازیم:

۲/۱- تورفگرهای هایپرولیک (۹)

این تورفگرها برای نمایش اطلاعات از قوانین علم هندسه سود می‌جویند. در ابتدا فضای اطلاعاتی بر روی یک سطح هایپرولیک قرار داده می‌شود و سپس بر روی یک منطقه دایره‌ای شکل ترسیم می‌گردد و نمایش داده می‌شود.

۲/۲- درخت‌های مخروطی (۱۰)

یکی از راه‌های نمایش دو بعدی و سه بعدی ساختارهای سلسله مراتبی، استفاده از درخت‌های مخروطی می‌باشد (کاربر، ۱۹۹۵).

از لحاظ مکانی، ریشه یا بالاترین نقطه در ساختار سلسله مراتبی در رأس مخروط قرار می‌گیرد و زیرشاخه‌های فرعی ساختار در فضای رأس پراکنده می‌شوند. هر کدام از این زیرشاخه‌ها به نوبه خود یک مخروط دیگر را تشکیل می‌دهند که زیرشاخه‌های دیگری در اطراف آن پراکنده می‌شوند و این تقسیم‌بندی به همین صورت ادامه پیدا می‌کند.

۲/۳- تنظیم هوشمند (۱۱)

این فن را می‌توان در ارتباطات موجود از نوع دو طرفه (دو بعدی) مورد استفاده قرار داد (بارترام، ۱۹۹۴). در این مدل، در ابتدا اطلاعات را به روش مشخصی خوشه‌بندی می‌کنند و سپس به هر خوشه یک نماد گرافیکی اختصاص داده می‌شود و نمادهای گرافیکی با نظم خاصی بر روی صفحه نمایش قرار می‌گیرند، در حالی که خطوط ارتباطی بین هر کدام بر روی صفحه نمایش قابل مشاهده می‌باشند.

۲/۴- عدسی جادویی (۱۲)

عدسی جادویی از جمله فنونی است که در آزمایشگاه «زیراکس پارک» (۱۳) گسترش یافته و منطقه خاصی از صفحه نمایش را با حالتی متفاوت از سایر نقاط نمایش می‌دهد (استون، ۱۹۹۴). اطلاعات اولیه به صورت کلی نمایش داده می‌شوند و کاربر می‌تواند عدسی را بر روی هر قسمت از این منطقه نمایشی اطلاعات قرار دهد؛ که در این صورت اطلاعات جزئی مربوط به آن منطقه خاص نمایش داده می‌شود. بنابراین با استفاده از این عدسی‌ها می‌توان به مطالعه و مشاهده جزئیات اطلاعات پرداخت و در همان حال، اطلاعات کلی در خارج از محدوده عدسی قابل مشاهده می‌باشند.

۲- روش «چه‌اویز» (۱۴)

روش «چه‌اویز» امکان مانور بهتر بر روی ساختارهای سلسله مراتبی بزرگ را فراهم می‌آورد. در این روش، کلیات اطلاعات در ساختارهای سلسله مراتبی از بین نمی‌رود و در همان زمان امکان دسترسی آسان و سریع به جزئیات نیز وجود دارد. در این مدل نیاز به خوشه‌بندی اطلاعات یا عملیات‌های گرافیکی و محاسباتی پیچیده برای نمایش اطلاعات نمی‌باشد.

بحث و نتیجه‌گیری

در این جا پنج سؤال مطرح شده در بخش سؤالات پژوهش به ترتیب مطرح می‌شوند و پاسخ آن‌ها براساس نتایج به دست آمده مورد بحث قرار می‌گیرند. بعد از آن، نتیجه‌گیری و پیشنهادها مطرح خواهند شد.

سؤال ۱. نحوه سازماندهی اطلاعات، ساخت فایل‌های شاخص و مقلوب و سایر فایل‌های کمکی در سیستم‌های بازیابی اطلاعات بهینه به چه صورت می‌باشد؟

امروزه با توجه به حجم انبوه اطلاعات و نگهداری اطلاعات تمام متن، ابزارها و شیوه‌های سنتی برای کنترل اطلاعات کافی نیستند. استفاده از روش‌های غیرخودکار در این زمینه غیراقتصادی است و با توجه به هوشمندتر شدن سیستم‌ها مقرون به صرفه نیست. هر چند که سازماندهی خودکار اطلاعات در سیستم‌های خارجی بازیابی اطلاعات

متداول است، اما در سیستم‌هایی که در داخل کشور طراحی شده‌اند، چیزی تحت عنوان سازماندهی اطلاعات وجود ندارد.

سازماندهی اطلاعات هسته مرکزی و ستون فقرات یک سیستم بازیابی اطلاعات را تشکیل می‌دهد. هر سیستم اطلاعاتی دارای یک مبنای خاص برای تجزیه و تحلیل اطلاعات می‌باشد، که سیستم براساس آن به تفسیر اطلاعات و مطابقت بین اقلام و درخواست‌های اطلاعاتی می‌پردازد. به لحاظ اهمیت سازماندهی اطلاعات، سیستم‌های بازیابی اطلاعات به دو دسته هوشمند و غیرهوشمند تقسیم می‌شوند.

هنگام پردازش متن‌هایی با زبان طبیعی، مسئله پیچیدگی زبان و بی‌قاعدگی‌هایی که در حوزه نحوی و معنایی وجود دارد خود را نشان خواهد داد که باید این مشکلات را مدنظر قرار داد. در این میان واژه‌نامه‌ها قادر نیستند به صورت کامل، ابهامات و

پیچیدگی‌های موجود در زبان را برطرف کنند؛ اما می‌توانند این بی‌قاعدگی‌ها را کاهش دهند. این واژه‌نامه‌ها شامل سیاهه بازدارنده، واژه‌نامه ریشه لغات و سیاهه پیوندها، واژه‌نامه عبارات، واژه‌نامه مفاهیم، برقراری روابط سلسله مراتبی بین مفاهیم، و ریشه‌یابی واژگان می‌شود.

به منظور سازماندهی واژه‌ها باید دامنه شمول واژه‌ها، نوع دسته‌بندی مفاهیم و مکان واژه‌ها را مدنظر قرار داد. پاسخگویی به مسائل فوق وابسته به سیاستگذاری سیستم در رابطه با ضرایب بازیابی و دقت می‌باشد. اگر بخواهیم ضریب بازیابی را بالا ببریم باید تمام واژه‌هایی که می‌توانند بنحوی در بازیابی مدارک موثر باشند، به واژه‌نامه اضافه شوند. هر چه واژه‌نامه عمومی‌تر باشد، ضریب بازیابی افزایش و ضریب دقت کاهش خواهد یافت. در مقابل اگر بخواهیم ضریب دقت بالاتری داشته باشیم و اقلام مرتبط‌تری مورد جستجو قرار گیرند، باید به سمت واژه‌نامه‌های تخصصی‌تر حرکت کنیم.

ساخت واژه‌نامه به صورت خودکار با دو روش تمام خودکار و نیمه‌خودکار قابل اجرا است، به دلیل این که عملی بودن روش تمام خودکار بیش‌تر از آن که به یک شاخه موضوعی خاص وابسته باشد به مجموعه مدارک مورد پردازش بستگی دارد؛ بنابراین در رویه‌های پردازشی تغییراتی اعمال می‌گردد و از قضاوت انسان‌ها استفاده می‌شود.

به منظور بالا بردن کارایی سیستم‌های بازیابی اطلاعات، از روش‌های محاسبه همبستگی و خوشه‌بندی اطلاعات نیز استفاده می‌گردد که محاسبه همبستگی به عنوان معیاری برای سنجش میزان مشابهت و همبستگی بین دو متن و خوشه‌بندی برای دسته‌بندی رکوردهای اطلاعاتی مشابه با هم، مورد استفاده قرار می‌گیرد.

سؤال ۲. چه امکانات و قابلیت‌هایی در سیستم‌های بازیابی اطلاعات بهینه لازم است تا کاربران بتوانند به بهترین شکل ممکن به بیان نیازهای اطلاعاتی خود بپردازند؟ به بیان دیگر، چه قابلیت‌ها و امکاناتی باید در سیستم وجود داشته باشد تا کاربران بتوانند یک راهبرد بازیابی صحیح و دقیق داشته باشند؟

فرمول‌بندی صحیح و دقیق نیازهای اطلاعاتی کاربران وابسته به امکانات بازیابی اطلاعات برای فرمول‌بندی می‌باشد. هر چه قابلیت‌ها و امکانات بهتری فراهم آورده شود، کاربر راحت‌تر و آسان‌تر می‌تواند به بیان نیازهای خود بپردازد. در ادامه به بررسی مختصر این امکانات می‌پردازیم:

۱- امکان استفاده از عملگرهای بولی: با استفاده از این کاربر عملگرها تصورات (AND, OR, NOT, XOR) می‌تواند بین و واژه‌هایی که برای بیان درخواست اطلاعاتی خود برگزیده است یک ارتباط منطقی برقرار کند.

۲- عملگرهای مجاورت: با ترکیب عملگرهای منطقی بولی و عملگرهای مجاورت می‌توانیم مجموعه‌های اطلاعاتی را مورد ارزیابی قرار دهیم

۳- جستجوی املائی: این روش هم در جستجوی واژه‌ها و هم در تجزیه و تحلیل ابتدایی مجموعه‌های اطلاعاتی به منظور ایجاد فایل‌های کمکی در سیستم، کاربرد دارد. به عنوان مثال، کاربر واژه Computer را برای جستجو وارد می‌کند و آنگاه واژه‌های Computer, Commpiter, Compute مورد جستجو قرار می‌گیرند. استفاده از این روش باعث افزایش ضریب فراخوانی و کاهش ضریب دقت می‌شود.

۴- کوتاه‌سازی واژگان: کوتاه‌سازی این امکان را فراهم می‌آورد که در هنگام جستجو، بدون در نظر گرفتن حرف یا حرف‌هایی از یک واژه یا یک واحد اطلاعاتی، به بازیابی بپردازیم. بدلیل ماهیت گوناگون اطلاعات و درخواست‌های اطلاعاتی، کوتاه‌سازی، بخصوص در سیستم‌هایی که ریشه‌یابی واژه‌ها (در مرحله تجزیه و تحلیل داده‌ها) و یکدست‌سازی اطلاعات بخوبی صورت نمی‌گیرد، اهمیت بسیار می‌یابد.

۵- تعیین محدوده در داده‌های کمی: بنابر ماهیت این نوع داده‌ها نیاز خواهد بود که علاوه بر یک مورد خاص، در محدوده‌های خاصی نیز جستجو صورت گیرد.

۶- گسترش مفاهیم و اصطلاحات: سیستم‌های بازیابی اطلاعات با استفاده از ابزارها و شیوه‌های مختلف به گسترش مفاهیم موجود در درخواست اطلاعاتی کاربر می‌پردازند و برای این منظور به استفاده از اصطلاحنامه‌ها و واژه‌نامه‌های موجود یا ایجاد آن‌ها بر پایه اطلاعات موجود در سیستم می‌پردازند.

۷- استفاده از ضریب وزنی: منظور از ضریب وزنی این است که کاربر به هر عضو دلخواه از فرمول جستجو، یک عدد صحیح که مشخص‌کننده اهمیت عضو در میان اعضا می‌باشد اختصاص می‌دهد و سیستم بازیابی اطلاعات با در نظر گرفتن این ضریب، بازیابی اطلاعات را انجام می‌دهد.

۸- سیاهه واژه‌ها: در سیاهه فوق بین هر جفت واژه، عملگر OR فرض می‌شود و هر کدام از آن‌ها دارای ضرایب وزنی یکسانی می‌باشند. در نهایت، رکوردها براساس بالاترین تعداد واژه‌های فوق که در رکوردها ظاهر می‌شوند، و همچنین براساس تعداد تکرار آن‌ها در رکوردها نمره‌گذاری می‌شوند و براساس همین نمره‌گذاری مرتب می‌شوند.

بعد از انجام جستجو، نمره‌گذاری مجموعه‌های بولی و اندازه‌گیری مشابهت می‌تواند در بازیابی بهتر اطلاعات کاربران بسیار مفید باشد:

۱- نمره‌گذاری مجموعه‌های بولی: در این روش نیز براساس تعداد تکرار اعضای فرمول جستجو در رکوردهای بازیابی شده، به هر کدام از رکوردها یک نمره اختصاص می‌یابد و در نهایت، رکوردها برحسب همین نمره در خروجی مرتب می‌شوند. این نمره نشان‌دهنده درصد ارتباط رکورد بازیابی شده حاضر، با درخواست اطلاعاتی کاربر می‌باشد.

۲. اندازه‌گیری مشابهت: دسترسی آسان کاربر به حداقل یک رکورد که بالاترین میزان ارتباط را با درخواست اطلاعاتی او داشته باشد، بسیار اهمیت دارد. اولین رکوردی که به این صورت توسط سیستم مورد بازبازی قرار گیرد، می‌تواند مبنای جستجوهای بعدی کار باشد.

سؤال ۳. بازخورد در سیستم‌های بازبازی اطلاعات به چه صورت می‌باشد؟ به بیان دیگر، سیستم بازبازی اطلاعات چگونه می‌تواند کاربران خود را به مرتب‌ترین اطلاعات، در رابطه با نیاز اطلاعاتیشان هدایت کند؟ بر مبنای چرخه دستورها و مطالعات انجام شده و با توجه به دستورهای انتخابی کاربر، می‌توان پیش‌بینی کرد که جستجوگران در چه مرحله‌ای از جستجو قرار دارند و هدف آن‌ها چیست. با گروه‌بندی و با توجه به ترتیب استفاده از آن‌ها و پارامترهای استفاده شده می‌توان به دسته‌بندی کاربران اطلاعاتی پرداخت و بازخورد متناسبی نشان داد. چرخه دستورها شامل مراحل زیر است: (۱) انتخاب یک بانک اطلاعاتی برای انجام جستجو، (۲) جستجو برای واژه‌های مورد نظر در بانک واژگان، (۳) ایجاد فرمول جستجو و انجام جستجو، (۴) نمایش و بازبینی رکوردها، (۵) سفارش مدارک، (۶) درخواست برای اطلاعاتی درباره سیستم بازبازی اطلاعات، (۷) برقراری پارامترهای نمایشی و ارتباطی با توجه به چرخه فوق و آنچه که در بخش دستورها به آن اشاره شد، بازخورد مناسب سیستم باید در چهار مقوله زیر برنامه‌ریزی و دنبال گردد. این چهار مقوله عبارت‌اند از:

۱. بهینه‌سازی انتخاب واژگان جستجو توسط کاربر،
۲. بهینه‌سازی فرمول جستجوی کاربر،
۳. بهینه‌سازی و تنظیم تعداد رکوردهای بازبازی شده،
۴. بهینه‌سازی ضریب دقت و بازبازی یا بهینه‌سازی کل بازبازی اطلاعات.

سؤال ۴. نحوه نمایش اطلاعات در خروجی‌های حاصل از نتایج جستجو در سیستم‌های بازبازی اطلاعات به چه گونه می‌تواند کاربر را در انتخاب اطلاعات مرتبط هدایت کند؟ به بیان دیگر، شیوه‌های بهینه نمایش اطلاعات در خروجی‌های حاصل از نتایج جستجو کدام‌اند؟

امروزه با بالا رفتن قدرت سیستم‌های سخت‌افزاری و نرم‌افزاری و راحت‌تر شدن برنامه‌نویسی برای ایجاد محیط‌های گرافیکی قوی، دیگر عذری برای عدم استفاده از شیوه‌های مختلف نمایش اطلاعات باقی نمی‌ماند. استفاده از جنبه‌های مختلف در نمایش گرافیکی اطلاعات مانند جهت‌دار کردن اطلاعات نمایشی، برجسته کردن اطلاعات، رنگ‌ها، ابعاد و اندازه‌ها، پیکره‌بندی تصاویر و تکرار در یک فضا به منظور نمایش بهینه را باید مورد توجه قرار داد. استفاده از شیوه‌های پیمایش و تنظیم اندازه، چند پنجره‌ای، راهبرد بازبینی نقشه‌های شیوه‌های تاکید بر محتوا، عدسی جادویی و روش «چه‌اویز»، تنظیم هوشمند، درخت‌های مخروطی و تورق‌گرهای هایپرولیک را نیز می‌توان به منظور نمایش بهینه اطلاعات مورد توجه قرار داد.

اگر رابطه در سیستم‌های نرم‌افزاری بر مبنای دو هدف تسریع در فرآیند ادراک و تسهیل در فرآیند تفکر طراحی شوند، قاعدتاً پیشرفت قابل ملاحظه‌ای در انتقال اطلاعات از رایانه‌ها به انسان صورت خواهد گرفت. حتماً شما هم این جمله معروف را شنیده‌اید که «یک تصویر برابر با هزار واژه می‌باشد».

راه‌های مختلفی برای به تصویر کشیدن اطلاعات در یک محیط گرافیکی وجود دارند، فهم دقیق فرآیند ادراک انسان، این بصیرت را برای طراحان سیستم‌ها ایجاد خواهد کرد که بدانند با کدام شیوه نمایش قادر خواهند بود فهم و انتقال اطلاعات را به حداکثر برسانند. در این جا باید این نکته را متذکر شد که برای این پرسش که بهترین روش برای به تصویر کشیدن اطلاعات کدام است، پاسخ واحدی وجود ندارد. پاسخ به این سؤال با توجه به اهداف نمایش اطلاعات و چهار متغیر فضا، زمان، ثبات، و بازبینی متفاوت می‌باشد. فنون متداول نمایش اطلاعات همانند پیمایش و تنظیم اندازه، روش چند پنجره‌ای، و راهبرد بازبینی نقشه‌ای، امکان نمایش ساختارهای اطلاعاتی کوچک را فراهم می‌آورند و یا بزرگ‌تر شدن ساختارها، بازبینی اطلاعات با استفاده از روش‌ها کارآمد نخواهد بود. در فنون تاکید بر محتوا مانند تورق‌گرهای هایپرولیک، درخت‌های مخروطی، تنظیم هوشمند و عدسی جادویی، امکان نمایش ساختارهای سلسله مراتبی بزرگ‌تر وجود دارد. توزیع مناسب ساختار در فضای نمایشی و ثبات، بازبینی مناسبی از اطلاعات را به وجود می‌آورد. اما نمایش اطلاعات در این فنون نیازمند محاسبات زیادی می‌باشد. در این صورت اگر سیستم از امکانات سخت‌افزاری پایینی برخوردار باشد، انجام محاسبات مدت زمان بیش‌تری را به خود اختصاص می‌دهد. در روش «چه‌اویز» امکان نمایش ساختارهای سلسله مراتبی بزرگ با در نظر گرفتن فضا، زمان و ثبات برای یک بازبینی خوب وجود دارد، اما مشکل ذاتی که در این روش وجود دارد این است که در هر لحظه فقط یک شاخه از اطلاعات قابل مشاهده می‌باشد.

در استفاده از این فنون و راهبردها همیشه یک خطر وجود دارد و آن این است که این استفاده به فهم مشاهده‌کننده، پیش‌زمینه‌های قبلی او و محتوای اطلاعات وابسته خواهد بود. ذهن انسان همیشه از آخرین اطلاعات برای ادراک اطلاعات جدید استفاده می‌کند. اگر مجموعه‌ای از اطلاعات به عنوان یک شیء نشان داده شوند، ذهن تمایل دارد که اطلاعات درون‌دای را همانند آن شکل تفسیر کند. تجارب قبلی ما بر قضاوت ما از تصاویر اثر خواهند گذاشت. این نکته را باید همیشه مد نظر قرار داد که: یک کاربر ممکن است یک تصویر را براساس تجارب قبلی خود تفسیر کند نه براساس آنچه طراح در نظر داشته است.

سؤال ۵. چگونه می‌توان یک سیستم بازبازی اطلاعات را مورد ارزیابی قرار داد؟ آنچه در هنگام ارزیابی سیستم‌های بازبازی اطلاعات در سطح دنیا مورد توجه قرار می‌گیرد، ضریب بازبازی و ضریب دقت می‌باشد. اما این زمانی صحیح است که سیستم‌ها از یک حداقل امکانات ممکن برخوردار باشند. در ادامه، عناصری که لازم است در یک سیستم بازبازی اطلاعات بهینه وجود داشته باشد در جدول ۱ خلاصه می‌شوند

و وضعیت نرم افزارهای خارجی و داخلی در رابطه با وجود یا نبود این عناصر نمایانده می شود.

جدول ۱. امکانات سیستمهای داخلی بازرایی اطلاعات

مقوله سازماندهی اطلاعات	زیرشاخه	نرم افزارهای ایرانی	نرم افزارهای خارجی
	سیاهه بازدارنده واژه نامه ریشه لغات و سیاهه پسوندها واژه نامه عبارات واژه نامه مفاهیم برقراری روابط سلسله مراتبی بین مفاهیم ریشه یابی واژگان امکان استفاده از روش های خودکار امکان استفاده از روش های نیمه خودکار محاسبه همبستگی خوشه بندی اطلاعات	وجود دارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد	وجود دارد وجود دارد وجود دارد وجود دارد وجود دارد وجود دارد وجود دارد وجود دارد
فرمول جستجو	امکان استفاده از عملگرهای بولی عملگرهای مجاورت جستجوی املائی کوته سازی تعیین محدوده در داده های کمی گسترش مفاهیم و اصطلاحات استفاده از ضریب وزنی سیاهه واژه ها مجموعه های بولی نمره گذاری شده اندازه گیری مشابهت	AND OR NOT XOR ADJ W N P S F ? * #	وجود دارد وجود دارد وجود دارد وجود ندارد وجود دارد وجود ندارد وجود دارد وجود ندارد وجود دارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد وجود ندارد
بازخورد سیستم	بهبود سازی انتخاب واژگان جستجو توسط کاربر بهبود سازی فرمول جستجوی کاربر بهبود سازی و تنظیم تعداد رکوردهای بازیابی شده بهبود سازی ضریب دقت و بازیابی با بهبود سازی کل بازیابی اطلاعات	وجود ندارد وجود ندارد وجود ندارد وجود ندارد	وجود دارد وجود دارد وجود دارد وجود دارد
نمایش اطلاعات	استفاده از جنبه های مختلف نمایش اطلاعات استفاده از شیوه های مختلف نمایش اطلاعات	وجود ندارد وجود ندارد	وجود دارد وجود دارد

1. Dialog
2. Oak
3. Dowquest
4. stop word list
5. information visualization
6. pan & zoom
7. map view strategy
8. focus+context
9. Hyperbolic Browser
10. Cone-Trees
11. intelligent zoom
12. Magic Lens
13. Xerox PARC
14. Cheops Apporach

منابع

1. Abraham. C.T. 1965. Techniques for Thesaurus Organization and Evaluation. Information Science 4(4).
2. Bartram, L. 1994. Contextual Assistance in User interfaces to Complex, Time-Critical Systems: The Intelligent Zoom. In: Graphics Interface. Banff, Albert: 216-224.
3. Beard, D.V. and Walker, J.Q. 1990. Navigating Techniques to Improve the Display of Large Two-Dimensional Spaces. Behavior and Information Technology. 9(6): 451-466.
4. Bonner, R.E. 1964. On Some Clustering Techniques. IBM Journal of Research 22-32.
5. Bookstein, A. 1983. Outline of a General Probabilistic Retrieval Model. Journal of Documentation. 39 (2): 63-72.
6. Boyce, B.R. and Kraft, D.H. 1985. Principles and Theories in Information Science. In: Annual Review of Information Science and Technology. Vol. 20. White Plains, Newyork: Knowledge Industry Publications.
7. Carrierre, J. and Kazman, R. 1995. Research Report Interacting With Huge Hierarchies: Beyond Cone Trees. Proceedings of Information Visualization. Allanta: GA.
8. Chapman, J.L. 1981. A State Transition Analysis of On-line Information-Seeking Behavior. Journal of the American Society for Information Science 35(2): 107-116.
9. Dennis, S.F. 1965. The Construction of a Thesaurus Automatically from a Sample of Text. Symposium on Statistical Association Methods for Mechanized Documentation. Natl. U.S. Misc. Publ. 369 (December).
10. Dovel, J.A. and Heald J.H. 1965. Project Lex Status. Proceedings of ADI Meeting. Santa Monica, Calif.
11. Lavery, F.1966. Organization et Consultation d'un Thesaurus. 1965 FID Congress, Spartan Books. Washington, D.C.
12. Lovins, J.B. 1968. Developing of Stemming Algorithm. Mechanical Translation and Computational Linguistics 11(1).
13. Penniman, W.D. 1975. A Stochastic Process Analysis of On-line Behavior. Proceeding of the American Society for Information Science 38th Annual Meeting. Washington D.C: ASIS.
14. Resiner, P.1965. Semantic, Diversity and a Growing Mac-Machine Thesaurus. Information Science 4(4).
15. Rose, R.1996. P1000 Science and technology and Strategy for Information Visulization. Version 2.
16. Salton, G.1964. A Document Retrival System For Mac-Machine Interaction. Proceeding 19th ACM Natl. Philadelphia, Pa.
17. Salton, G. 1965. Progress in automatic Information Retrieval. IEEE Spectrum 2(8).
18. Salton, G. 1966. Data Manipulation and Programming Problems in Automatic Information Retrieval. Commun . ACM 9(3).
19. Salton, G. and McGill, M. 1983. Introduction to Modern Information Retrieval . New York: McGraw-Hill.
20. Sparck-Jones, K. 1965. Experiments in Semantic Classification. Mech. Transl 8(4).
21. Stone, M.C. and Fishkin, K. 1994. The Moveable filter as a User Interface Tool. Proceedings of CHI CONFERENCE on Human Factors in Computing Systems. Boston: ACM.