

اصول و روش‌های نمایه‌سازی رایانه‌ای

احمد یوسفی^۱

چکیده: مقاله حاضر، علاوه بر ارائه تعریف و اصول نمایه‌سازی رایانه‌ای، به بیان نمایه‌سازی به کمک رایانه، انواع روش‌های نمایه‌سازی رایانه‌ای، شامل نمایه‌سازی رایانه‌ای استخراجی، نمایه‌سازی رایانه‌ای تخصیصی (اسنادی)، نمایه‌سازی رایانه‌ای انتهای کتاب، و مزایا و هزینه‌های نمایه‌سازی رایانه‌ای می‌پردازد. در انتهای بحث، پیش‌بینی‌ها و پیشنهادهایی برای آینده نمایه‌سازی رایانه‌ای، بهبود روش انجام کار، جلوگیری از اتلاف وقت و هزینه ارائه می‌شود.

مقدمه

از جمله عوامل مثبت و اصلی مؤثر بر خدمات نمایه‌سازی و چکیده‌نویسی استفاده از فن‌آوری رایانه‌ای به منظور ایجاد پایگاه‌های اطلاعاتی ماشینی خوان است. تا زمانی که این پیشرفت از سرعت فعلی افزایش هزینه‌های منابع چاپی بکاهد، سودمندی‌های بسیار بیشتری، از جمله قابلیت استفاده از همان پایگاه اطلاعاتی برای ارائه سایر خدمات اطلاع‌رسانی، اشاعه اطلاعات، کاوش گذشته‌نگر، تولید انتشارات خاص، نمایه‌نامه‌ها، چکیده‌نامه‌ها و مانند آن در زمان کوتاه و حجم بالا را موجب می‌شود.

نمایه‌سازی رایانه‌ای دارای اصول، اهداف، انواع، مزایا، هزینه‌ها، و مشکلات خاصی است. برای اقدام به نمایه‌سازی رایانه‌ای باید به کلیه مسائل پیش‌گفته به خوبی آگاه بود تا از اقدام نابه‌جا، اتلاف وقت، هزینه، نیروی انسانی، فن‌آوری و امثال آن جلوگیری شود و نتیجه‌ای مطلوب عاید گردد. به همین دلیل، در مقاله حاضر سعی بر این است که علاوه بر مسائل بالا، پیش‌بینی‌ها و پیشنهادهایی برای آینده نمایه‌سازی رایانه‌ای ارائه شود.

۱. کارشناس ارشد مرکز اطلاع‌رسانی و خدمات علمی جهادسازندگی

تعریف نمایه‌سازی رایانه‌ای^۱

فرایند استخراج مجموعه‌ای از مدخل‌های نمایه‌ای که نشان‌دهنده موضوع متن هستند توسط رایانه از متن ماشین‌خوان را نمایه‌سازی خودکار یا نمایه‌سازی رایانه‌ای می‌نامند. به عبارت دیگر، نمایه‌سازی با انتقال و تخصیص واژه‌های نمایه موضوعی به متن مربوط است. در استخراج واژه‌های نمایه‌ای از متن می‌توانیم از کلماتی که در متن حضور دارند استفاده کنیم. ولی هنگام اختصاص واژه‌نمایه به متن، گاهی اوقات می‌توان از واژه یا واژه‌هایی استفاده کرد که مجموعه‌ای از واژه‌های نمایه‌ای را مشخص می‌کند و ممکن است در متن نیز نیامده باشد^(۱).

اصول نمایه‌سازی رایانه‌ای

نمایه‌سازی رایانه‌ای از دو دستورالعمل عمده و کلی پیروی می‌کند: (۱) تحلیل آماری متن، (۲) تحلیل نحوی متن. در تحلیل آماری متن، عناصری چون بسامد واژه‌های مهم محاسبه می‌شود و برای نمایه‌سازی اهمیت دارد. معمولاً واژه‌های غیراطلاعاتی شامل نام اشیاء، حروف اضافه، حروف ربط، و واژه‌های رایج، از قبیل "چون" و "مانند" و مانند آن در نظر گرفته نمی‌شود. ممکن است واژه‌ها براساس بسامد و تکرارشان رتبه‌بندی شوند. سایر شیوه‌های آماری، هم حضوری یا هم وقوعی واژه‌ها را در جملات یا چند کلمه قبل و بعد از واژه مورد نظر مبنای کار قرار می‌دهند. این شیوه در دهه ۱۹۵۰ آغاز شد و باکسندال^۲ و لون^۳ پیشگامان آن بودند^(۲). بدین ترتیب، نمایه‌سازی رایانه‌ای مستلزم این است که: (۱) متن مدرک قابل خواندن با ماشین باشد، (۲) ماشین بتواند هر یک از کلمات را به‌طور انفرادی و جداگانه شناسایی کند، و (۳) بتوانیم قوانینی برای انتخاب کلمات و عبارات مورد نظر توسط رایانه تدارک ببینیم^(۳).

نمایه‌سازی به کمک رایانه^۲

اغلب نظام‌های نمایه‌سازی خودکار به مفهوم جانشین رایانه به جای انسان واقعاً خودکار نیستند، ولی قصد دارند تا به متخصصان نمایه‌سازی کمک نمایند. اصطلاح مناسب‌تر برای این نظام "نمایه‌سازی به کمک رایانه" است. به‌طور کلی، برای نمایه‌سازی به کمک رایانه دو شیوه عمده را می‌توان تعیین کرد:

(۱) از رایانه برای تهیه انواع گوناگون یادآوری‌ها و نمایش پیوسته به منظور کمک به نمایه‌ساز استفاده می‌شود. خط‌هایی که نمایه‌ساز مرتکب می‌شود (برای مثال استفاده از اصطلاحات

1. Automatic indexing or computerized indexing

2. Baxendale

3. Luhn

4. Computer-aided indexing

غیراستاندارد یا ترکیب سرعنوان یا سرعنوان فرعی غیرمعتبر) ممکن است بلافاصله شناسایی شده و فوراً به اطلاع نمایه‌ساز برسد. بدین نحو، رایانه به فرد کمک می‌کند تا از اصطلاحات و توصیفگرهای کنترل شده استفاده کند.

۲) برنامه‌های رایانه‌ای به منظور خواندن متن (گاهی فقط عناوین و یا چکیده‌ها) و به منظور انتخاب اصطلاحات نمایه‌ای با استفاده از شیوه‌های استخراجی یا تخصیصی (اسنادی) به کار می‌رود. اصطلاحاتی که بدین ترتیب انتخاب شده‌اند توسط نمایه‌ساز بازیابی می‌شود و نمایه‌ساز ممکن است نقاط بازیابی یا توصیفگرهای بیشتری که برنامه‌های رایانه‌ای قادر به تعیین آن نبوده را اضافه کند و یا اصطلاحاتی را که اشتبهاً اختصاص یافته حذف و توصیفگرهای صحیح را جایگزین آن‌ها کند^(۴). بنابراین، در این شیوه رایانه فقط در انتخاب توصیفگر، متخصص نمایه‌سازی را یاری می‌کند.

همچنین می‌توان واژه‌ها را پس از مقایسه با برخی از انواع فرهنگ لغت‌های ذخیره شده که حاوی واژه‌های قابل قبول هستند از متن استخراج کرد. این شیوه مبنای کار مهم نمایه‌سازی به کمک ماشین است که در دهه ۱۹۷۰ در مرکز سندآرایی دفاعی^۱ امریکا اجرا شد. در اصل، زنجیره واژه‌های موجود در عناوین و چکیده‌ها با زبان طبیعی پایگاه اطلاعاتی^۲ مقایسه می‌شدند. واژه‌هایی که همانند واژه‌های موجود در فرهنگ لغت ذخیره شده بودند برای واژه‌نمایه‌ای انتخاب می‌شدند. کلینگیل^۳ و رینکر^۴ نتایج نمایه‌سازی به کمک رایانه را با نتایج نمایه‌سازی دستی مقایسه کردند. نتیجه سه مطالعه موردی نشان داد که نمایه‌سازی ویرایش شده به کمک ماشین می‌تواند با سطح جامعیت نمایه‌سازی دستی هم‌تراز باشد و مانعیت به دست آمده در نمایه‌سازی به کمک رایانه نیز به خوبی مانعیت به دست آمده در نمایه‌سازی دستی می‌باشد. جامعیت نمایه‌سازی رایانه‌ای ویرایش شده مساوی با نمایه‌سازی دستی و مانعیت آن بهتر از نمایه‌سازی دستی است. بنابراین، شایان ذکر است که شیوه‌های نمایه‌سازی به کمک رایانه اصطلاحات زبان طبیعی را به مدرک اختصاص می‌دهند، نه اینکه توصیفگرها توسط نمایه‌سازی دستی تعیین شوند. بنابراین، مطالعات انجام شده توسط کلینگیل و رینکر مقایسه‌های خالصی از نمایه‌سازی دستی در مقابل نمایه‌سازی به کمک رایانه نیستند؛ بلکه بیشتر مقایسه نمایه‌سازی زبان طبیعی در مقابل نمایه‌سازی کنترل و ازگان می‌باشند^(۵).

1. Defense Documentation Center

2. Natural Language Data Base (NLDB)

3. Klingbiel

4. Rinker

انواع نمایه‌سازی رایانه‌ای

۱. نمایه‌سازی رایانه‌ای استخراجی^۱

در نمایه‌سازی استخراجی، کلمات و عبارات متن استخراج شده و برای نمایاندن محتوای متن به کار می‌روند. نمایه‌سازان سعی می‌کنند عبارات متنی را که به نظر می‌رسد شاخص‌های خوبی برای نشان دادن موضوع مدرک هستند انتخاب کنند. عبارات متنی براساس بسامد حضور واژه در مدرک و همچنین براساس مکان حضور واژه - در عنوان، چکیده، شرح تصاویر، و امثال آن - و همچنین براساس بافت آنها متغیر بوده و نوسان دارند.

اگر متن به شکل ماشین‌خوان باشد، واضح است که رایانه می‌تواند با استفاده از همان معیارهای بسامدی، مکانی، و بافتی، نمایه‌سازی استخراجی را انجام دهد. پیشینه نمایه‌سازی خودکار براساس بسامد واژه را می‌توان در دهه ۱۹۵۰ و آثار لون (۱۹۵۷) و بکسندال (۱۹۵۸) ردیابی نمود. زمانی که متن با سیاهه باز دارنده^۲ به منظور حذف واژه‌های بی‌اهمیت (مانند اشیاء، حروف اضافه، حروف ربط، و امثال آن) مقایسه شود و پس از آن واژه‌ها براساس بسامد حضورشان رتبه‌بندی و مرتب شوند، با نوشتن برنامه‌ای ساده می‌توان میزان حضور واژه‌ها را در متن شمارش کرد.

واژه‌هایی که در رأس فهرست قرار دارند به عنوان "واژه‌های نمایه‌ای"^۳ برای مدرک انتخاب می‌شوند. حدنهایی و نقطه قطع انتخاب واژه‌ها ممکن است براساس چندین شاخص، از جمله تعداد واژه معین و قطعی، تعداد واژه بر مبنای طول متن، یا واژه‌هایی که بسامد آنها بیش از حد مشخصی باشد، در نظر گرفته شود. برنامه‌ای که کمی پیشرفته‌تر و پیچیده‌تر باشد می‌تواند عباراتی را که غالباً به‌طور قابل ملاحظه یا معنی‌داری در متن حضور می‌یابند استخراج کند. بدین ترتیب، ترکیبی از واژه‌ها و عبارات می‌تواند نمایانگر مدرک باشد. در نظر گرفتن معیار بسامد برای انتخاب عبارات، کمتر از واژه‌های مهمی که انتخاب می‌شوند لازم الاجرا است.

می‌توان برنامه را طوری نوشت که به جای انتخاب واژه‌ها و عبارات، ریشه لغات را انتخاب کند. بدین ترتیب، ریشه beat ممکن است به جای اشکال beat، beating، و beated انتخاب و ذخیره شود. برنامه‌های خودکار انتخاب ریشه را می‌توان فقط برای حذف انتهای واژه‌های انتخاب شده (برای مثال "tion"، "ed"، "ing"، و "al") به کار برد. البته می‌توان به کلیه واژه‌ها، عبارات، یا ریشه‌ها ارزش وزنی، که نشان‌دهنده بسامد حضور آنها در مدرک باشد، اختصاص داد. برای مثال، می‌توان به ریشه beat براساس میزان حضور آن در متن، به عنوان مثال ۱۲ مرتبه،

1. Automatic extraction indexing

2. Stop List

3. index terms

یک ارزش عددی اختصاص داد.

معیار بسامد را می‌توان با معیارهای دیگر کامل و همراه نمود. برای مثال، بکسندال (۱۹۵۸) پیشنهاد نمود که فقط اولین و آخرین جمله هر بند یا پاراگراف پردازش شود، زیرا یکی از مطالعات وی نشان داد که اولین جمله در ۸۵ درصد مواقع و جمله آخر در ۷ درصد مواقع "جمله اصلی"^۱ است. نظر بر این که "جمله اصلی" بیشترین اطلاعات را در مورد نظر محتوا ارائه می‌دهد. در این نوع نمایه‌سازی خودکار شیوه‌های گوناگون دیگری نیز برای شناسایی بخش‌های "اطلاعاتی"^۲ متن پیشنهاد و آزموده شده است. برنامه‌های رایانه‌ای عناصری از قبیل: عبارات اضافی، کلمات راهنمای متن مانند "نتیجه‌گیری" و "خلاصه" و بخش‌هایی از متن مانند سرعنوان‌ها را جست‌وجو می‌کند.

یکی از معایب آشکار استفاده از بسامد واژگان یا عبارات ساده در انتخاب اصطلاحات این است که حتی پس از استفاده از سیاهه بازدارنده، برخی از کلماتی که به گرات در مدرک حضور می‌یابند ممکن است وجه تمایز خوبی برای تشخیص این رکورد از سایر رکوردهای پایگاه اطلاعاتی نباشد، زیرا آنها به طور متناوب و به گرات در پایگاه اطلاعاتی حضور می‌یابند. مثالی آشکار اینکه کلمات "کتابخانه"، "اطلاعات‌رسانی"، "کتابداری" و "اطلاعات" وجه تمایز خوبی برای رکوردهای خاص در پایگاه اطلاعاتی کتابداری و اطلاعات‌رسانی نیست. بدین ترتیب، برای مثال ممکن است در مدرکی خاص کلمه "کتابخانه" ۱۲ مرتبه تکرار شود در حالی که کلمه "پنبه‌نسوز" فقط ۴ مرتبه تکرار شده باشد. با وجود این، کلمه "پنبه‌نسوز" وجه تمایز بهتری است، زیرا اصطلاحی است که به ندرت در منابع کتابداری و اطلاعات‌رسانی حضور می‌یابد. این کلمه حتی اگر در مدرک فقط یک بار هم حضور یافته باشد باز هم بسیار مهم است.

بسامد حضور واژه در مدرک تنها بسامد مورد نظر در پردازش رایانه‌ای متن نیست. بلکه بسامد حضور واژه در کل پایگاه اطلاعاتی عاملی مهم‌تر است. بدین معنی که، کلماتی بهترین وجه تمایز و شاخص هستند که حضور آنها در مجموعه غیرمنتظره و کمیاب است؛ برای مثال "پنبه‌نسوز" در پایگاه اطلاعاتی کتابداری و اطلاعات‌رسانی و واژه "کتابخانه" در پایگاه اطلاعاتی پنبه‌نسوز. در اطلاعات واقعی، نیازی به شمردن بسامد حضور واژه در کل متن پایگاه اطلاعاتی نیست بلکه فقط بسامد حضور واژه در فایل مقلوب^۳ که برای کاوش متن بکار رفته، کافی است (یعنی تعداد حضور واژه به تعداد حضور کلیه واژه‌ها در فایل مقلوب مربوط است). پس از آن، به جای بسامد مطلق حضور واژه در مدرک، می‌توان از شیوه بسامد نسبی برای انتخاب واژه

استفاده نمود (اسوالد^۱ و دیگران، ۱۹۵۹). در این شیوه، زمانی واژه‌ها یا عبارات انتخاب می‌شوند که در مدرک بسامد حضور بیشتری نسبت به میزان حضور آنها در کل پایگاه اطلاعاتی داشته باشند. این شیوه کمی پیچیده‌تر از شیوه بسامد و مطلق است زیرا در این شیوه باید تعداد بسامد حضور هر واژه در پایگاه اطلاعاتی (نسبت به تعداد کل حضور واژه در پایگاه اطلاعاتی) نگهداری و حفظ شود. و سپس این میزان حضور باید با میزان حضور واژه در مدرکی خاص مقایسه شود.

فهرست و سیاهه واژه‌ها و عبارات استخراج شده از مدرک بر مبنای بسامد نسبی با فهرست بدست آمده بر مبنای بسامد مطلق کمی متفاوت است. اغلب واژه‌ها همان‌گونه باقی می‌مانند. واژه‌های جدید اندک آن‌هایی خواهند بود که بسیار کم در مدرکی خاص حضور می‌یابند، شاید فقط یک مرتبه. به طور کلی اگر آن واژه فقط ۵ مرتبه در یک پایگاه اطلاعاتی با ۱۰ میلیون واژه حضور یافته باشد، یک مرتبه حضور در میان ۵۰۰۰ واژه از یک مقاله نشریه بسیار مهم است! قدر مسلم، واژه‌هایی که نادیده گرفته می‌شوند واژه‌هایی هستند که در حالی که ممکن است به دفعات در یک مدرک حضور یابند، به دفعات در کل پایگاه اطلاعاتی نیز حضور می‌یابند.

البته واژه‌هایی که بر مبنای بسامد نسبی انتخاب شده‌اند نباید با واژه‌هایی که براساس بسامد مطلق انتخاب می‌شوند تفاوت اساسی داشته باشند. برای بازیابی کارآمد اطلاعات، هم واژه‌هایی که وجه تمایز و شاخص خوبی برای مدرک هستند و هم واژه‌هایی که رده‌های مؤثر مدارک را تشکیل می‌دهند مورد نیاز است. ممکن است تنها یک مدرک در پایگاه اطلاعاتی در مورد خطرات استفاده از پنبه‌نسوز در سقف‌های کتابخانه بحث کند. اما شخص بخواهد مدارک مرتبط را نیز بازیابی کند. کلماتی چون خطرات ممکن است مانند پنبه‌نسوز در پایگاه اطلاعاتی کتابداری کاملاً نایاب نباشد، اما برای بازیابی گروه خاصی از مدارک که ممکن است مورد علاقه برخی از استفاده‌کنندگان باشد مفید است.

بنابراین، شاخص‌ها و معیارهای استخراج واژه از مدرک شامل بسامد مطلق، بسامد نسبی، ترکیبی از هر دو، و همچنین معیار مکانی و نحوی است. البته اگر برای انتخاب واژه از شیوه بسامد نسبی استفاده می‌شود، در واقع نیازی به سیاهه بازدارنده نیست. حروف اضافه، حروف ربط، و نام اشیا به همراه اسامی که واژه‌های رایج هستند (مانند کلمه "کتابخانه" در حوزه کتابداری) مکرر در یک مدرک بکار می‌رود. این کلمات در کل پایگاه اطلاعاتی نیز به دفعات حضور می‌یابند. بنابراین در نمایه پذیرفته نمی‌شوند^(۶).

در آن نوع نمایه‌سازی که نتایج آن صرفاً و مطلقاً استخراجی باشد (یعنی نمایه مساوی واژه‌ای به جای مفهومی) معایبی از جمله پراکندگی هم معنی‌ها، عدم حل مشکلات تضاد و ابهام هم نویسه‌ها^۱، و سایر مشکلات ناشی از به کارگیری لغات و زبان خاص توسط نویسنده به چشم می‌خورد.

بنابراین، گام مهم بعدی در تجزیه و تحلیل خودکار متون برای امور پردازش اسناد، توسعه فنون ماشینی برای تشخیص و دسته‌بندی واژه‌های هم معنی و متمایز نمودن معانی گوناگون هم نویسه‌ها است. علاوه بر این، برخی از کارهای نظارت تجویزی واژگان را می‌توان به وسیله انتخاب خودکار واژه‌های مرجع و بوسیله تهیه و اتخاذ تدابیر ایجاد پیوند، مانند ارجاعات "نگاه کنید به" و "نیز نگاه کنید به" عرضه کرد.

نمایه‌سازی رایانه‌ای تخصیصی اسنادی^۲

در نمایه‌سازی رایانه‌ای تخصیصی برای انتخاب اصطلاحات از ابزارهای کنترل واژگان، مانند اصطلاحنامه‌ها، استفاده می‌شود. شایان ذکر است که نمایه‌سازی دستی نیز از همین شیوه و ابزارها مانند اصطلاحنامه، استفاده می‌کند. یعنی نمایه‌سازی دستی همان نمایه‌سازی تخصیصی (اسنادی)، ولی بدون دخالت و استفاده از رایانه است. استخراج واژه‌ها و عبارات از مدارک (نمایه‌سازی رایانه‌ای استخراجی) کاری است که رایانه‌ها می‌توانند آن را به خوبی انجام دهند. استخراج خودکار دو برتری و مزیت آشکار بر استخراج دستی دارد: ۱) سرعت بالا (۲) یکدستی و یکپارچگی کامل. اما نمایه‌سازی تخصیصی و اجرای آن توسط رایانه کاری دشوار و پیچیده است.

شیوه ساده و بدیهی برای اجرای نمایه‌سازی تخصیصی توسط رایانه عبارت از ایجاد بایگانی از واژه‌ها یا عباراتی که گرایش و احتمال حضور مکرر در مدارک مورد نظر برای نمایه‌سازی دارند و ارجاع خودکار این واژه‌ها و اصطلاحات به اصطلاح مرجع است. برای مثال، این نوع بایگانی برای اصطلاح "باران اسیدی" ممکن است شامل اصطلاحاتی چون "بارش اسیدی"، "آلودگی هوا"، "دی‌اکسید سولفور"، و امثال آن باشد. اگر هر اصطلاحی در کنترل واژگان همراه به بایگانی باشد، می‌توان از برنامه‌های رایانه‌ای برای مقایسه عبارات مهم در مدرک به‌طور مشخص آنهایی که براساس معیار بسامدی که قبلاً ذکر شد استخراج می‌شوند با این مجموعه بایگانی‌ها استفاده کرد و هنگامی که بایگانی مدرک با بایگانی واژگان بیش از آستانه معینی جور و همسان باشد، اصطلاح را به مدرک اختصاص می‌دهند.

در حالی که نمایه‌سازی تخصیصی خودکار در مدت ۲۵ سال گذشته توسعه چشمگیری یافته است، هنوز به مرحله‌ای نرسیده‌ایم که بتوان اصطلاحات اصطلاحنامه بزرگی (برای مثال تقریباً دارای ۱۰/۰۰۰ توصیفگر) را کاملاً به طور خودکار و بدون دخالت انسان به مدرک اختصاص دهیم. امروزه، جز برای تولید نمایه چاپی، کمتر به نمایه‌سازی تخصیصی خودکار توجه می‌شود. در ۲۵ سال گذشته این نوع نمایه‌سازی از اهمیت بیشتری برخوردار بود. زیرا ذخیره و پردازش حجم زیادی از متن توسط رایانه بسیار پرهزینه بوده و هر روشی که متن را کوتاه‌تر می‌کرد قابل قبول بود. البته، در حال حاضر اگر متن کامل به شکل ماشین‌خوان باشد و یا اگر چکیده مناسبی داشته باشد، کمتر نیاز به نمایه‌سازی حس می‌شود؛ مگر اینکه قرار باشد از پایگاه اطلاعاتی نمایه چاپی تولید شود^(۷).

نمونه‌ای از اصطلاحنامه که حداقل از روش تجزیه و تحلیل نیمه خودکار برای کارهای نمایه‌سازی و کنترل زبان اسناد استفاده می‌کند توسط شولتز^۱ و شفرد^۲، با برگزاری مقالات کنفرانس فدراسیون انجمن‌های زیست‌شناسی تجربی آمریکا^۳ در سال ۱۹۶۰ شروع به کار کرد. از نویسندگان این مقالات خواسته شد که نمایه اصطلاحات پیشنهادی را به همراه دست‌نویس ارائه کنند. هنگامی که برای پردازش ماشینی آماده شد، این اصطلاحات پیشنهادی در برابر اصطلاحنامه رایانه‌ای اجرا شدند، به نحوی که اصطلاحات عام‌تر جایگزین اصطلاحات بسیار خاص می‌شدند (برای مثال اگر نویسنده اصطلاحات "iodines" یا "fluorides" را به کار برده باشد رایانه اصطلاح "halogens" را جانشین آن اصطلاحات می‌کند و همچنین به نحوی که اصطلاحات و واژه‌های هم معنی ذیل تنها یک اصطلاح مرتجح با یکدیگر دسته‌بندی می‌شوند. ارجاعاتی، مانند "نگاه کنید به" و "نیز نگاه کنید به" نیز به طور خودکار تهیه می‌شوند.

باکسندال نمونه‌ای از عملیات نمایه‌سازی اشتقاقی^۴ مبتنی بر استخراج واژگان و نقش معنایی آنها را ارائه کرده است. ابتدا بوسیله تشخیص خودکار نقطه، ویرگول، دو نقطه و نقطه ویرگول، و به وسیله راهنماهای مشابه فرازبانی عنوان (یا جمله‌ای از چکیده) که قبلاً وارد رایانه شده است، به عبارات، جمله، و واژه‌ها تقطیع و تقسیم می‌شود. در مرحله دوم، با نگاه به فهرست، می‌توان واژگان وارد شده معینی را که به عنوان عملگرهای نحوی کمیته‌نما، معرف عبارت، حروف اضافه، و امثال آن عمل می‌کنند شناسایی کرد. در مرحله سوم، به واژه‌هایی که رایانه می‌تواند آنها را باز هم منقطع یا تقسیم کند، از قبیل حذف کمیته‌نما، افعال کمی، یا کلماتی

1. Schultz

2. Shepherd

3. Federation of American Societies for Experimental Biology

4. Derivative indexing

که به "ed" یا "ing" ختم می‌شوند و کلماتی که پس از خود افعال کمکی دارند، گداهای ویژه‌ای اختصاص می‌دهد. پس از آن، رایانه روابط از پیش تعیین شده معین بین کلمات باقی مانده را که احتمالاً اسامی یا عبارات اسمی هستند جست‌وجو می‌کند^(۸).

نمایه‌سازی رایانه‌ای انتهای کتاب^۱

شکل خاصی از نمایه چاپی، نمایه انتهای کتاب است. پیشینه تهیه این نوع نمایه توسط رایانه به ۲۵ سال پیش باز می‌گردد. آرتانندی^۲ (۱۹۶۳) با استفاده از رایانه در حوزه شیمی نمایه انتهای کتاب تهیه کرد. برای هر مدخل نمایه (اصطلاحات توصیفی^۳ مرجح) سیاهه‌ای از عبارات همبسته (اصطلاحات تشخیصی^۴ و نامرجح) را استخراج کرد. حضور هر یک از اصطلاحات نامرجح یا مرجح در صفحه متن منجر به انتخاب مدخلی نمایه‌ای مرجح برای آن صفحه شد. آرتانندی معتقد بود که کیفیت این نوع نمایه با نمایه دستی برابری می‌کند ولی پرهزینه‌تر است. با وجود این، بخش عمده هزینه مربوط به ماشین‌خوان نمودن متن است. امروزه بدین علت که تقریباً کلیه انتشارات از طریق داده‌های ماشین‌خوان تولید می‌شود، عوامل مربوط به هزینه، دیگر دلیل عاقلانه‌ای برای نمایه‌سازی دستی نیست. با این همه، مشکلات تولید نمایه انتهای کتاب به صورت خودکار دشوارتر از آن است که آرتانندی به آن اشاره می‌کند. حتی در حوزه موضوعی محدود نیز به اصطلاحنامه بسیار بزرگی از اصطلاحات توصیفی مرجح نیاز است و همچنین برای هر یک از اصطلاحات توصیفی مرجح، تعداد اصطلاحات تشخیصی احتمالی و نامرجح نیز بسیار زیاد است. علاوه بر این، هر دو اصطلاحنامه باید به منظور انعکاس پیشرفت‌های جدید و واژه‌های متغیر در آن حوزه، روزآمد نگه‌داشته شود.

البته، آرتانندی به نمایه‌سازی تخصصی (اسنادی) مبادرت ورزیده بود. ولی استفاده از شیوه نمایه‌سازی استخراجی، یعنی استخراج عبارات از متن کتابی که برای استفاده به عنوان مداخل نمایه‌ای مناسب است کار آسان‌تری است. ارل^۵ (۱۹۷۰) شیوه تولید نمایه انتهای کتاب با استفاده از رایانه را که مستلزم استخراج عبارات اسمی است توصیف می‌کند. وی معتقد است با استفاده از شیوه خودکار، به همراه ویرایش پس از آن برای حذف اصطلاحات غیر ضروری، می‌توان نمایه‌های مناسبی برای انتهای کتاب تولید کرد. سالتون^۶ (۱۹۸۹) شیوه‌های به کارگیری تحلیل نحوی برای ایجاد عبارات مناسب جهت استفاده در نمایه‌های انتهای کتاب را تشریح کرد^(۹).

1. Automatic back of the book indexing

2. Artandi

3. Expression term

4. Detection terms

5. Earl

6. Salton

مزایا و هزینه‌های نمایه‌سازی رایانه‌ای

امروزه، چون از همان ابتدا منابع و مدارک به شکل الکترونیکی تولید می‌شوند، این ویژگی تأثیر به‌سزایی در کاهش هزینه‌ها و افزایش مزایا دارد. وقتی مدارک از همان ابتدا به صورت الکترونیکی تولید و منتشر شوند، هزینه ماشینی‌کردن و ورود اطلاعات به رایانه، که هزینه عمده و زیادی است، کاهش می‌یابد. مزایا و هزینه‌های مدارک الکترونیکی عبارتند از:

۱) مزایای نمایه‌سازی مدارک الکترونیکی

الف. مزایای مالی

- تولید فزاینده

- کاهش نیاز به کارکنان و متخصصان

- افزایش دستیابی به اطلاعات جاری با هزینه اندک

- ارائه خدمات بهینه و ارزان به مشتری

- کاهش خطاهای مرتب کردن به صورت دستی

ب. مزایای قانونی

پ. مزایا برای کارکنان

- روزآمد بودن اطلاعات کاری

- کنترل مدرک

- دستیابی از راه دور

- دستیابی همزمان

- کاهش زمان آموزش

ت. مزایای ارزش افزوده

- بهبود خدمات به مشتری

- مزیت رقابت

- مزیت جلب توجه

۲) هزینه‌های نمایه‌سازی مدارک الکترونیکی

- هزینه‌های فن‌آوری

- هزینه‌های بازیابی

- هزینه‌های روزآمد سازی^(۱۰)

- هزینه‌های نرم‌افزاری و برنامه‌نویسی

البته، باید توجه داشت که این هزینه‌ها فقط برای مدارک الکترونیکی است و اگر مدارک چاپی باشند، باید هزینه نسبتاً زیادی برای ماشینی و وارد کردن متن مدرک به رایانه نیز برآورد شود.

پیش‌بینی‌هایی برای آینده نمایه‌سازی رایانه‌ای

۱. مزایای بسیاری برای ادغام منابع ردیف اول و دوم (مانند چکیده‌نامه‌ها و نمایه‌نامه‌ها) بیان شده است. نویسندگان عهده‌دار مسئولیت بیشتری برای نمایه‌سازی و چکیده‌نویسی مقالات خود در نظام‌های پیوسته خواهند شد و در نتیجه نیاز به متخصصان نمایه‌سازی و چکیده‌نویسی کاهش خواهد یافت. این فرآیند، دستیابی به منابع اطلاعاتی را مقرون به صرفه می‌کند و تأخیر در تولید پایگاه‌های اطلاعاتی ردیف دوم را کاهش می‌دهد (دروت^۱ و دیگران، ۱۹۹۷).

۲. ممکن است به واسطه مرکز پردازش و ویراستاری، ادغام انتشار منابع ردیف اول و دوم سریع‌تر و کارآمدتر صورت پذیرد. مرکز پردازش و ویراستاری از مزایای عملیات متمرکز برای انتشار نشریات رایانه‌ای استفاده خواهد کرد و همچنین به تولید محصولات جانبی ماشین‌خوان و انتشار منابع ردیف دوم می‌پردازد.

۳. آینده خدمات نمایه‌سازی و چکیده‌نویسی به توانایی این خدمات در مواجهه با مشکلاتی بستگی دارد که در حال حاضر با آن مواجه هستند. اگر خدمات با فشارها و اثرات تحول فن‌آوری، سازگار و هماهنگ شود هر دو مشکل رشد مداوم منابع ردیف اول و کاهش همزمان تقاضای خدمات منابع ردیف دوم به شکل چاپی حل خواهد شد (بارلو^۲، ۱۹۷۸؛ ویلیام^۳ و براندهورست^۴، ۱۹۷۹).

۴. انتظار می‌رود که در سال‌های آتی، گذر از کاغذ به سوی منابع الکترونیکی تسریع شود. دلایل بسیاری برای چنین انتظاراتی وجود دارد. آشکارتر از همه اینکه، افزایش سریع‌تر هزینه‌های ارائه خدمات منابع ردیف دوم به شکل چاپی نسبت به هزینه‌های ارائه این منابع به صورت الکترونیکی اجتناب‌ناپذیر و حتمی است^(۱۱).

۵. با توجه به گسترش روزافزون تولید اطلاعات بر روی محمل‌ها و رسانه‌های الکترونیکی، به جای تولید بر روی کاغذ، نمایه‌سازی رایانه‌ای نیز رواج بیشتری خواهد یافت. تحقق این امر بدین علت است که در فرایند نمایه‌سازی رایانه‌ای، یکی از مهمترین مراحل، تبدیل اطلاعات چاپی به اطلاعات الکترونیکی است که این فرایند خود دشواری‌ها و هزینه‌های بسیاری را در بر دارد. ولی اگر از همان ابتدا اطلاعات، به جای تولید بر روی کاغذ، بر روی محمل‌ها و

1. Drott

2. Barlow

3. Williams

4. Brandhorst

رسانه‌های الکترونیکی تولید شود راه برای انجام نمایه‌سازی رایانه‌ای هموارتر شده، رواج بیشتری یافته و با سرعت بیشتری انجام خواهد پذیرفت، البته شایان ذکر است که رعایت و پیروی از استانداردهای یکسان و واحد و یا استفاده از استانداردهای سازگار و هماهنگ با یکدیگر در امر تولید، پردازش، و بازیابی اطلاعات، و همچنین استفاده از نرم‌افزارهای نمایه‌سازی رایانه‌ای یکسان، شرط اول این تسریع و رواج روزافزون است.

۶. ممکن است به مرور زمان خدمات نمایه‌سازی و چکیده‌نویسی دستی به‌طور کلی بر چیده و تعطیل شود. زمانی مقالات نشریات چکیده نداشتند و باید برای آنها چکیده تهیه می‌شد. ولی به مرور زمان برخی از نویسندگان خود چکیده مورد قبولی برای مقالاتشان نوشتند و با ورود رایانه به عرصه خدمات اطلاع‌رسانی و ظهور پایگاه‌های اطلاعاتی نقش فرد فقط وارد کردن چکیده مقالات به رایانه است. یعنی چکیده‌نویسی دستی اهمیت خود را از دست می‌دهد و وظیفه ما تبدیل کردن چکیده به شکل ماشین‌خوان خواهد بود. پس از وارد کردن چکیده با استفاده از نرم‌افزارها و برنامه‌های رایانه‌ای می‌توان اقدام به نمایه‌سازی رایانه‌ای کرد. و رفته رفته رایانه نقش نمایه‌ساز و چکیده‌نویس را کم رنگ می‌کند. بدین ترتیب، به مرور نقش رایانه افزایش می‌یابد و به همان میزان نقش نمایه‌ساز و چکیده‌نویس کاهش می‌یابد. ادامه این امر به جایی می‌رسد که رایانه کاملاً جای نمایه‌ساز و چکیده‌نویس را می‌گیرد و زمانی که کلیه نشریات به‌صورت الکترونیکی منتشر شده و وارد پایگاه‌های اطلاعاتی شوند، نمایه‌سازی و چکیده‌نویسی رایانه‌ای کاملاً جایگزین نمایه‌سازی و چکیده‌نویسی دستی خواهد شد.

پیشنهادها

۱. هدف از نمایه‌سازی بازیابی است، اگر مدارک بازیابی نشوند مثل این است که اصلاً وجود ندارند.
۲. بهتر است سعی در خرید یا تهیه و طراحی نظامی باشید که برای کاربران شما مناسب باشد نه اینکه سعی در تغییر کاربران برای هماهنگی نظام نمایه‌سازی داشته باشید.
۳. آموزش استفاده از نظام نمایه‌سازی به کاربران ضروری است.
۴. برای جلوگیری از وارد کردن مجدد اطلاعات در رایانه و نمایه‌سازی اضافی، در مورد چگونگی استفاده از اطلاعاتی که از قبل به شکل الکترونیکی دارید فکر کنید. ممکن است بتوانید به‌طور خودکار برخی از فیله‌های نمایه‌ای را ایجاد کنید.
۵. به منظور کاهش هزینه‌های آتی و تسریع در امر نمایه‌سازی رایانه‌ای، بهتر است که از همان ابتدا سعی در تولید منابع به‌طور الکترونیکی داشته باشید.
۶. برای خرید نظام‌ها و نرم‌افزارهای نمایه‌سازی رایانه‌ای بهتر است با فروشندگان معتبر، که

استانداردها را رعایت می‌کنند و از زبان‌های برنامه‌نویسی غیراختصاصی استفاده می‌کنند کار کنید.

۷. قبل از اقدام به نمایه‌سازی رایانه‌ای حتماً مطالعات امکان‌سنجی صورت پذیرد تا از اتلاف هزینه و وقت جلوگیری شود.

۸. باید کلیه هزینه‌های مستقیم و غیرمستقیم نمایه‌سازی رایانه‌ای برآورد، و از نظر هزینه - بهره‌وری بررسی شود.

۹. توصیه می‌شود که اگر مدارک شما به صورت الکترونیکی نیست، از نمایه‌سازی رایانه‌ای آنها خودداری کنید. زیرا هزینه ورود اطلاعات و ماشینی کردن مدارک، با استفاده از هر یک از شیوه‌های موجود، بسیار پرهزینه خواهد بود و دشواری‌های زیادی را به دنبال خواهد داشت. از طرف دیگر، انجام این کار از نظر هزینه، و سودمندی مقرون به صرفه نیست.

منابع

1. Dugan, James A. and Minker, Jack. Automatic data processing: Library and information center applications. in encyclopèdia of library and information science, New York: Marcel Dekker, 1969. V.2, P.220.
2. Ibid.
3. Vickery, B.C. Analysis of information in encyclopedia of library and information science. New York: Marcel Dekker, 1968. V.1, P.362
4. Lancaster, F.W. Indexing and abstracting in theory and practice. Illinois: University of Illinois, 1991. P.228-229.
5. Ibid: P.224.
6. Ibid: P.221-224.
7. Ibid: P.224-228.
8. Stevens, Mary Elizabeth. Automatic analysis in encyclopedia of library and information science, New York: Marcel Dekker, 1969. V.2, P.156-158.
9. Lancaster, F.W.: P.228.
10. 1998 indexing digital (electronic) documents: it's not an option; pay now or pay (more)... [Onlin] available: WWW. gslis. Utxas. edu
11. Lancaster, F.W.: P.253-255.