

قلمروهای نو در بازیابی اطلاعات

• فرزاد دادرسی

کارشناس ارشد کتابداری و اطلاع‌رسانی

می‌تواند دید کلی درخصوص کتاب و موضوعات مورد بحث ارائه دهد و چشم‌اندازی را برای خوانندگان در ابتدای مطالعه کتاب ترسیم نماید.

نقص دوم، فقدان متن مبسوطی است درباره زندگی‌نامه نویسندگان و پیشینه علمی و فنی آنها و کسانی که در تهیه این کتاب حضور داشته‌اند، که در نسخه ترجمه شده حذف شده است. وجود این قسمت در نسخه ترجمه شده، می‌توانست دید کامل‌تری از محدوده تجربیات، توانایی‌ها و دانش مؤلفان کتاب به خواننده ارائه دهد.

نقص سوم، نبود یادداشت مترجم است. هرچند که در فهرست مطالب کتاب گزینه «یادداشت مترجم» دیده می‌شود، اما در متن کتاب اثری از آن نمی‌توان یافت. وجود آن می‌توانست در اثربخشی کتاب بسیار موثر باشد.

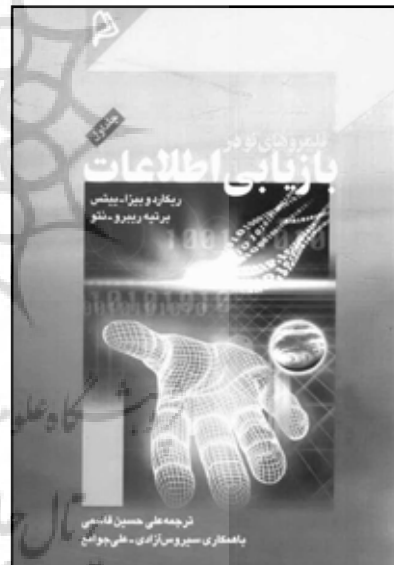
اعتبار کتاب

(الف) ناشر:

نکته شایان توجهی که یکی از ویژگی‌های اصلی و نقاط قوت کتاب شمرده می‌شود، اعتبار ناشر آن است. کتاب از سوی انتشارات ACM، یکی از معتبرترین انجمن‌ها و ناشران فنی در جهان در حوزه فناوری اطلاعات منتشر شده است و این به تنهایی کافی است که نه تنها اعتبار کتاب، بلکه اعتبار نویسندگان کتاب را روشن کند. اما متأسفانه در کتاب ترجمه شده، هیچ‌جا اثری از نام ناشر اصلی کتاب دیده نمی‌شود، در حالیکه در متن اصلی یک بند درباره انتشارات ACM و اعتبار آن توضیح داده شده که به‌طور یقین می‌توانست به اطلاعات خوانندگان در خصوص ناشران فنی بیفزاید.

(ب) بیوگرافی نویسندگان اصلی

ریکاردو بیزا بیس مدرک کارشناسی کامپیوتر خود را در سال ۱۹۸۳ از دانشگاه شیلی دریافت کرد. در سال ۱۹۸۵ مدرک کارشناسی ارشد خود را در مهندسی کامپیوتر و در سال ۱۹۸۹ دکترای خود را از دانشگاه واترلو کانادا اخذ کرد. او بین سال‌های ۱۹۹۲-۱۹۹۵ و ۱۹۹۷-۱۹۹۸، رئیس انجمن کامپیوتر شیلی بود.



■ بیزا - بیس، ریکاردو؛ ریبرو - نتو، بریته. قلمروهای نو در بازیابی اطلاعات. ترجمه علی حسین قاسمی با همکاری سیروس آزادی و علی جوامع. تهران: چاپار، چاپ اول، ۱۳۸۵. ۲۵۰ ص.

پیش از پرداختن به «محتوا»، اجازه دهید در مورد «ساختار» خود کتاب صحبت کنیم و چند نکته بسیار مهم را - به عنوان نکاتی که مترجمان محترم کتاب می‌توانند آن را در ویرایش‌های بعدی اصلاح کنند - بیان کنیم.

کتاب ترجمه شده دارای سه نقص کلیدی است که دو مورد مهم آن در متن اصلی کتاب وجود نداشته است.

نقص اول، نبود معرفی کوتاهی در مورد کتاب است، که یکی از قسمت‌های مهم کتاب اصلی به شمار می‌رود. این معرفی کوتاه



وی در سال ۱۹۹۱ در نوشتن دستنامه «الگوریتم‌ها و ساختمان‌های داده‌ها» از انتشارات آدی سان‌وسلی^۱ به عنوان نویسنده همکار و در سال ۱۹۹۲ برای نوشتن کتاب «بازیابی اطلاعات: الگوریتم‌ها و ساختمان‌های داده» از انتشارات پرنیس هال^۲ به عنوان نویسنده همکار فعالیت داشت. همچنین چند مقاله در ژورنال‌های تخصصی SIAM, ACM, IEEE منتشر کرده و عضو انجمن‌های SIAM, IEEE, EATCS, AMS, ACM نیز می‌باشد.

نویسنده همکار، برتیه ریبروتو مدارک کارشناسی خود را در رشته‌های ریاضیات و مهندسی برق، مدرک کارشناسی ارشد خود را در رشته کامپیوتر از دانشگاه فدرال میتاس گرایس برزیل و در سال ۱۹۹۵ مدرک دکترای خود را از دانشگاه کالیفرنیا در لس آنجلس اخذ کرد. زمینه‌های مورد علاقه وی در حوزه‌های بازیابی اطلاعات کتابخانه‌های دیجیتال و ... می‌باشد. او در پروژه‌های بزرگی که در وزارت علوم و فناوری برزیل انجام شده شرکت داشته است و همچنین عضو IEEE, ASIS, ACM بوده و در پروژه‌های بزرگی با ریکاردو بیزا بیترس در امریکا و برزیل همکاری داشته است.

وی در سال ۱۹۹۱ در نوشتن دستنامه «الگوریتم‌ها و ساختمان‌های داده‌ها» از انتشارات آدی سان‌وسلی^۱ به عنوان نویسنده همکار و در سال ۱۹۹۲ برای نوشتن کتاب «بازیابی اطلاعات: الگوریتم‌ها و ساختمان‌های داده» از انتشارات پرنیس هال^۲ به عنوان نویسنده همکار فعالیت داشت. همچنین چند مقاله در ژورنال‌های تخصصی SIAM, ACM, IEEE منتشر کرده و عضو انجمن‌های SIAM, IEEE, EATCS, AMS, ACM نیز می‌باشد.

نویسنده همکار، برتیه ریبروتو مدارک کارشناسی خود را در رشته‌های ریاضیات و مهندسی برق، مدرک کارشناسی ارشد خود را در رشته کامپیوتر از دانشگاه فدرال میتاس گرایس برزیل و در سال ۱۹۹۵ مدرک دکترای خود را از دانشگاه کالیفرنیا در لس آنجلس اخذ کرد. زمینه‌های مورد علاقه وی در حوزه‌های بازیابی اطلاعات کتابخانه‌های دیجیتال و ... می‌باشد. او در پروژه‌های بزرگی که در وزارت علوم و فناوری برزیل انجام شده شرکت داشته است و همچنین عضو IEEE, ASIS, ACM بوده و در پروژه‌های بزرگی با ریکاردو بیزا بیترس در امریکا و برزیل همکاری داشته است.

محورهای اصلی کتاب

توصیف کتاب از دیدگاه نویسنده: «باید گفت که این کتاب تصویری کلی از مسائل و مشکلات حوزه بازیابی اطلاعات را با رویکردی پژوهشی از دید یک متخصص علوم رایانه ارائه می‌دهد. بنابراین توجه این کتاب بر فنون و الگوریتم‌های رایانه‌ای است که در سیستم‌های بازیابی اطلاعات برای بهبود یا حل مسائل و مشکلات این حوزه مورد استفاده قرار می‌گیرد» (رویکرد کامپیوتر - محور).

این تأکید نشان می‌دهد که نویسندگان اصلی کتاب مانند بیشتر کارشناسان تکنوکرات حوزه علوم رایانه، محور اصلی حرکت خود را بر مبنای مسائل فنی حوزه بازیابی اطلاعات قرار داده‌اند و کمتر بر مسائل و مشکلات انسانی بازیابی اطلاعات توجه داشته‌اند.

جالب است که در این کتاب و برخی کتاب‌های دیگر این مسئله به تصریح بیان شده است که کارشناسان علوم کتابخانه و اطلاعات

کتابخانه) تأکید شایانی شده است. کتاب، علاوه بر مسائل فنی و تکنیکی بازیابی اطلاعات، به انسان - محوری و مسائل انسانی بازیابی اطلاعات پرداخته، که در این قسمت به نقش کارشناسان کتابداری و اطلاع‌رسانی (علوم مطالعات و کتابخانه) تأکید شایانی شده است.

کلیاتی دیگر درباره کتاب

سناریوی کتاب و ترتیب مطرح شدن موضوعات آن سیر منطقی بسیار جالبی را برای انتقال دانش و تجربه مؤلفان به مخاطب پیموده است. کتاب مخصوصاً در چند فصل اولیه از نمودارهای بسیار رسا و روشنی برای انتقال مفاهیم استفاده کرده، که کارایی و اثربخشی مطالعه کتاب را دو چندان کرده است.

همچنین استفاده از موضوعات کاربردی که جنبه‌های علمی مطالب نظری کتاب را برای خواننده به تصویر می‌کشد یکی دیگر از ویژگی‌های برجسته کتاب است این ویژگی‌ها در کنار استفاده مناسب از پانوشته‌های مربوط به اصطلاحات انگلیسی واژه‌ها، کارایی بیشتری به کتاب داده است.

در ادامه معرفی کوتاهی از فصل‌های کتاب ارائه می‌شود

فصل اول به مقدمات می‌پردازد. در این بخش به محورهای قانونی کتاب، همچون تأکید بر بازیابی اطلاعات به جای بازیابی داده‌ها، مفاهیم اساسی در بازیابی اطلاعات و گذشته، حال و آینده این حوزه موضوعی پرداخته است. همچنین اشاراتی نیز به بحث بازیابی اطلاعات در کتابخانه‌ها و کتابخانه‌های دیجیتالی شده است. بررسی فرایند بازیابی اطلاعات از یک نگاه کلی از جالب‌ترین بخش‌های فصل

اول کتاب می‌باشد. این فصل با توصیف ساختار کتاب (موضوعات و فصل‌های کتاب) و چگونگی استفاده از کتاب پایان می‌پذیرد. فصل دوم به مدل‌سازی فرایندهای بازیابی اطلاعات پرداخته و در این مسیر ابتدا مدل‌های مختلف بازیابی اطلاعات بررسی شده است. همچنین برخی مفاهیم بنیادین بازیابی اطلاعات به عنوان زیرساخت‌های نظری فصل‌های بعدی کتاب به تفصیل مورد بحث قرار گرفته‌اند.

برخی از مدل‌های بررسی شده در این فصل عبارتند از: مدل‌های کلاسیک (بولی، برداری، احتمالی) گونه‌های احتمالی نوین، بازیابی متون ساخت‌یافته و مدل‌های برای مرور (فرامتن‌ها).

به دنبال طرح بحث مدل‌های بازیابی، بحث کارایی عملکرد آنها به ذهن می‌رسد. به این ترتیب فصل سوم به بحث درباره «ارزیابی مسئله بازیابی» می‌پردازد. برای نمونه در سیستم‌های بازیابی اطلاعات ارزیابی میزان ربط مدارک بازیابی شده به نیاز اطلاعاتی کاربر یک موضوع جدی است و در این فصل روش‌های ارزیابی مسئله و مسائلی نظیر آن مورد بحث قرار گرفته است.

در بازیابی اطلاعات به شیوه سنتی، کاربران پرسش‌های خود را به صورت مجموعه‌ای از کلیدواژه‌ها بیان می‌کنند که روش بسیار ساده‌ای است؛ اما مشکلات فراوانی را در بازیابی اطلاعات مناسب ایجاد می‌کند. پس طبیعی است که در فصل چهارم درباره مشکلات مربوط به انواع پرسش‌های طرح شده توسط کاربران بحث شده و «انواع مختلف زبان‌های پرس‌وجو» بررسی شود.

گاه ممکن است کاربر در پرسش خود به ساختار مدرک نیز اشاره کند؛ پس این ساختار (یعنی گنجانیدن توصیف محتوای یک مدرک و ساختار آن) باید با یک زبان متنی مانند زبان نشانه‌گذاری تعمیم یافته استاندارد^۲ تعریف شود. این موضوع در فصل شش با عنوان «زبان‌های متنی» مورد بحث قرار گرفته است.

عملیات پرس‌وجو موضوع بحث فصل پنجم است. گفته شد که در شیوه بازیابی سنتی - که به صورت ارائه کلیدواژه از سوی کاربر انجام می‌شود - کیفیت بازیابی ممکن است به دو دلیل پایین باشد. نخست اینکه به احتمال، پرسش کاربر ساختار ضعیفی داشته باشد؛ یعنی هم از تعداد اصطلاحات کم تشکیل شده و هم به لحاظ مفهومی ضعیف باشد. این شکل می‌تواند با ایجاد تغییراتی در پرسش، مثل انبساط پرسش^۳ یا با استفاده از بازخوردهای دریافت شده از کاربر بهبود یابد که با عنوان عملیات پرس‌وجو در فصل پنجم مورد بررسی قرار گرفته است.

دلیل دوم پایین بودن کیفیت بازیابی در سیستم کلیدواژه‌ای است که در فصل هفتم مورد بررسی قرار می‌گیرد. بیشتر مجموعه کلیدواژه‌های استخراج شده برای بازنمایی محتوای معنایی مدارک (نمایه‌ها)، در خلاصه کردن محتوای معنایی درک نارسا هستند. این شکل را می‌توان با انجام عملیاتی همچون ریشه‌گیری^۵ شناسایی گروه‌های اسمی و با استفاده از اصطلاحنامه بهبود بخشید که موضوع بحث فصل هفتم با عنوان «عملیات متنی» است.

پس از دریافت پردازش - پرسش به گونه‌ای که در فصول گذشته به آن اشاره شد - سیستم باید بتواند اسناد مرتبط با پرسش پردازش

شده را ارزیابی کند. در نتیجه برای سرعت بخشیدن به این کار بویژه در حجم‌های بسیار بالا، باید فنون مطلوب نمایه‌سازی و جست‌وجو در سیستم مورد استفاده قرار گیرد که در فصل هشتم و نهم مورد بحث قرار گرفته‌اند. تمرکز فصل نهم بر این مطلب است که می‌توان با استفاده از فنون بازیابی موازی و توزیع شده به پردازش و پاسخ‌دهی به پرسش شتاب و سرعت بیشتری داد.

فصل دهم شعاع روشن‌کننده‌ای به فضای بازیابی می‌افکند که در مباحث بازیابی اطلاعات در داخل کشور کمتر مورد توجه قرار گرفته است. اشاره به مبهم بودن پرسش (چه به لحاظ مفهومی و معنایی و چه به لحاظ کمی) این معنا را روشن می‌کند که اصولاً در فضای بازیابی اطلاعات - به پارادایم‌های جدیدی در عرصه محیط رابط کاربر نیاز داریم که در فرایند تدوین پرسش به کاربر کمک کنند و راهکارهایی را برای دیداری کردن اطلاعات مجموعه‌های بزرگ فراهم آورند و این مسائل در فصل دهم به تفصیل مورد بحث قرار گرفته‌اند.

بررسی فنون بازیابی اطلاعات در داده‌های چندرسانه‌ای، موضوع بحث فصل یازدهم کتاب است. محور اصلی بحث بر این استوار شده که چگونه مدارک ساخت‌یافته‌ای را که در بردارنده مواد چندرسانه‌ای مانند صدا و تصویر هستند؛ مدل‌سازی، نمایه‌سازی و جست‌وجو کرد. همچنین در این فصل سیستم‌های بازیابی اطلاعات اداری و پزشکی مورد بحث قرار گرفته‌اند.

فصل دوازدهم به نمایه‌سازی و جست‌وجوی مؤثر و کارآمد برای مواد چند رسانه‌ای پرداخته است.

سه فصل آخر کتاب متوجه کاربردهای بازیابی اطلاعات نوین است که به ترتیب بخش‌های وب، سیستم‌های کتاب‌شناختی و کتابخانه‌های دیجیتالی را مورد بحث قرار می‌دهد.

فصل سیزدهم به معرفی وب و مشکلات اصلی مربوط به جست‌وجو در وب برای یافتن اطلاعات سودمند پرداخته و به طور خلاصه مشهورترین موتورهای کاوش را بازخوانی و بررسی کرده است.

فصل چهاردهم به پایگاه‌های داده‌ای مربوط به اسناد تجاری و فهرست‌های دسترسی عمومی درون‌خطی (اپک‌ها) پرداخته است و این سیستم‌ها را به عنوان بزرگ‌ترین سیستم‌های بازیابی اطلاعات مورد بحث قرار داده است.

و سرانجام فصل پانزدهم به بحث درباره کتابخانه‌های دیجیتالی نوین پرداخته و مدل‌ها، نمونه‌ها، استانداردها و موضوعات ساختاری آنها را بررسی کرده است.

پی‌نوشت‌ها

1. Addi Sonwesley
2. Prentice Hall
3. SGML
4. Query Expansion
5. Stemming