

استراتژی جستجو

محمد رضا داورپناه

امروزه پایگاه‌های اطلاعاتی پیوسته یکی از مهمترین ابزارهای تکنولوژی موجود در زمینه اطلاع‌رسانی است. پیش از این محققان و دانش‌پژوهان برای کسب اطلاعات از مراجع چاپی بهره می‌جستند، لکن با ظهور خدمات کامپیوتری مرجع منابع و شیوه جستجوی اطلاعات نیز تغییر یافته و به دلیل سرعت، دقت، قابلیت انعطاف، توان ترکیب منطقی، روزآمدی، و دلایل دیگر تعداد استفاده کنندگان از جستجوهای پیوسته روبه افزایش است. اما نظر به هزینه زیاد برخورداری از پایگاه اطلاعاتی و تکنولوژی نوین به منظور تضمین استفاده از هزینه صرف شده باید نهایت بهره‌برداری از نظام به عمل آید. از آنجا که یکی از عوامل محاسبه هزینه جستجوهای پیوسته مدت زمان جستجو است، هرچه زمان جستجو افزایش یابد به تناسب هزینه نیز افزایش پیدا می‌کند و موجبات خروج بیش از پیش ارزش فراهم می‌گردد. لذا چنانچه مقتضیات یک جستجوی پیوسته به جستجوگران گوشزد و یک روند سریع منطقی جستجو از پیش طرح‌ریزی شود، در حداقل زمان بهترین برون‌داد بدست خواهد آمد.

جستجوی کامپیوتری

فرض کنید عنوان «اثر تکنولوژی جدید در انتخاب و فراهم‌آوری پیاپی‌ها» را در دست داشته باشیم. این عنوان ممکن است روی کارت منگنه، نوار کاغذی یا نوار مغناطیسی به همین صورت اما به شکل ماشین‌خوان ذخیره شده باشد. چنانچه برنامه‌ریزی چنان باشد که کامپیوتر نوار مربوط به چنین عنوانی را جستجو کند، ماشین کار جستجوی واژه‌ها یا ترکیب‌های واژه‌ها را بر اساس بررسی نوار از ابتدا تا انتها انجام می‌دهد. این چنین شیوه‌ای که در آن رکوردهای اطلاعات به صورت ترتیبی قرار گرفته‌اند و برای دسترسی به هر کدام می‌بایستی کلیه رکوردهای ماقبل آن گذرانده شود به عنوان جستجوی ترتیبی^۱ شناخته شده است، این شیوه در حقیقت یک پردازش خطی است. زیرا کامپیوتر بایستی تمام نوار یا نوارها و همه استنادهای پایگاه اطلاعاتی را به طور کامل جستجو کند تا برون‌داد را ارائه دهد.

به دلیل گندی عملیات جستجو و صرف وقت زیاد در روش دستیابی ترتیبی، شیوه‌های دستیابی دیگری به منظور ذخیره بخش‌های قابل جستجوی ارجاعات با استفاده از فایل‌های مقلوب، مشابه نمایه‌سازی‌ها، طراحی گردیده است. البته در اینجا

بعضی از سیستم‌ها به گونه‌ای طراحی شده‌اند که

جستجوی ارجاعات را به دوزبان مثلاً انگلیسی و

فرانسه انجام می‌دهند.





فرمولبندی جستجو

آنچه در ذیل به عنوان فرمولبندی جستجویم آید اصولی کلی و عمومی برای جستجو است نه مختص به سیستمی خاص یا موضوعی ویژه، زیرا شیوه اختصاصی جستجو به دلیل اینکه خدمات اطلاعاتی از نرم افزارهای متفاوتی استفاده می کنند از یکدیگر متمایز است. البته روند واقعی جستجو در یک پایگاه اطلاعاتی توسط کامپیوتر صورت می گیرد لکن استراتژی و فرمولبندی ای که در اینجا از آن سخن می رود، پاره ای کوشش هایی است که قبل از برقراری تماس با کامپیوتر به منظور تحلیل پرسش انجام می پذیرد.

۱. نوشتن موضوع مورد جستجو

پیش از اینکه از کامپیوتر برای جستجوی اطلاعات استفاده شود، ابتدا می بایستی موضوع مورد جستجو یا پرسش مطرح شده نوشته شود. این نوشتن باید به گونه ای باشد که کامپیوتر قادر به خواندن و تشخیص آن باشد، به عبارت دیگر نوشته به شکل ماشین خوان تهیه شود. استراتژی جستجو اساساً انعکاس شکل ماشین خوان به داخل ورودیهای کامپیوتر مثل نوار پانچ، کارت، نوار مغناطیسی یا دیسک است. در اغلب سیستمها روند جستجو عوامل انسانی زیادی را به منظور برقراری ارتباط بین استفاده کننده و کامپیوتر به خود اختصاص داده است. بعضی سیستمها نیز این اجازه را می دهند که استفاده کننده از پایگاه اطلاعاتی با استفاده از ترمینال پیوسته مستقیماً جستجو کند (البته در صورتیکه ترمینال مستقیماً به کامپیوتر مرکزی متصل باشد).

موضوع می بایستی در قالب یک جمله روشن به زبان پایگاه اطلاعاتی تهیه شود. بعضی از سیستمها به گونه ای طراحی شده اند که جستجوی ارجاعات را به دوزبان مثلاً انگلیسی و فرانسه انجام می دهند. اما آنچه مسلم است چنانچه سیستم قادر به ارائه اطلاعات تمام متن باشد، اطلاعات تمام متن را به همان زبانی که ذخیره شده عرضه می کند. زیرا به دلیل حجم عظیم اطلاعات (به عنوان مثال ذخیره سرویس اطلاعاتی «Dialog» بیش از ۲۰۰ میلیون رکورد است) و نیاز به سرمایه گذارهای کلان و ضعف سیستمهای مترجم تاکنون هیچیک از تهیه کنندگان خدمات اطلاعاتی به ترجمه متن اصلی موضوعات نپرداخته اند و فقط عناوین را با توجه به خط مشی پایگاه اطلاعاتی به یک یا دو زبان بین المللی عرضه می دارند.

فرضاً موضوع مورد علاقه همان موضوع «اثر تکنولوژی جدید در انتخاب و فراهم آوری پایندها» است. و با فرض اینکه پایگاه اطلاعاتی جستجوی ارجاعات را به زبان انگلیسی انجام می دهد، لذا ترجمه موضوع فوق به زبان انگلیسی چنین می شود:

Impact of new technology on serials acquisition



هرچه زمان جستجو افزایش یابد به تناسب هزینه نیز افزایش پیدا می کند.

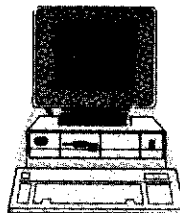
ارجاعات به جای نوار روی دیسکها جای می گیرند. شیوه جستجو در این حالت «جستجوی مستقیم یا تصادفی»^۲ نام گرفته است. زیرا کامپیوتر فقط از میان بخش کوچک مرتبط در پایگاه اطلاعاتی به جستجویم پردازد. در این شیوه هر رکورد دارای کد مخصوصی است که این کد به نشانی رکورد در روی دیسک تبدیل می گردد. بنابراین برای دسترسی به اطلاعات رکورد با استفاده از نشانی مستقیماً به سراغ آن می رود.

فرضاً برنامه ریزی می کنیم که کامپیوتر همه ارجاعات شامل واژه New technology را جستجو کند. کامپیوتر معمولاً به گونه ای طرح ریزی شده که ارجاعات حاوی واژه یا واژه های خواسته شده را بعد از جستجو نمایش دهد.

در این صورت امکان دارد صدها عنوان حاوی واژه مورد نظر یافته شود اما قطعاً بسیاری از آنها با آنچه در پی آن هستیم بی ارتباط خواهد بود. به منظور اینکه برونداد واقعاً مفید و مورد استفاده قرار گیرد، بایستی «دقت» زیادی اعمال شود و به گونه ای عناوین ناخواسته محدود گردند.

در این مورد خاص محدودیت را با به کارگیری واژه های Serials و Acquisitions اعمال می کنیم. این بار قطعاً عناوین بازیابی شده از تعداد قبلی بسیار کمتر خواهد بود و آنچه بازیابی می شود با موضوع مورد نظر ارتباط نزدیکی خواهد داشت. البته هنوز هم امکان دارد عناوینی بازیابی شود که مورد نظر ما نباشد و یا با موضوع ما ارتباطی نداشته باشد. اما یک نظام بازیابی خوب می بایستی «بازیافت» بالا و «دقت» زیاد داشته باشد و نباید عناوین مفید زیادی را حذف کند و بنا به قولی ریزش کاذب ایجاد کند. البته در عمل دستیابی به ۱۰۰ درصد دقت و ۱۰۰ درصد بازیافت امکان پذیر نخواهد بود و در هر بازیابی عملاً تعدادی عناوین ناخواسته بدست خواهد آمد و تعدادی ارجاعات مفید نیز از دست خواهد رفت.

این نکته را نیز نباید از خاطر دور داشت که نتیجه جستجو همیشه سازش بین «دقت» زیاد و «بازیافت» بالا است. امکان دارد جستجوگر به بازیافت بیشتر از دقت اهمیت دهد یا برعکس.



اصطلاحنامه به عنوان بهترین ابزار گزینش می تواند در انتخاب اصطلاحات موضوعی و واژه های مترادف مورد استفاده قرار گیرد.

نوع مدرک استفاده می کنند. با فرض اینکه درباره موضوع مورد نظر مطالبی به زبان انگلیسی می خواهیم، بنابراین جستجو را به زبان انگلیسی محدود می کنیم.

(a)	(b)	(c)	(d)
New technology	Serials	Acquisition	English language
Computer	Periodicals	Selection	
Automation	Journals	Subscription	
Computer program		Collection development	
Electronic processing			

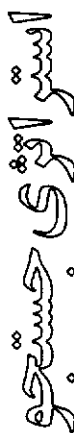
۵. بریدن یا باز گذاشتن انتهای کلید واژه ها

از آنجایی که کامپیوتر کلیدواژه های درون داد را حرف به حرف مقایسه می کند، لذا اصولاً از این شیوه برای مد نظر قرار دادن صورتهای مفرد و جمع واژه ها یا واژه هایی که در بخش پایانی متفاوت هستند، استفاده می شود. ناگفته نماند که بنا به امکانات سیستم عمل بریدن در ابتدا، انتها و یا حتی در میانه واژه ها به کار گرفته می شود. علائمی که بدین منظور مورد استفاده قرار می گیرند در سیستمهای مختلف متفاوت است. بعضی از این علائم عبارتند از: ستاره (*)، علامت سؤال (?)، یا علامت دلار (\$) .

به عنوان مثال می خواهیم کل واژه هایی که با کلمه (Computer) شروع شده اند مورد جستجو قرار گرفته و بازیابی شوند. بدین جهت جستجوگر کلید واژه را با توجه به علامت به کار گرفته شده در سیستم، فرضاً ستاره، بدین صورت می نویسد *Computer. چنانچه بدین شکل وارد کامپیوتر گردد، علاوه بر واژه Computer واژه هایی همانند Computers و Computerised نیز مورد جستجو و بازیابی قرار می گیرند.

۶. اعم و اخص نمودن کلید واژه ها

به عبارتی مفاهیم نزدیک به یکدیگر را با هم دسته بندی می کنیم. و در حقیقت اولویت بندی می کنیم. این امر بدان لحاظ است که چنانچه در پایگاه اطلاعاتی درباره کلیدواژه های اخص عناوین اندکی یافته شد یا اصلاً عنوانی بازیابی نگردید واژه های



در نوشتن کلمات دقت لازم می بایست صورت پذیرد؛ زیرا جا انداختن یک حرف یا جابجا نوشتن حروف و یا اشتباه نوشتن یک حرف باعث عدم بازیابی یا بازیابی غلط می گردد.

۲. تجزیه و تحلیل موضوع مورد جستجویه مفاهیم جداگانه

منظور این است که مفاهیم اساسی مورد نظر و کلیدواژه ها را مشخص کنیم. این امر بدان دلیل انجام می پذیرد که در یک جستجو، کامپیوتر ابتدا تمام ارجاعات هریک از مفاهیم را جداگانه پیدا سپس آنها را براساس فرمول داده شده ترکیب می کند. مفاهیم اساسی در این موضوع خاص عبارتند از:

(a)	(b)	(c)
New technology	Serials	Acquisition

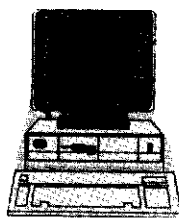
۳. تقسیم کلیدواژه ها و مفاهیم اساسی به پارامترهای مختلف

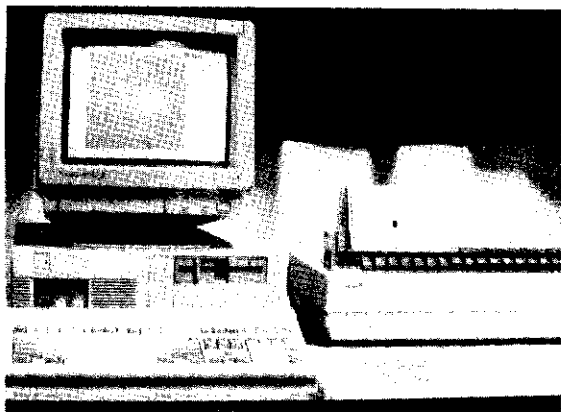
درواقع اصطلاحات موضوعی و واژه های در ارتباط با هریک از مفاهیم را انتخاب می نمایم. بدین منظور اصطلاحنامه به عنوان بهترین ابزار گزینش می تواند در انتخاب اصطلاحات موضوعی و واژه های مترادف مورد استفاده قرار گیرد. این اقدام خصوصاً در مورد سیستمهایی که تراوس ندارند و مهار واژگانی در آنها اعمال نمی شود بسیار مؤثر واقع می شود؛ زیرا احتمال دارد مفاهیم مورد نظر تحت عنوان واژه هایی دیگر ذخیره شده باشند. واژه های مترادف هریک از مفاهیم به صورت زیر خواهد بود:

(a)	(b)	(c)
New technology	Serials	Acquisitions
Computer	Periodicals	Selection
Automation	Journals	Subscription
Computer program		Collection development
Electronic processing		

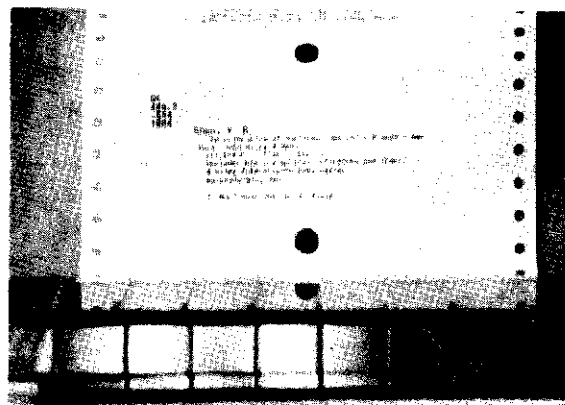
۴. اعمال محدودیت بر موضوع

بر اساس امکانات سیستم در صورت لزوم مشخصه های محدودکننده را تعیین می کنیم. مثلاً محقق به زبان چینی یا ژاپنی آشنایی ندارد و ترجمه مطالب نیز برای وی گران تمام می شود، لذا منطقی است که حوزه جستجو را به زبانی که آشنایی دارد محدود سازد؛ یا اینکه اصولاً اطلاعات و عناوین قدیمی به کار او نمی آید و به دنبال عناوین جدید است، لذا شایسته است از نظر زمانی اعمال محدودیت کند. البته سیستمهای مختلف به دلیل رقابت، جلب مشتری، یا دلایل دیگر امکانات متفاوتی تدارک می بینند، ولی عموماً از محدودگرهایی همانند زبان، سال انتشار، چکیده و





استراتژی جستجو



لازم به ذکر است که این مرحله می‌تواند در هنگام جستجو کامل‌تر گردد و یا احتمالاً تغییر یابد. زیرا به کارگیری یا ترکیب کلیدواژه‌ها یا مفاهیمی که در پایگاه اطلاعاتی دارای تعداد رکوردهای زیادی هستند نتیجه مطلوب‌تری بدست خواهد داد.

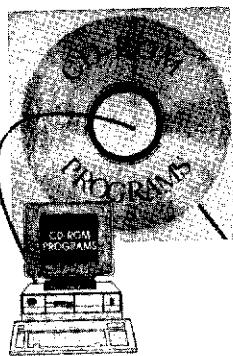
عامتراً مرحله به مرحله بکار بندیم. به عنوان مثال:

P _۱	P _۲	P _۳
New technology	Electronic processing	Serials
Computer *	Computer program	Periodical
Automation		Journal

P _۴	P _۵	P _۶
Acquisition	Subscription	English language
Selection	Collection development	

۷. این مرحله که در حقیقت آخرین مرحله برنامه‌ریزی قبل از جستجو است، فرمولبندی و ترکیب مفاهیم یا کلیدواژه‌ها براساس اهمیت و ارتباط با استفاده از این عملکردهای منطقی است: AND (ضرب منطقی است و جستجو را محدود و انحصار می‌سازد و اشتراک مفاهیم را می‌رساند)، OR (یا جمع منطقی موجب گسترش جستجو گردیده و حوزه مفاهیم را وسعت می‌بخشد) و NOT (تفریق منطقی است که برای مستثنی کردن مفاهیم مورد استفاده قرار می‌گیرد) باتوجه به آنچه گذشت به عنوان نمونه می‌توانیم موضوع مورد نظر را به شکل زیر فرمول‌بندی کنیم:

مرحله اول — P_۱ AND P_۲ AND P_۴ AND P_۶
 (New technology OR Computer*) AND Serial*
 AND acquisition AND LA = English
 مرحله دوم — P_۲ AND P_۳ AND (P_۴ OR P_۵) AND P_۶
 Electronic processing AND Serial*
 AND (Acquisition OR subscription) AND
 LA = English



یادداشت‌ها:

۱. مهرداد، جعفر؛ «بازیابی کامپیوتری اطلاعات»، مجموعه مقالات اولین سمینار کاربردی مراکز و سازمانهای پژوهش فرهنگی و اطلاع‌رسانی، تهران، سازمان مدارک فرهنگی انقلاب اسلامی، ۱۳۷۰.
۲. سینیایی، علی؛ تحریرات کلاسی، ۱۳۷۱.
۳. بخشی، نطف‌علی؛ آشنایی با کامپیوتر و داده‌پردازی، تهران، جهاد دانشگاهی، ۱۳۷۰.

4. Lancaster, F.W., *Information retrieval systems: characteristics testing and evaluation*.
5. Soergel, *Organizing information; principle of data base and retrieval systems*, Academic press, 1985.
6. *Searching DIALOG: the complete guide*. Palo Alto, CA: DIALOG information services, inc, 1989.