



بررسی اجمالی تاریخچه نظام‌های بازیابی اطلاعات^۱

علی گزنی^۲

چکیده: نظام‌های بازیابی اطلاعات (نرم‌افزار) عبارتند از: نظام‌هایی که به منظور بازیابی و پردازش بر روی داده‌های ساختار نیافته طراحی شده و به لحاظ نوع سازماندهی، ساختار پایگاه‌ها، راهبرد بازیابی، گروه‌های خدمات گیرنده و فرآیندی که در طی آن به درخواست‌های اطلاعاتی کاربران پاسخ می‌دهند منحصر به فرد می‌باشند. در مقاله حاضر به بررسی اجمالی تاریخچه نظام‌های بازیابی اطلاعات پرداخته شده است و سیر تحولات این نظام‌ها در طول دهه‌های مختلف مورد بررسی قرار گرفته است.

کلیدواژه‌ها: نظام‌های بازیابی اطلاعات، تاریخچه

مقدمه

در چند دهه اخیر بازیابی رایانه‌ای مورد توجه خاصی قرار گرفته است و علت آن را می‌توان در چهار عامل تشریح کرد:
۱- افزایش اهمیت زمان: مدت زمانی که در طی آن اطلاعات مورد نیاز جمع‌آوری شده و سپس برپایه آن تصمیم‌گیری و کنترل صورت می‌گیرد، بسیار حائز اهمیت می‌باشد. این اهمیت با توجه به تشدید رقابت‌ها، گسترش فعالیت‌های بین‌المللی و تغییر در عقاید عمومی می‌تواند موجب بهبود وضعیت اقتصادی، اجتماعی، نظامی، و سیاسی گردد.

۲- تغییر در کمیت و کیفیت اطلاعات: افزایش فزاینده اطلاعات مسائل زیر را بدنبال داشته است:

الف - پردازش بر حجم انبوه اطلاعات از لحاظ اقتصادی مقرون به صرفه نمی‌باشد.

ب - ابزار و شیوه‌های سنتی مورد استفاده در کتابخانه‌ها برای کنترل اطلاعات کافی نیستند.

۳- تغییر در ماهیت نیازهای اطلاعاتی: تخصصی شدن شاخه‌های موجود در دانش بشر و حل مسائل اجتماعی نیاز به دامنه گسترده‌ای از اطلاعات دارد. در این خصوص استفاده از جزء جزء اطلاعات موجود در منابع مختلف لازم به نظر می‌رسد.

۴- تغییر در اهمیت منابع اطلاعاتی: افزایش سریع

سازمان‌های جهانی، صنعتی، تربیتی و سیاسی موجب شد تا نقش اطلاعات در انجام تصمیم‌گیری‌ها آشکار گردد و در همین زمان لزوم دستیابی به منابع اطلاعاتی بدون در نظر گرفتن فاصله جغرافیایی آنها اهمیت پیدا کرد.

این عوامل تغییر و توسعه در نظام‌های بازیابی اطلاعات را بدنبال داشت و موجب مطرح شدن بحث‌هایی از جمله لزوم تغییر در ابزارهای اطلاعاتی، ایجاد و استفاده از نظام‌های ارتباطی جدید، تغییر معانی اطلاعات و اشاعه اطلاعات گردید.

بازیابی اطلاعات

لنکستر^۳ (۱۹۷۹) بازیابی اطلاعات را عبارت از: فرآیند

۱. برگرفته از: علی گزنی، "طراحی سیستم‌های بازیابی اطلاعات

بهینه در نرم‌افزارهای کتابخانه‌ای و اطلاع‌رسانی" به راهنمایی

دکتر عبدالحسین فرج‌بیلو، پایان‌نامه کارشناسی ارشد کتابداری و

اطلاع‌رسانی، دانشگاه شهید چمران اهواز، ۱۳۷۹.

۲. کارشناس ارشد کتابداری و اطلاع‌رسانی در کتابخانه منطقه‌ای علوم و

تکنولوژی شیراز

۱. Lancaster



فیزیکی اطلاعات ذخیره شده)

۳) راهبرد بازیابی

۴) فرآیند بازیابی

۵) عملکرد بازیابی

فیدل^۲ و سُرگیل^۳ (۱۹۸۳، صص ۱۶۲ - ۱۸۰) پس از انجام یک سری تحقیقات مشخصه‌های اصلی پایگاه‌های اطلاعاتی و بخش جستجو در نظام‌های بازیابی اطلاعات را در اقلام زیر ارائه می‌دهند:

الف. پایگاه اطلاعاتی

۱) گروه خدمات گیرنده

۲) زمینه موضوعی تحت پوشش

۳) خدماتی که به وسیله کارگزار در پایگاه فراهم‌آور می‌شود

۴) ساختار و نحوه سازماندهی فایل‌های شاخص

۵) هزینه و میزان پیچیدگی پایگاه

ب. بخش جستجو

۱) فیلدهایی که برای انجام جستجو تعیین شده‌اند

۲) امکانات نظام جستجو

۳) نحوه نمایش اطلاعات

۴) پیچیدگی و هزینه نظام جستجو

۵) مکانیزم شارژ شده پایگاه

چری^۴ (۱۹۸۳) نظام‌های بازیابی اطلاعات را از نظام‌های پرسش و پاسخ^۵ جدا دانسته و چنین اظهار می‌دارد که هدف در این سیستم‌ها یافتن مدارک مرتبط با نیازهای اطلاعاتی کاربران می‌باشد نه جواب دادن به پرسش‌ها. در همین رابطه مارون^۶ (۱۹۸۴، صص ۱۴۵ - ۱۷۶) می‌نویسد که سیستم‌های بازیابی اطلاعات بنابر ماهیت ذاتیشان، دارای حوزه و دامنه نامعینی می‌باشند و با صرف‌نظر کردن از بعضی وجوه تمایز

جستجو در میان مجموعه‌ای از مدارک با هدف تعیین آن دسته از مدارک که در حیطه موضوعی درخواست شده باشند می‌داند. چهار عنصر اصلی در این تعریف قابل مشاهده می‌باشد:

۱) نیاز اطلاعاتی که نقش محرک در بازیابی را ایفا می‌کند.

۲) هدف از بازیابی اطلاعات که همان یافتن اطلاعات مورد نیاز می‌باشد.

۳) جستجو در میان مجموعه‌ای از اطلاعات ذخیره شده صورت می‌پذیرد، رکوردهای اطلاعاتی با استفاده از اطلاعات تولید شده توسط پدیدآورندگان ایجاد می‌گردد و سپس در بانک‌های اطلاعاتی سازماندهی می‌شوند. بدون آنکه بدانند در آینده چه کسانی و در چه شرایطی آنها را مورد مطالعه قرار می‌دهند.

۴) فرآیند جستجو فرآیندی تعاملی می‌باشد، تعامل به این معنا که کاربر بر فرآیند جستجو کنترل داشته و در این مسیر دست به انتخاب می‌زند. پرسش کاربر ممکن است در طی این فرآیند با توجه به بازخوردهای حاصل از نتایج بازیابی یک سیر تکاملی را طی کند و انجام انتخاب مستلزم صرف هزینه، وقت و تلاش می‌باشد.

کنترل بر فرآیند بازیابی و انتخاب اطلاعات مرتبط دو جزء لازم در بازیابی اطلاعات می‌باشند. کپی کردن یک فایل بر روی دیسکت بازیابی اطلاعات نامیده نمی‌شود و یا نگاه کردن به یک برنامه اخبار تلویزیون بازیابی اطلاعات نمی‌باشد، زیرا در طول آن بیننده هیچ کنترلی بر آنچه که پخش می‌شود، ندارد. یک کتابخانه بهترین مثال برای نشان دادن بازیابی اطلاعات می‌باشد. جستجوگر کل مجموعه را برای بازیابی اطلاعات مورد مطالعه قراز نمی‌دهد، او اغلب مدارکی را انتخاب می‌کند که بالاترین درجه ارتباط را با نیازهای اطلاعاتی او دارند.

نظام‌های بازیابی اطلاعات

سات‌کلیف^۱ (۱۹۹۲، صص ۴۶۷ - ۴۹۰) نقاط و

مشخصه‌های اصلی یک نظام بازیابی اطلاعات را با توجه به اقلام زیر بر می‌شمارد:

۱) مشکل از مجموعه‌ای از مدارک

۲) نوع سازماندهی و ساختار اطلاعات (شکل منطقی و

1. Tague - Sutcliffe

2. Fidel

3. Sorgel

4. Gerrie

5. Question Answering System

6. Maron



می‌توان نظام‌های پرسش و پاسخ را جزئی از نظام‌های بازیابی اطلاعات به حساب آورد. الیس^۱ (۱۹۹۲، صص ۴۵ - ۶۴) نظام‌های بازیابی اطلاعات را دارای دو بعد می‌داند:

(۱) بعد سیستمی: ساختار داخلی نظام‌های بازیابی اطلاعات در این مقوله قرار می‌گیرند.

(۲) بعد درکی: کاربران و تأثیر مثبت و مفید آنها بر نتایج بازیابی در این مقوله قرار می‌گیرند.

در نهایت می‌توان تعریف زیر را برای نظام‌های بازیابی اطلاعات ارائه داد:

نظام‌هایی که به منظور بازیابی و پردازش بر روی داده‌های ساختار نیافته طراحی شده و به لحاظ نوع سازماندهی، ساختار پایگاه‌ها، راهبردهای بازیابی، گروه‌های خدمات‌گیرنده و فرآیندی که در طی آن به درخواست‌های اطلاعاتی کاربران پاسخ می‌دهند، منحصر بفرد می‌باشند.

خدمات بازیابی اطلاعات را می‌توان به دو دسته تقسیم کرد:

(۱) سرویس‌های از راه دور

(۲) سرویس‌های محلی

در سرویس‌های از راه دور خدمات‌دهی از طریق پایانه میزبان صورت می‌گیرد، کاربر با استفاده از یک ایستگاه کاری و یک خط ارتباطی به این پایانه متصل شده و ترمینال میزبان با تجزیه و تحلیل درخواست اطلاعاتی کاربر به بازیابی اطلاعات مرتبط می‌پردازد. از نقطه نظر کاربران تفاوت اندکی بین نظام‌های از راه دور و محلی وجود دارد، هر دوی این نظام‌ها در حکم یک رابط تعاملی بین کاربر و بانک‌های اطلاعاتی می‌باشند، تفاوت این دو نظام در هزینه جستجو در آنها می‌باشد.

نظام‌های مدیریت پایگاه داده‌ها

نظام‌هایی که برای پردازش بر روی اقلام داده‌ها طراحی شده‌اند را می‌توان به دو گروه کلی تقسیم کرد: "نظام‌های بازیابی اطلاعات" و "نظام‌های مدیریت پایگاه داده‌ها"^۲ مشکل درک تمایز این نظام‌ها از آنجا آغاز می‌گردد که هر دو با کنترل داده‌ها سروکار دارند، اما نظام‌های بازیابی اطلاعات دارای ویژگی‌ها و توابع خاصی برای کنترل و پردازش بر روی

"اطلاعات" هستند. در این مفهوم اطلاعات با دانش بشری سروکار دارد، که به نوبه خود از دامنه نامعین و حوزه بسیار وسیعی برخوردار است، کاربرد واژه نامعین به این دلیل است که، حداقل استاندارد یا کنترل بر پدیدآورندگان این اطلاعات وجود دارد. در مقابل نظام‌های مدیریت پایگاه داده‌ها برای پردازش بر روی داده‌های ساختار یافته بهینه‌سازی شده‌اند.

یک نویسنده سعی می‌کند تصورات و ایده‌هایش را مبتنی بر ذهنیات و افکارش بیان کند، به همین دلیل حداقل یکدستی در شیوه نگارش و واژه‌های بکار رفته در نوشته‌ها مشاهده می‌گردد، افراد مختلف جامعه انتزاع‌های مختلفی برای مفاهیم مشترک دارند، ماهیت مبهم و گسترده زبان نیز به این پیچیدگی بیشتر دامن می‌زند، بنابراین یک جستجوگر برای یافتن اطلاعات مورد نیاز خود بناچار می‌بایست با تمام واژه‌هایی که در آن حیطه بکار رفته‌اند آشنا باشد.

در مقابل داده‌های ساختار یافته دارای حوزه‌های معین و مشخصی بوده و به خوبی در بانک‌های جدول قابل نمایش می‌باشند. حوزه معنایی در نظام‌های مدیریت پایگاه داده‌ها مشخص و طبقه‌بندی شده می‌باشد، بنابراین هنگام درون‌داد اطلاعات به راحتی می‌توان اطلاعات را در فیلدهای اطلاعاتی مربوطه قرار داد. به عنوان مثال در این جداول هیچ تداخلی بین فیلد اطلاعاتی "نام کارمند" و "دستمزد کارمند" وجود ندارد.

درون‌داد و برون‌داد داده‌های ساختار یافته با قالب‌های معینی صورت می‌پذیرد، بالعکس در نظام‌های بازیابی اطلاعات هنگام جستجوی اطلاعات، در بسیاری از اوقات تمامی واژه‌های درون‌داد شده توسط کاربر مورد بازیابی قرار نمی‌گیرد. فقط با تغییر و تکرار متناوب فرمول جستجو، کاربر می‌تواند به اطلاعات خاص مورد نظرش دست یابد. به همین دلیل این فرآیند، جستجوی متناوب نامیده می‌شود.

در کنار مسائل فوق خیلی از اوقات کاربر نمی‌داند که چه می‌خواهد، نمی‌داند که از چه واژه‌هایی باید استفاده کند. روش کار با نظام بازیابی اطلاعات را نمی‌داند. حتی اگر موارد بالا حل شده باشند از واژه‌ها و نحوه سازماندهی آنها در درون

۱. Ellis

۲. Database Management System



نظام اطلاع ندارد، از این جهت اهمیت نظام‌های بازیابی کارآمد مشخص می‌شوند.

دهه ۱۹۵۰

در این زمان فعالیت‌های پایه و ابتدایی که زیربنای مطالعات و گسترش‌های بعدی نظام‌های بازیابی اطلاعات بودند، صورت گرفت. لوهن^۷ بر مبنای نظام نمایه کوئیک^۸ یک نظام بازیابی اطلاعات را طراحی کرده و مورد استفاده قرار داد. این روش جزء روش‌های نمایه‌سازی گردان است، در این روش رایانه عناوین مجلات را به صورت افقی به گونه‌ای حرکت می‌دهد که واژه‌های مشابه در مجاورت یکدیگر قرار می‌گیرند^۹.

به نظر می‌آید واژه "بازیابی اطلاعات"^{۱۰} به عنوان یک موضوع برای اولین بار توسط تئوریسینی بنام مؤثر^{۱۱} در سال ۱۹۵۱ بکار رفت. او معتقد بود که یک نظام بازیابی اطلاعات باید امکان نمایه‌سازی و جستجوی اطلاعات موجود در کتابخانه را فراهم آورد، وی با استفاده از روش‌های تئوریک و بررسی ساختار تلگراف توانست برای کدگذاری اطلاعات از راهی مؤثر و کم خطا استفاده کند.

در سال ۱۹۵۳ تویی^{۱۲} تصور موزونی از نمایه‌سازی ارائه کرد. وی به هر مدرک با توجه به محتوای اطلاعاتی آن تعدادی توصیفگر اختصاص داد به نحوی که با دانستن هر کدام از

1. Bush

۲. این مقاله در اینترنت با آدرس

<http://ebbs.english.vt.edu/btbl/AS-WE-MAY-THINK.html> قابل دسترسی

می‌باشد.

3. Speech Recognition

4. Optical Scanner

5. Optical Character Recognition

۶. واحدی برای اندازه‌گیری حافظه که تقریباً با یک ترینون بایت برابر است. یک ترابایت برابر ۱۰۰۰ کیگابایت می‌باشد.

7. H.P. Luhn

8. Keyword-in-Context (KWIC)

۹. این روش برای اولین بار توسط خود لوهن ابداع شد.

10. Information Retrieval

11. Mooers

12. Toobi

تاریخچه نظام‌های بازیابی اطلاعات

در ادامه به بررسی سیر تحول و تکامل نظام‌های بازیابی اطلاعات خواهیم پرداخت. در همین زمینه این تاریخچه در پنج دهه ۱۹۵۰، ۱۹۶۰، ۱۹۷۰، ۱۹۸۰، ۱۹۹۰ مورد بررسی قرار خواهد گرفت.

تعیین یک مبدأ زمانی دقیق برای گسترش نظام‌های بازیابی اطلاعات کار دشواری است. اما با این وجود گروه زیادی از صاحب‌نظران نظام‌های بازیابی اطلاعات، گسترش و عمومی شدن این نظام‌ها را از زمان ارائه مقاله بوش^۱ (۱۹۴۵، عرص ۱۰۱ - ۱۰۸) می‌دانند. بوش در مقاله^۲ خود به آینده‌نگری در مورد سخت‌افزارها و نرم‌افزارها در حیطه نظام‌های اطلاعاتی می‌پردازد. در حال حاضر و بعد از گذشت ۵۵ سال می‌توانیم پیش‌بینی‌های وی را مورد بررسی قرار دهیم. در یک نگاه می‌توان گفت: در حیطه سخت‌افزار فراتر و در حیطه نرم‌افزار عقب‌تر از پیش‌بینی‌های مورد نظر بوش قرار داریم. به عنوان مثال: در حیطه نرم‌افزار، بوش طراحی نظام‌های تشخیص صحبت^۳ را پیش‌بینی کرده بود، که هنوز به آن حد نرسیده‌ایم. همچنین پیش‌بینی‌های او در مورد قلم‌های نوری^۴ و تشخیص نوری حروف^۵ به صورت ۱۰۰٪ عملی نشده است. در حیطه سخت‌افزار، پیش‌بینی‌های بوش در دهه ۱۹۶۰ در دسترس بشر قرار گرفت، رایانه‌ها عملیات‌های ریاضی پیچیده‌ای را انجام دادند و داده‌ها برای استفاده‌های آتی در رایانه‌ها قابل ذخیره بود.

پیش‌بینی او در مورد منابع ذخیره رایانه‌ای هنوز هم تحقق نیافته است. او می‌نویسد اگر هر فرد در رایانه شخصی خود روزانه ۵۰۰۰ صفحه اطلاعاتی را ذخیره کند، ۲۰۰ سال برای تکمیل ظرفیت منبع ذخیره طول خواهد کشید. اگر هر صفحه حدود ۳۰ کیلوبایت فضا اشغال کند و ۲۵۰ روز از یکسال ذخیره‌سازی صورت گیرد به یک منبع ذخیره ۷/۵ ترابایتی^۶ نیاز داریم این در حالی است که حجم منابع ذخیره فعلی در حد کیگابایت است که در مقابل ترابایت نسبت کوچکی را به خود اختصاص می‌دهد.



کلیدواژه‌ها قادر باشیم مدرک را مورد ارزیابی قرار دهیم. در همین دهه اتحاد جماهیر شوروی سابق اولین ماهواره خود را به فضا پرتاب کرد. پرتاب موفق اسپوتنیک^۱ موجب بهبود در انتقال اطلاعات علمی در سطح جهان گردید.

دهه ۱۹۶۰

اولین نظام‌های اطلاعاتی در مقیاس بزرگ مانند نظام‌های تجاری بی - آر - اس^۲ و دایالوگ^۳ در این زمان تجربه شدند. مفهوم ارزیابی نظام‌های بازیابی اطلاعات و بدنبال آن مفاهیم ضریب فراخوانی^۴ و ضریب دقت^۵ مورد توجه قرار گرفتند. در همین زمان رشته بازیابی اطلاعات به صورت مستقل از علوم رایانه شروع به فعالیت کرد. نمایه‌های چایی موجود به صورت عملی در نظام‌های بازیابی اطلاعات درون‌داد شده و امکان جستجوی بولی بر روی آنها میسر شد. برخلاف محدودیت‌های نمایه‌های چایی، عملگرهای منطقی قابلیت‌های قابل ملاحظه‌ای را به همراه داشتند. به عنوان مثال: استفاده از عملگر AND در نمایه‌های چایی غیرممکن بود، اما نظام رایانه‌ای امکان بازیابی رخدادهای همزمان دو واژه و امکان پرداختن به جزئیات بیشتر در مورد یک مدرک را ممکن ساختند.

با توجه به بالا رفتن حجم اطلاعات، نمایه‌سازی مدارک غیراقتصادی به نظر می‌رسید. بنابراین ایده جستجوی متن‌های آزاد^۶ مطرح شد. در این روش در بازیابی اطلاعات کل واژه‌های موجود در مدارک مورد استفاده قرار می‌گرفتند، این ایده مورد استقبال قرار گرفت تا اینکه مشخص گردید نمایه‌سازی فقط انتخاب واژه‌ها برای جستجو نمی‌باشد، بلکه انتخاب واژه‌های صحیحی است که یا در رابطه با متن مهم باشند یا نشان دهنده موضوع متن مورد نظر باشند.

اولین کتاب تئوریک در زمینه نظام‌های بازیابی اطلاعات در سال ۱۹۶۱ توسط ویکری^۷ ارائه شد. در این کتاب نحوه ایجاد نمایه‌ها و توصیف موضوعی مدارک به صورت گسترده‌تری مورد بررسی قرار گرفت. انگیزه اصلی ویکری از تألیف این کتاب، رشد سریع نوشته‌ها و پیچیدگی روش‌های نمایه‌سازی مورد استفاده در کتابداری بود.

در سال ۱۹۶۵ ویرایش دوم این کتاب به چاپ رسید.

روش‌های قبلی که توسط ویکری ارائه شده بودند، باعث پیچیدگی نمایه‌ها می‌گردید و این روش‌ها برای مجلدات بزرگ کاربردی نداشتند. مسائل مهم و جدیدی از جمله: تحلیل کمی بازیابی‌های سودمند، هزینه‌ها و مقرون به صرفه بودن به بحث‌های مطرح شده در کتاب اول اضافه گردید.

ویرایش کتاب "تئوری نظام‌های بازیابی" ویکری که در سال ۱۹۶۵ ارائه شده بود توانست مبنای تئوریک دقیقی در این زمینه ارائه دهد. عدم وجود یک ساختار خاص، عدم ارائه قواعد زیربنایی قابل تعمیم، و عدم ارائه قواعد عمومی کاربردی همه از کاستی‌های برهان‌ها و قضایای مطرح شده در این کتاب بودند، بنابراین، این کتاب نمی‌توانست راهنمای طراحان و یا تحلیل‌گران نظام‌های بازیابی اطلاعات باشد. در این کتاب موضوعاتی مانند: نحوه استفاده از منطق ریاضی در سازماندهی توصیفگرها به منظور کمک به مراجعه‌کنندگان؛ منطق بولی؛ مطالعات زبان‌شناختی، انتخاب توصیفگرها، و راهبردهای بازیابی مورد بحث و بررسی قرار گرفت.

در این دهه موضوع کنترل واژگان و یکدست‌سازی واژه‌های مورد استفاده در نمایه‌سازی به منظور بازیابی بهتر اطلاعات نیز مطرح شد. همزمان با مطرح شدن ایده استفاده از متن‌های آزاد، مسائلی چند مورد توجه قرار گرفت از جمله: آیا استفاده از متن‌های آزاد کیفیت لازم و مورد نظر در جستجو را به همراه دارد؟ اگر بخواهیم از واژگان کنترل شده استفاده کنیم آیا راهی برای ترجمه و ایجاد آنها وجود دارد. در همین رابطه در سال ۱۹۶۶ به منظور ارزیابی نظام‌های بازیابی اطلاعات کلوردون^۸ ضریب فراخوانی (نسبت مدارک مرتبط بازیابی شده به کل مدارکی که احتمال بازیابی آنها وجود داشته) و ضریب دقت (نسبت مدارک مرتبط بازیابی شده به کل مدارک) را مطرح کرد.

- | | |
|--------------|----------------|
| 1. Sputnik | 2. BRS |
| 3. Dialog | 4. Recall |
| 5. Precision | 6. Free - Text |
| 7. Vickery | 8. Cleverdon |



تمام نظام‌های بازیابی اطلاعات در این دوره دارای نظام جستجوی ساده و محدودی بودند. برای مثال در سرویس^۸ OCLC چهار حرف اول نویسنده و چهار حرف اول عنوان برای جستجوی فهرست برگه مورد نظر بکار می‌رفت. در خدمات بهتر مانند دایالوگ جستجو به عملگرهای منطقی AND/OR/NOT محدود می‌شد. نمایه‌های مورد استفاده از میان واژه‌های موجود در متن و با کمترین کنترل استخراج می‌شدند.

در همین دهه نظام‌های تمام متن ظاهر شدند. لگزیس^۹ که در حال حاضر توسط رد - الزویر^{۱۱} کنترل می‌شود از همین نمونه می‌باشد. این پایگاه اطلاعاتی امکان دسترسی به متن‌های کامل حقوقی را فراهم می‌آورد. در دهه‌های ۱۹۷۰ و ۱۹۸۰ تحقیقات گسترده‌ای بر روی بانک‌های اطلاعاتی و سیستم‌های خودکار پیوسته صورت گرفت.

با گسترش دامنه تحقیقات مسئله نظام‌های بازیابی اطلاعات احتمالی به وسیله رایجس‌برگن^{۱۱} در دانشگاه گلاسکو^{۱۲} مطرح شد. در این نظام‌ها با استفاده از دفعات رخداد کلمات میزان مرتبط بودن یا نبودن مدارک، با بازیابی جاری مشخص می‌شد. الگوریتم‌های محاسباتی که بعداً تحت

سالتون^۱ در سال ۱۹۶۸ ایده بازخورد متناسب^۲ را مطرح کرد. او سعی در گسترش دامنه مفاهیم، اصطلاحات، و پرسش‌های مطرح شده^۳ توسط کاربر با افزودن واژه‌های متناسب داشت. در این دهه مسائل دیگری مانند پردازش زبان طبیعی از طریق هوش مصنوعی نیز مطرح شد. علت اصلی مطرح شدن این بحث این بود که بیان پرسش به زبان طبیعی نیاز به یادگیری نداشته و مانند زبان مکالمه روزمره افراد می‌باشد.

در اواخر این دهه جنبه‌های جدیدی در نظام‌های بازیابی اطلاعات مورد توجه قرار گرفت. از جمله می‌توانیم: تأکید بر کاربران به عنوان یک عامل مهم در بازیابی اطلاعات و توجه به بعد درکی و تعاملی کاربران در این نظام‌ها را نام ببریم. در طول این دهه و دهه‌های بعدی الگوی تطبیقی^۴ که الگوی متداول و رایج در نظام‌های بازیابی اطلاعات بود مورد نقد قرار گرفت.

تایلور^۵ (۱۹۶۸، صص ۱۷۸ - ۱۹۴) با اشاره به بخش مرجع و فرآیندی که در طی آن به نیازهای اطلاعاتی کاربران پاسخ داده می‌شود اظهار می‌دارد که گاهی اوقات، کاربر دقیقاً از نیاز اطلاعاتی خود آگاهی ندارد، تا اینکه این نیازها توسط کتابدار مرجع تشخیص داده شده، تشریح می‌گردد و در نهایت پاسخ‌های لازم ارائه می‌شود.

دهه ۱۹۷۰

در این دهه واژه‌پردازها و انجام حروف‌چینی توسط رایانه‌ها در ابعاد وسیع باعث گردید تا اطلاعات زیادی در فرم قابل خواندن به وسیله ماشین در دسترس قرار گیرند. در این دهه غالب خروجی‌های چاپی، حاصل درون‌دادهای رایانه‌ای بود. حجم انبوه اطلاعات موجب افزایش اهمیت نظام‌های بازیابی اطلاعات شد. نظام‌های بازیابی اطلاعات امکان کنترل و پردازش اطلاعات، بخصوص نمایه‌ها را فراهم آوردند.

از دیگر فن‌آوری‌هایی که نقش مهمی در تکامل نظام‌های بازیابی اطلاعات ایفا کرد، استفاده از نظام‌ها اشتراک زمانی^۶ بود. از این زمان به بعد پردازش بر روی پرسش به عنوان یک عمل پشت‌زمینه‌ای^۷ صورت نمی‌گرفت و با پردازش مستقیم بر روی پرسش کاربر، نتایج جستجو سریع‌اً اعلام می‌شد.

- | | |
|--|-----------------------|
| 1. Salton | 2. Relevance Feedback |
| 3. فرمول جستجوی که توسط کاربر برای بیان نیازهای اطلاعاتی او بیان می‌شود. | |
| 4. بلکین و کرافت (۱۹۸۷، صص ۱۰۹ - ۱۴۵) پردازش تطبیقی را بازیابی و منظم کردن مدارک بازیابی شده با استفاده از تکنیک‌های مختلفی که شامل جستجوی بولی، در نظر گرفتن احتمالات، مدل‌های برداری، کوتاه‌سازی واژه‌ها، شبکه‌های عصبی و... می‌باشد، معرفی می‌کنند. | |
| 5. Taylor | 6. Time Sharing |
| 7. این گونه پردازش‌ها به صورت ضمنی در نظام صورت می‌گیرد و در مقایسه با پردازش‌های مستقل از سرعت کمتری برخوردار است. | |
| 8. Ohio College Library Center | |
| 9. Lexis | 10. Reed - Elsevier |
| 11. Rijsbergen | 12. Glasgow |



عنوان ضریب وزنی مطرح شد، گسترش داده شده این مبحث بودند.

در دهه ۱۹۷۰ شاهد تحقیقات محققان هوش مصنوعی در زمینه بازیابی اطلاعات بودیم، پردازش بر روی پرسش‌های مطرح شده به زبان‌های طبیعی و جداسازی اطلاعات با استفاده از الگوهای مختلف از این نمونه می‌باشند.

سوانسون^۱ (۱۹۷۷، صص ۱۲۸ - ۱۴۸) نظام‌های بازیابی اطلاعات را به عنوان یک فرآیند آزمایش و خطا در نظر می‌گیرد. وی معتقد است که ضرایب بازیابی و دقت به تنهایی نمی‌توانند دلیل کیفیت و کارآمدی نظام باشند. وی می‌گوید در مراحل اولیه نیازهای اطلاعاتی بسیاری از کاربران خام می‌باشد، آیا واقعاً کاربر می‌داند که چه چیزی را می‌خواهد؟ آیا می‌داند آن را چگونه بدست آورد؟ آیا وقتی نتایج جستجو توسط نظام نمایش داده می‌شود کاربر قادر خواهد بود تا اطلاعات نامرتب را جدا کند؟ وی معتقد است که نظام باید از حداکثر انعطاف‌پذیری برای انجام آزمایش و خطاها برخوردار باشد به نحوی که کاربر بتواند به نتایج دلخواه خود دست یابد.

کوپر^۲ (۱۹۷۹، صص ۳۶۷ - ۳۷۵) اظهار می‌دارد مفید بودن یک نظام بازیابی اطلاعات فقط به ضرایب بازیابی و دقت وابسته نمی‌باشد بلکه عامل مهم در این زمینه مرتبط بودن اطلاعات بازیابی شده با نیازهای واقعی مورد درخواست کاربر می‌باشد.

هارتر^۳ در سال ۱۹۷۹ یک مدل برای سیستم‌های بازیابی اطلاعات و تعامل کاربران با این نظام‌ها ارائه می‌دهد. وی فنون مورد استفاده در نظام‌های بازیابی اطلاعات را به چهار دسته تقسیم می‌کند که عبارتند از: فنون نمایشی، فنون مربوط به ساختار فایل‌ها، فنون مربوط به فرمول بازیابی، و فنون مربوط به انتخاب واژگان.

دهه ۱۹۸۰

در طول دهه ۱۹۸۰ باز هم بر تعداد و اهمیت واژه‌پردازها افزوده شد و در مقابل قیمت منابع ذخیره رایانه‌ای کاهش یافت. هر روز بر حجم اطلاعات قابل خواندن توسط ماشین افزوده می‌شد. استفاده از نظام‌های بازیابی اطلاعات پیوسته در دو جنبه مختلف گسترش پیدا کرد. جنبه اول گسترش

نظام‌های تمام‌متن بود، تا قبل از این دهه در بازیابی اطلاعات فقط از نمایه‌ها و چکیده‌ها استفاده می‌شد. جنبه دوم گسترش نظام‌های پیوسته مورد استفاده غیرمتخصصین بود. مانند کتابخانه‌هایی که فهرست‌برگه‌های خود را با اوپک^۴ جایگزین کردند. در این دهه کارهای جدیدی بر روی تجزیه و تحلیل کلمات صورت گرفت. ظهور سی.دی.رام^۵ باعث شد تا نحوه توزیع اطلاعات متحول گردد.

در دهه ۱۹۸۰ استفاده از اوپک در بسیاری از کتابخانه‌ها مرسوم شد. گسترش استفاده از اوپک مرهون پایه‌گذاری این نظام‌ها توسط OCLC در دهه ۱۹۷۰ می‌باشد. تعداد نظام‌های بازیابی اطلاعات رو به افزایش بود. در همین ایام پروژه ریمارک^۶ (شامل مشخصات تمام کتاب‌های کتابخانه کنگره) به صورت پیوسته در دسترس جامعه استفاده‌کننده قرار گرفت. در انتهای این دهه نرم‌افزارهای اوپک توسط کارگزاران تجاری ارائه شدند (مانند: GEAC, NOTIS).

در این دهه استفاده از نظام‌های تمام‌متن رواج یافت. تعدادی از مجلات و روزنامه‌ها به صورت پیوسته و متنی مورد استفاده قرار گرفتند. برای مثال نظام پیوسته نشریه انجمن شیمی آمریکا^۷ (به صورت تمام‌متن) که در حال حاضر با نام شبکه علوم و تکنولوژی^۸ موجود می‌باشد، مورد استفاده قرار گرفت.

تلاش‌های جدیدی برای استفاده از تکنیک‌های جدید در بازیابی اطلاعات آغاز شد. انواع جدید مواد خواندنی به بازار عرضه شد. در سیستم‌های رایانه‌ای نیاز به انجام پردازش‌های مستقل در ایستگاه‌های کاری احساس شد. اینها همه از مجموعه عواملی بودند که نیاز به الگوریتم‌های جدید را به وجود آوردند. کرافت^۹ در دانشگاه ماساچوست و فاکس^{۱۰} در دانشگاه ویرجینیا در این زمینه شروع بکار کردند.

- | | |
|-------------|-----------|
| 1. Swanson | 2. Cooper |
| 3. Harter | 4. OPAC |
| 5. CD - ROM | 6. REMARC |
| 7. CJO | 8. STN |
| 9. Croft | 10. Fox |



کردن اطلاعات در خروجی‌های حاصل از نتایج جستجو را مورد بررسی قرار دادند آنها دریافتند که درصد خطاها با توجه به شیوه برجسته کردن اطلاعات در صفحات نمایشی متفاوت می‌باشد.

دهه ۱۹۹۰

هر روزه اطلاعات بیشتری از طریق پایگاه‌های مستنی پیوسته مورد استفاده قرار می‌گرفتند. حجم منابع ذخیره اطلاعات رایانه‌ای افزایش فوق‌العاده‌ای یافت، تصاویر به همراه اطلاعات ارائه شد و راه نمایش اطلاعات به گونه‌ای متفاوت تعقیب گردید.

تحول دیگر استفاده از پوششگرها^۱ برای درون‌داد اطلاعات بود. ناشران سعی کردند علاوه بر ارائه انتشارات خود با قالب متنی، قالب گرافیکی انتشارات را نیز، در دسترس جامعه قرار دهند. ارزان‌تر شدن قیمت سی.دی.رام‌ها و دیگر منابع ذخیره اطلاعات راه را برای استفاده از پوششگرها هموار ساخت و ناشران سعی کردند علاوه بر نظام‌های پیوسته تمام‌متن از طریق سی.دی.رام‌ها صفحات پوشش شده را، در اختیار همگان قرار دهند. اینها همه از عواملی بودند که بر اهمیت نظام‌های بازیابی اطلاعات می‌افزودند.

سالتون (۱۹۹۲) در زمینه الگوی تطبیقی اینگونه اظهار می‌دارد که:

(۱) فرمول جستجو ممکن است واقعاً بیان‌کننده نیازهای اطلاعاتی کاربران نباشند.

(۲) روش‌هایی که برای متمایز کردن اطلاعات مرتبط از نامرتب به کار می‌روند کامل نیستند.

(۳) مجموعه‌هایی که مورد آزمایش قرار می‌گیرند بسیار کوچک می‌باشند.

پری^۶ (۱۹۹۴، صص ۱۴۹ - ۱۵۹) عوامل مؤثر در

در این دهه تکنولوژی سی.دی.رام به صورت گسترده‌ای مورد استفاده قرار گرفت. در انتهای این دهه در اکثر کتابخانه‌ها حداقل یک درایو سی.دی.رام وجود داشت و از آن برای خدمات اطلاع‌رسانی استفاده می‌شد. از لحاظ اقتصادی تولید و توزیع اطلاعات از طریق سی.دی.رام‌ها مقرون به صرفه بود (این سی.دی.رام‌ها در حال حاضر در اندازه‌های ۶۵۰ مگابایتی و بالاتر، با قیمت بسیار پایینی تولید می‌شوند). این به آن معنی است که بانک‌های اطلاعاتی تمام متن بزرگ به راحتی قابل استفاده بودند.

در پروژه INSTANT بلکین، سیگر، و ورسینگ^۱ (۱۹۸۳، صص ۱۵۳ - ۱۶۷)، و ورسینگ و هنینگز^۲ (۱۹۸۴، صص ۷ - ۳۰) سعی کردند تا یک ساختار جدید برای نظام‌های بازیابی اطلاعات ارائه کنند. در این پروژه جنبه‌های مختلفی همانند: نیازهای اطلاعاتی کاربران، مشکلات کاربران حین تعامل با سیستم، مقاصد و پیش‌زمینه‌های قبلی کاربران مورد توجه خاصی قرار گرفت.

هازتر (۱۹۸۴، صص ۱۱۰ - ۱۱۷) فرآیند بازیابی اطلاعات را مشابه یک پرسشنامه می‌داند و بعضی از وجوه تشابه بین این دو را بدین صورت برمی‌شمارد:

(۱) در هر دو صورت مسئله‌ای وجود دارد که می‌بایست به آن پاسخ داده شود.

(۲) در هر دو مفاهیم و تصورات به وسیله واژه‌هایی بیان می‌گردند.

(۳) در هر دو عوامل هزینه تمام شده و دقت از اهمیت برخوردار است.

(۴) در هر دو از یک روش خاص برای تعیین مسیر جهت پاسخ به مسئله استفاده می‌گردد.

(۵) هر دو فرآیند از خصوصیات تکراری و تقریبی بودن برخوردار هستند.

بُرگمن^۳ (۱۹۸۶، صص ۴۷ - ۶۴) دریافت که مهارت فنی کاربران نقش عمده‌ای در عملکرد نظام ایفا می‌کند. کاربرانی که در تعامل با نظام می‌باشند از توانایی‌های ذهنی و هوشی متفاوتی برخوردار هستند. بنابراین تفاوت‌های فردی افراد در رفتار آنها با نظام بازیابی اطلاعات تأثیر خواهد گذاشت.

داویز و شاو^۴ (۱۹۸۹، صص ۳۲۵ - ۳۳۴) شیوه برجسته

1. Belkin & Seeger & Wersig

2. Hennings

3. Borgman

4. Davis & Shaw

5. Scanner

6. Barry



قضاوت بر روی اطلاعات مرتبط را اینگونه معرفی می‌کند:

(۱) محتوای اطلاعاتی

(۲) تجربه و زمینه‌های قبلی کاربر

(۳) قالب‌های ذهنی و پیش‌داوری‌های کاربر

(۴) اطلاعات و منابع دیگر در محیط اطلاعاتی مربوط

(۵) مأخذ مدرک

(۶) شکل مدرک

(۷) وضعیت کاربر

هارتر (۱۹۹۶، صص ۳۷ - ۴۹) علاوه بر اینکه یکی از معیارهای ارزیابی مهم در نظام‌های بازیابی اطلاعات را بازیابی و قضاوت در رابطه با اطلاعات مرتبط می‌داند، قضاوت کاربران مختلف را در این زمینه از گروهی به گروه دیگر متفاوت می‌داند.

تحول مهمی که در این دهه به وقوع پیوست ظهور اینترنت بود، تا حدود سال‌های ۱۹۹۴ فقط ۱۵٪ از افرادی که دارای رایانه شخصی بودند، به وسیله مودم با شبکه‌های پیوسته ارتباط برقرار می‌کردند. ولی امروزه روزانه میلیون‌ها انسان با شبکه اینترنت ارتباط برقرار می‌کنند. تا قبل از این تحول تعداد ناشران اطلاعاتی بسیار محدود بود ولی در حال حاضر افراد زیادی در سرتاسر جهان اطلاعات مربوط به خود را تولید و دسته‌بندی می‌کنند و از طریق صفحات خانگی^۱ خود آنها را در اختیار دیگران قرار می‌دهند.

نتیجه‌گیری

امروزه حجم انبوه اطلاعات به خصوص اطلاعات موجود بر روی شبکه جهانی وب موجب ظهور جنبه‌های جدیدی در بازیابی اطلاعات گردیده است. حجم این اطلاعات به صورت تصاعدی در حال گسترش می‌باشد و این در حالی است که تعداد افراد کم تجربه‌ای که از موتورهای کاوش وب استفاده می‌کنند نیز زیاد می‌باشند. شرکت Cyveillance اعلام کرد که تا تاریخ ۱۰ جولای سال ۲۰۰۰ معادل ۲/۱ میلیارد صفحه اطلاعاتی در وب شمارش شده است. هر روزه حدود ۷ میلیون صفحه به اطلاعات موجود بر روی وب افزوده می‌شود. موتورهای کاوش می‌بایست به ایندکس‌سازی ده‌ها و صدها میلیون صفحه اطلاعاتی بپردازند و روزانه به میلیون‌ها

نیاز اطلاعاتی پاسخ گویند. به عنوان مثال تعداد صفحات ایندکس شده توسط سایت [www. google.com](http://www.google.com) در حدود ۱/۳۲۶/۰۰۰/۰۰۰ اعلام شده است. جوابگویی به نیازهای اطلاعاتی کاربران با توجه به حجم انبوه اطلاعات با استفاده از فنون الگوهای تطبیقی موجب بالا رفتن ضریب فراخوانی و کاهش ضریب دقت و در نتیجه پایین آمدن کیفیت نتایج جستجو خواهد شد. در این مسیر نیاز به فنون مکمل و راهکارهای جدید در جهت افزایش ضریب دقت لازم به نظر می‌رسد و این خود مستلزم انجام تحقیق و پژوهش در این زمینه می‌باشد.



20. Taylor, R.S. Question Negotiation and Information Seeking in Libraries. **College & Research Libraries**, No. 29 (1968): 178-194.

21. Wersig, G., & R.D. Hennings. The Intellectual Architecture of Information System : A Broad Range Research Agenda. **IN Representation and Exchange of Knowledge as a Basis of Information Processes**, edited by H.J. Distachmann. International Research forum in Information Science, 5th. Amsterdam: North-Holland (1984): 7-30.

1. Barry, C.L. User - Defined Relevance Criteria: An Exploratory Study. **Journal of American Society for Information Science** 45, No. 3 (1994): 149-159.

2. Belkin, N.J. and Croft, W.B. Retrieval Techniques. **Annual Review of Information Science & Technology**, No.22 (1987): 109-145.

3. Belkin, N.J., T. Seeger, & G. Wersig. Distributed Expert Problem Treatment as a Model for Information System Analysis and Design. **Journal of Information Science**, No.5 (1983): 153-167.

4. Borgman, C.L. The User's Mental Model of an Information Retrieval System: An Experiment on a Prototype On-line Catalog. **International Journal of Man-machine Studies**, No. 24 (1986): 47-64.

5. Bush, V. As We May Think. **Atlantic Monthly** 176, No.1 (1945): 101-108.

6. Cleverdon, C. and Keen, M. **Factors Affecting the Performance of Indexing Systems**. Vol. 2. ASLIB Cranfield Project Research Project, 1966.

7. Cooper, W.S. The Paradoxical Role of Unexamined Documents in the Evaluation of Retrieval Effectiveness. **Information Processing & Management**, No. 12 (1979): 367-375.

8. Davis, C.H. and Shaw D. Comparison of Retrieval System Interfaces Using an Objective Measure of Screen Design Effectiveness. **Library & Information Science Research**, No. 11 (1989): 325-334.

9. Ellis, D. The Physical and Cognitive Paradigm in Information Retrieval Research. **Journal of Documentation** 48, No.1 (1992): 45-64.

10. Fidel, R. and Soergel D. Factors Affecting On-line Bibliographic Retrieval: A Conceptual Framework for Research. **Journal of the American Society for Information Science** 34, No. 3 (1983): 163-180.

11. Gerrie, B. **On-line Information System: Use and Operating Characteristics, Limitations, and Design Alternatives**. Arlington, VA: Information Resources Press, 1983.

12. Harter, S.P. Variations in Relevance Assessments and the Measurement of Retrieval Effectiveness. **Journal of the American Society for Information Science** 47, No. 1 (1996): 37-49.

13. Harter, S.P. Scientific Inquiry : A Model for On-line Searching. **Journal of the American Society for Information Science** 35, No.2 (1984): 110-117.

14. Lancaster, F. W. **Information Retrieval Systems: Characteristics, Testing, and Evaluation**. 2nd ed. New York: Wiley, 1979.

15. Maron, M.E. Probabilistic Retrieval Models. **Progress in Communications Sciences**, NO. 5 (1984): 145-176.

16. Salton, G. **Automatic Information Organization and Retrieval**. New York: McGraw - Hill, 1968.

17. Salton, G. The State of Retrieval System Evaluation. **Information Processing & Management**, No. 28 (1992): 441-449.

18. Swanson, D. R. Information Retrieval as a Trial and Error Process. **Library Quarterly** 47, No. 2 (1977): 128-148.

19. Tague-Sutcliffe, J. The Pragmatics of Information Retrieval Experimentation, Revisited. **Information Processing & Management**, No. 28 (1992): 467-490.

