

نظام‌های بازیابی اطلاعات: ویژگی‌ها، آزمون، و ارزیابی

فصلنامه کتاب از ارائه نقد بر متون منتشر شده در حوزه کتابداری و اطلاع‌رسانی استقبال می‌کند. البته نشر این گونه نقدها لزوماً به معنای تأیید مطالب آنها نیست. این نوشته برای دکتر مهرداد ارسال شده است که در صورت دریافت پاسخ از ایشان در شماره بعد درج خواهد شد.

فردریک ویلفرید لانکاستر. نظام‌های بازیابی اطلاعات: ویژگی‌ها، آزمون، و ارزیابی. ترجمه جعفر مهرداد. شیراز: نوید، ۱۳۷۹. ۳۰۸ ص. ۱۶۵۰۰۰ ریال.

علیرضا بهمن‌آبادی^۱

ذخیره و بازیابی اطلاعات یکی از محورهای اصلی و زمینه‌های مهم علم اطلاع‌رسانی به شمار می‌آید. در سال‌های اخیر این حوزه از پیشرفت‌های چشمگیری برخوردار و در عین حال با چالش‌ها و مسائل پژوهشی گوناگونی نیز روبرو بوده است. با وجود اهمیت بسیار این حوزه، متأسفانه مباحث مهم و اصلی آن چنانکه باید و شاید در مدارس کتابداری ما تدریس نمی‌شود و ما به واقع با مفاهیم و نکات اساسی این موضوع بیگانه هستیم. بی‌گمان نبود کتاب‌های درسی و خواندنی مناسب به زبان فارسی در این خصوص به این بیگانگی بیشتر کمک کرده است.

خوشبختانه سال‌های ۱۳۷۹ و ۱۳۸۰ از نظر انتشار کتاب در زمینه بازیابی اطلاعات، سال‌های پرباری بوده‌اند و تا جایی که نگارنده اطلاع دارد به جز کتابی که در اینجا به نقد و بررسی آن می‌پردازیم دو کتاب دیگر نیز ترجمه و به زور طبع آراسته شده است.^۲

کتاب نظام‌های بازیابی اطلاعات، اثر یکی از صاحب‌نظران معتبر علم اطلاع‌رسانی، توسط

۱. عضو هیئت علمی مرکز اطلاع‌رسانی و خدمات علمی جهاد کشاورزی

۲. مشخصات این دو کتاب عبارتند از میراندا لی پائو. مفاهیم بازیابی اطلاعات. ترجمه اسدالله آزاد و رحمت‌الله فتاحی. مشهد: دانشگاه فردوسی مشهد، ۱۳۷۹. ۵۰۰ ص.

استون پولیت. نظام‌های ذخیره و بازیابی اطلاعات: خاستگاه، توسعه، و کاربردها. ترجمه محمد حسین دبانی و جعفر مهرداد. شیراز: کتابخانه منطقه‌ای علوم و تکنولوژی شیراز، ۱۳۸۰. ۲۴۰ ص.

یکی از "اساتید" حوزه کتابداری و اطلاع‌رسانی کشور ترجمه شده است. انتشار چنین اثری به گفته مترجم محترم آن با هدف کمک به "دانشجویان گروه‌های آموزشی علوم کتابداری و اطلاع‌رسانی که در اصل به مدیریت اطلاعات در نظام‌ها و مراکز اطلاع‌رسانی نظر دارند و اعضای هیأت علمی که دروس ذخیره‌سازی و بازیابی اطلاعات را تدریس می‌کنند و ممکن است بخش‌های برگزیده‌ای از کتاب را مفید تشخیص دهند" صورت پذیرفته است.

با توجه به کمبود متون مناسب در این زمینه، تلاش مترجم محترم را باید ستود که نسبت به این حوزه پر اهمیت توجه نشان داده و به ترجمه این کتاب همت ورزیده‌اند.

کتاب شامل شانزده فصل و واژه‌نامه و نمایه است. فصول کتاب عبارتند از "فعالیت‌های اساسی بازیابی اطلاعات"، "نمایه‌سازی موضوعی، واژه‌های موضوعی و واژگان‌های کنترل شده: ابزارهای بازیابی سنتی"، "نمایه‌سازی موضوعی، واژه‌های موضوعی و واژگان‌های کنترل شده: همارایی مفاهیم"، "فایل‌های جست‌وجو و راهکارهای کاوش"، "خدمات آگاهی‌رسانی جاری"، "معیارهای عملکرد نظام‌های بازیابی اطلاعات"، "عوامل مؤثر بر عملکرد نظام بازیابی اطلاعات"، "زبان‌های نمایه، اجزاء و خصوصیات"، "به سوی نظام‌های کاملاً مکانیزه"، "ارزیابی بازدهی عملیات نظام بازیابی اطلاعات"، "تحلیل داده‌های آزمون"، "تفسیر و کاربرد داده‌های آزمون"، "ارزیابی بازده اقتصادی نظام بازیابی اطلاعات"، "تعامل میان نظام بازیابی و استفاده‌کنندگان آن"، "راهکارهای کاوش"، "نتیجه، ملاحظات دربارہ طراحی و عملیات نظام بازیابی اطلاعات".

کتاب حاضر در واقع برگردانی از کتاب "Information Retrieval Systems: Characteristics, Testing, and Evaluation" است که در سال ۱۹۶۸ منتشر شده و در سال ۱۹۷۰ جایزه بهترین کتاب علم اطلاع‌رسانی را از سوی انجمن علوم اطلاع‌رسانی آمریکا^۱ از آن خود کرده است. ویرایش دوم این کتاب در سال ۱۹۷۹ و ویرایش سوم از آن نیز با عنوانی متفاوت^۲ در سال ۱۹۹۳ منتشر گردیده است. با اطلاع از این نکته اولین مطلبی که به ذهن خطور می‌کند این است که مترجم محترم چه اصراری برای ترجمه و ویرایش اول داشته‌اند؟ پاسخ این سؤال متأسفانه در مقدمه مترجم پیدا نمی‌شود. خوانندگان و به‌ویژه مترجم محترم، به‌عنوان یکی از معدود "اساتید" کتابداری کشور، کاملاً مطلع هستند که بخش قابل توجهی از مباحث بازیابی اطلاعات متأثر از فن‌آوری‌های رایانه‌ای است و این فن‌آوری‌ها از سال‌های ۱۹۶۸ به بعد رشد و تحول

1. American Society for Information Science (ASIS)

2. F.W. Lancaster & A.J. Warner. **Information Retrieval Today**. Arlington, VA: Information Resources Press, 1993.

چشمگیری داشته‌اند. علاوه بر آن مطالعات متعددی در خصوص تعامل کاربران با نظام‌های بازیابی اطلاعات و همین‌طور پژوهش‌ها و بررسی‌های متعدد دیگری در این زمینه طی سه دهه اخیر - از زمان انتشار کتاب به زبان اصلی - انجام شده و بر اساس این پژوهش‌ها گرایش‌های نوینی نیز شکل گرفته است. بدین ترتیب ترجمه کتابی به این کهنگی چگونه قابل توجیه است؟ خواننده این کتاب در فضایی سیر می‌کند که در آن هنوز "نظام مدلاز در کتابخانه ملی پزشکی با استفاده از رایانه مدل Honeywell 800 و تجهیزات جانبی قادر است جست‌وجوی دسته‌ای حاوی ۴۰ فرمول‌بندی جست‌وجوی کاملاً پیچیده را در مقابل ۷۰۰/۰۰۰ رکورد با تهیه و تولید خروجی رایانه‌ای مربوط به هر یک از کاوش‌ها که زمان پردازش به‌طور کلی هشت ساعت به‌طول می‌کشد به انجام برساند!" (صفحه ۶۰). همین خواننده نتایج پژوهشی را در واقع مطالعه می‌کند که به قول خود مؤلف برگرفته از همکاری وی با پروژه کرانفیلد اسلیپ (دهه ۱۹۶۰ و ۱۹۵۰) می‌باشد (مقدمه نویسنده)، پروژه‌ای که با همه اهمیت، گستردگی، و اثرگذاری آن مربوط به سه دهه قبل بوده و نشان‌دهنده پیشرفت‌های این حوزه از علم در آن دوره زمانی است. انتخاب چنین متنی و در نظر گرفتن آن برای درس ذخیره و بازیابی اطلاعات می‌تواند معنای دیگری هم داشته باشد و آن عبارت از کهنه بودن نظام آموزشی و بی‌توجه بودن گروه‌های آموزشی کشور (و به‌ویژه گروه‌هایی که مسئولیت آموزش دانشجوی در مقطع دکتری را دارند) به پیشرفت‌های جاری در علم اطلاع‌رسانی است.

از مسئله کیفیت انتخاب این کتاب برای ترجمه که بگذریم، موضوع مهم دیگر کیفیت خود ترجمه است. در این خصوص چند نکته قابل ذکر است:

ترجمه نشدن نمونه‌ها و نمودارها: بسیاری از مثال‌ها، نمونه‌ها، و نمودارهایی که مؤلف کتاب برای فهم بیشتر مطالب ذکر کرده است عیناً در متن آمده و هیچگونه تلاشی برای برگرداندن آنها به زبان فارسی صورت نگرفته است. با توجه به اهمیت نمونه‌ها و نمودارها در متون درسی، ترجمه نشدن آنها، می‌تواند درک مطالب و مباحث کتاب را دشوار کند. در واقع این بدان معنی است که خواننده کتاب همواره باید یک واژه‌نامه انگلیسی فارسی در کنار دست خود داشته باشد تا با رجوع به آن معنی مثال‌ها را دریابد(!). سهل است که در متونی از این دست، برای راحتی درک مطالب توسط خواننده فارسی زبان، چه بسا لازم باشد از نمونه‌ها و مثال‌های فارسی به جای نمونه‌ها و مثال‌های لاتین استفاده شود. باقی گذاشتن مثال‌ها به همان شکل اصلی خود در واقع نوعی رفع تکلیف از مترجم به حساب می‌آید.

ویرایش نشدن متن ترجمه: به‌عنوان مسئله بارز بعدی ویراسته نشدن متن ترجمه قابل ذکر است. این مطلب باعث شده است تا خواننده با متنی ناروان و دشوار روبرو باشد. برای مثال به

دو نمونه زیر که از قسمت‌های نسبتاً خوانای کتاب انتخاب شده توجه کنید:

به علت زیرنظام نمایه سازی دو نوع مشخص نقطه ضعف وجود دارد: (۱) نقطه ضعف‌هایی به خاطر اشتباهات و خطاهای نمایه‌ساز، (۲) نقطه ضعف‌هایی به علت سیاستگذاری درباره تعداد میانگین اصطلاحات انتخاب شده در نمایه سازی. خطاهای نمایه سازی خود بر دو نوع است: (۱) حذف و نادیده گرفتن اصطلاح یا اصطلاحات لازم برای توصیف یک موضوع مهم که در مقاله مورد بحث قرار گرفته است، و (۲) استفاده از اصطلاحی که برای توصیف موضوع مقاله نامناسب به نظر می‌رسد. حذف اصطلاحات معمولاً به ناتوانی‌های بازیافت منجر می‌شود، در حالی که استفاده از یک اصطلاح نامناسب می‌تواند یا ناتوانی دقت بازیافت، (به این معنی که جست‌وجوگر این اصطلاح را در یک استراتژی بکار می‌برد و اثری نامربوط بازیابی می‌کند)، یا ناتوانی بازیافت (یعنی، جست‌وجوگر از اصطلاحات صحیح استفاده می‌کند و مدرک دلخواهی را به این علت که با یک اصطلاح نادرست برچسب زده شده است، از دست می‌دهد) به بار می‌آورد (صفحه ۱۸۶).

متن فوق ترجمه جملات ذیل است:

There are two distinct type of failure due to the indexing subsystem: (a) those due to indexer errors and (b) those due to a policy decision regarding the average number of terms assigned in indexing. Indexer errors are themselves of two types: (a) omission of a term or terms necessary to describe an important topic discussed in an article, and (b) use of a term that appears inappropriate to the subject matter of the article. Omissions will normally lead to recall failure, while use of an inappropriate term (i.e., sheer misindexing) can cause either a precision failure (the searcher uses this term in a strategy and retrieves an irrelevant item) or a recall failure (the searcher uses the correct terms and a wanted document is missed because labeled with an incorrect term).

شاید متن مزبور با اندکی حوصله بیشتر می‌توانست به شکل زیر ترجمه و ویرایش شود:

در ارتباط با زیرنظام نمایه سازی دو نوع نقص جداگانه وجود دارد: نقص‌هایی که به اشتباهات نمایه‌ساز می‌گردد؛ و اشتباهاتی که ناشی از نوع سیاستگذاری درباره تعداد متوسط اصطلاحاتی است که به مدارک تعلق می‌گیرد. اشتباهات نمایه‌ساز خود بر دو نوع است: الف) حذف اصطلاح یا اصطلاحات لازم برای تشریح موضوع مهمی که در مقاله‌ای مورد بحث قرار گرفته است؛ ب) و استفاده از اصطلاحی که برای موضوع مقاله نامناسب به نظر می‌رسد. حذف اصطلاحات معمولاً منجر به افت جامعیت می‌شود؛ درحالی که استفاده از اصطلاح نامناسب (یعنی نمایه‌سازی بد-محض) می‌تواند سبب کاهش مانعیت (بدین معنی که کاربر از

این اصطلاح در جست‌وجوی خود استفاده کرده و موارد نامرتبلی را بازیابی می‌کند) یا کاهش جامعیت گردد (بدین معنی که کاربر از اصطلاحاتی درست استفاده می‌کند اما چون مدرک یا اصطلاحی نادرست معرفی شده بازیابی نمی‌شود).

یا به‌عنوان نمونه‌ای دیگر می‌توان به مورد ذیل توجه کرد:

جامعیت فرمول بندی جست‌وجو مسلماً به سطح همارایی مربوط می‌شود، اما ممکن است میان جامعیت و سطح همارایی رابطه انفرادی وجود نداشته باشد. برای مثال "کمبود پروتئین" و "دیسانتری" و "میکروارگانسیم‌های روده‌ای" یک همارایی سه-اصطلاحی است که از این نظر جامعیت دارند که تمام دیدگاه‌های مربوطه را که به وسیله تقاضای درخواست شده است، پوشش می‌دهند، اما همینطور "کمبود پروتئین" و "دیسانتری" از نظر باسیل، که یک همارایی دو-اصطلاحی است (صفحه ۱۹۴).

که ترجمه بند ذیل است:

Exhaustivity of a search formulation is obviously related to the coordination level (i.e., the number of index term required to co-occur before an article can be retrieved), but there may not be a strict one-to-one relationship between exhaustively and coordination level. For example, "protein deficiency" and "dysentery" and "intestinal microorganisms" is a three-term coordination that is exhaustive in that it covers all the related notions demanded by the requester, but so also does "protein deficiency" and "dysentery, bacillary", which is a two-term coordination.

همانطور که ملاحظه می‌شود بند مزبور اساساً در پایان رها شده است و مبتدای جمله آخر بدون خبر مانده است. متن مزبور نیز می‌توانست با اندکی دقت و تأمل بیشتر به شکل زیر ترجمه و ویرایش شود:

کل‌نگری در تدوین فرمول جست‌وجو مسلماً با "سطح همارایی" (یعنی تعداد اصطلاحات نمایه‌ای لازم برای هم‌وقوعی قبل از بازیابی مقاله) در ارتباط است، اما ممکن است رابطه یک به یک جداگانه‌ای میان کل‌نگری و سطح همارایی وجود نداشته باشد. مثلاً "کمبود پروتئین" و "اسهال خونی" و "میکروب‌های روده‌ای" یک همارایی سه‌واژه‌ای است که به لحاظ پوشش دادن تمامی ذهنیات مورد تقاضای درخواست‌کننده کل‌نگر است، اما این کل‌نگری در مورد "کمبود پروتئین" و "اسهال خونی، باسیل" که هماراسته‌ای دو‌واژه‌ای است نیز صادق است (صفحه ۱۹۴).

از جمله موارد دیگر که در ارتباط با کلیت این ترجمه قابل ذکر است عبارتند از:

ترجمه نشدن برخی عبارات‌ها: علاوه بر نمونه‌ها و مثال‌ها، عبارات‌هایی همچون

Item-on-term, term-on-item یا فرمول چهاربزه‌ای پیشنهادی رانگاناتان (صفحه ۲۴) نیز ترجمه شده و به همان شکل اصلی در متن فارسی باقی مانده‌اند. می‌توان این‌طور استنباط کرد که مترجم محترم مسئولیت فهم آنها را مستقیماً برعهده خود خوانندگان کتاب و نهاده‌اند!

معادل‌گذاری‌های نه چندان دقیق: به نظر می‌رسد مترجم محترم حساسیت لازم را نسبت به بعضی از معادل‌ها به خرج نداده‌اند. واژه‌های "دقت بازیافت" (در برابر Precision) و "بازیافت" (در برابر Recall) از این مقوله هستند. دو واژه اخیر از جمله واژه‌های پربسامد و بسیار کلیدی در متون بازیابی اطلاعات محسوب می‌شوند و از این نظر لازم است توجه بیشتری در معادل‌گذاری‌های فارسی آنها صورت گیرد. در این خصوص برخی از مترجمان معادل‌های دقت و بازیافت را در برابر واژه‌های فوق پیشنهاد کرده‌اند و برخی دیگر از معادل‌های جامعیت و مانعیت استفاده می‌کنند که به نظر می‌رسد به لحاظ مفهومی رساتر و نزدیکتر به معنای واقعی این واژه‌ها باشد. در صورت استفاده از برابرنهاده‌های اخیر ممکن است دربارهٔ برابرنهادهٔ "جامعیت" به جای Exhaustivity سوال شود که پیشنهاد نگارنده قراردادن عبارت "کل‌نگری" در برابر واژه مزبور و "خاص‌نگری" در برابر Specificity است. اما معادل‌گذاری "دقت بازیافت" و "بازیافت" به‌طور کلی از این نظر اشتباه است که خواننده با مشاهده آنها ممکن است تصور کند که "دقت بازیافت" زیرمجموعه‌ای از "بازیافت" محسوب می‌شود. درحالی‌که مترجم محترم نیک می‌دانند این دو، مقوله‌هایی جدا از هم به حساب می‌آیند.

ترجمه‌های غلط: در مواردی برخی از عبارت‌ها غلط ترجمه شده یا معادل انتخاب شده تناسبی با متن ندارند. به نظر می‌رسد مترجم محترم در موارد متعددی اولین معنی هر لغتی را که در واژه‌نامه ذکر شده و بدون توجه به بافتی که کلمه در آن مورد استفاده قرار گرفته است، برگزیده‌اند. مثلاً در (صفحه ۴۰) Custom-built به جای "سفارشی" به "مشتري خور" (!)؛ در (صفحه ۵۷) واژه Manual به جای "دستی" به "دستورالعمل"؛ در (صفحه ۸۰) عبارت Management Requirement به جای "نیازهای مدیریت" به "شرایط مدیریت"؛ و در (صفحه ۱۵۷) واژه Sources به جای "دلایل، منشأها" به "منابع" ترجمه شده و در مورد اخیر باعث شده است تا عبارت Sources of system failure به "منابع و ناتوانی‌های ضعف نظام" ترجمه شود.

نمایه کتاب نیز موضوع بعدی است که در ارتباط با این کتاب قابل ذکر است. جای خوشحالی و امیدواری فراوان است که مترجم یا ناشر محترم کتاب نسبت به تهیه نمایه توجه به خرج داده و صفحاتی از کتاب را به این مهم اختصاص داده‌اند. در عین حال باید توجه داشت فراهم آوردن نمایه برای کتاب کار ظریفی است که به دقت و توجه فراوان نیاز دارد. تهیه نمایه انتهای کتاب برای منابع کتابداری و اطلاع‌رسانی توسط معلمان و دست‌اندرکاران این رشته

نیازمند دقت و توجهی مضاعف است، چراکه این نمایه‌ها علاوه بر جنبه راهنمایی و اطلاعاتی که دارند به لحاظ آموزشی نیز قابل اهمیت هستند. مروری گذرا بر نمایه این کتاب نشان می‌دهد که نمایه مزبور اولاً فاقد جامعیت است و درج اصطلاحاتی که امکان رجوع به آنها وجود دارد مورد غفلت قرار گرفته است. به‌عنوان مثال خواننده هیچ اصطلاح نمایه‌ای را در باره "استفاده‌کننده"، "فنون آماری در بازیابی اطلاعات"، "طراحی نظام‌های بازیابی" و... نمی‌یابد. ثانیاً نمایه کتاب فاقد هرگونه ارجاع "نگاه کنید" یا "نیز نگاه کنید" است. مدخل‌های "زبان نمایه" با واژگان کنترل شده و "زمان پاسخگویی" با "ارزیابی نظام" از جمله مدخل‌هایی هستند که در همین شکل فعلی نمایه امکان برقراری ارجاع "نیز نگاه کنید" میان آنها وجود داشت. ثالثاً سیاست واحدی در انتخاب مدخل‌های نمایه‌ای اعمال نشده است و از این لحاظ مدخل‌های نمایه‌ای کتاب فاقد یکدستی لازمند، به‌عنوان مثال شما هم می‌توانید "نمایه سازی خودکار" و هم "نمایه سازی اتوماتیک"؛ هم "فهرست سرعنوان‌های موضوعی" و هم "سرعنوانهای موضوعی" را در نمایه این کتاب بیابید. در حالی که با استفاده از ارجاعات "نگاه کنید" مناسب این اشکال به راحتی قابل رفع بود. مطلب چهارم و مهمترین نکته درباره نمایه این کتاب مسئله پراکنده شدن مدخل‌هاست. به‌عنوان مثال مدخل‌های "ارزیابی ربط"، "استانداردهای ربط"، "ربط"، "نسبت ربط" در چهار مکان مختلف این نمایه و "استراتژی جست‌وجو"، "جست‌وجوی تقاضا"، "جست‌وجوی تکراری"، "جست‌وجوی شخصی"، "جست‌وجوی فراگیر"، "جست‌وجوی رایانه‌ای"، "جست‌وجوی گذشته‌نگر"، "جست‌وجوی واگذارشده"، "جست‌وجوی هم‌شیر"، "فرم جست‌وجو"، "فرم برونداد جست‌وجو"، در چندین جای مختلف نمایه پراکنده شده‌اند. این مطلب در مورد بسیاری از مدخل‌های نمایه‌ای دیگر کتاب صادق است. در حالیکه این موضوع نیز می‌توانست فرضاً با انتخاب یک مدخل اصلی و استفاده از بیانگرها به گونه‌ای مناسب حل شود. در مجموع می‌توان گفت نمایه کتاب حاضر، نمایه‌ای درخور و شایسته نبوده و در تدوین آن بدهی‌ترین اصول نمایه‌سازی نادیده گرفته شده است.

شاید پاسخ برخی از این سئوالات و اشکالات را مترجم محترم می‌توانستند در مقدمه کوتاه خود بیان کنند اما افسوس که مقدمه هم به اندازه کافی نارسا و بی‌خاصیت است و کلی‌گویی‌های آن، به‌ویژه در بند اول، درخور کتابی عمومی و نه کاملاً تخصصی در حوزه اطلاع‌رسانی است که به قلم فردی فاقد مطالعات کافی در این حوزه نوشته شده باشد.