

مروری بر نمایه‌سازی خودکار و نرم‌افزارهای رایج در تولید آن

عباس دولانی^۱

محمد رضا فرهادپور^۲

چکیده

گسترش روزافزون منابع اطلاعات علمی، باعث گرایش متخصصان اطلاعات به فشرده‌گویی و استفاده از راهکارهای آسان‌سازی جست‌وجوی اطلاعات شده است. در این بین، نمایه‌سازی یکی از باصرفه‌ترین راه‌های میانبر جهت رسیدن به اطلاعات است. در بین روش‌های بی‌شمار نمایه‌سازی که هر یک دارای نقاط ضعف و قوت خاص خود هستند، نمایه‌سازی خودکار یکی از روش‌هایی است که علاوه بر جست‌وجوی پذیر نمودن اطلاعات موجود، باعث افزایش توان آدمی در برابر پدیده انفجار اطلاعات و افزایش بی‌وقفه داده‌ها، به‌خصوص در قالب الکترونیکی شده است. با پیشرفت و تکامل نرم‌افزارهای نمایه‌سازی خودکار در افزایش سرعت فرایند نمایه‌سازی، تلفیق راهکارهای نمایه‌سازی خودکار و دستی، و کاهش زمان و هزینه‌ها اقبال و گرایش مدیران پایگاه‌های اطلاعاتی را جهت استفاده از این فناوری‌ها بیش از پیش افزایش داده است. این مقاله شامل درآمدی بر نمایه‌سازی خودکار و ارائه نظرات متخصصان در قالب بحث کوتاهی در مقایسه ضمنی دو روش نمایه‌سازی خودکار و دستی و در پایان معرفی جامع نرم‌افزارهای نمایه‌سازی خودکار و نحوه عملکرد آنها می‌باشد.

کلیدواژه‌ها

نمایه‌سازی خودکار، نمایه‌سازی دستی، نرم‌افزارهای نمایه‌سازی، اطلاعات الکترونیکی، بازیابی اطلاعات.

مقدمه

ایتنرنیتی از سوی دیگر، بر حجم اطلاعات قابل دسترس افزوده شده است. برای دسترس‌پذیری اطلاعات (چه به صورت

امروزه، با گسترش و توسعه اطلاعات الکترونیکی از یکسو و رشد سریع اطلاعات

۱. کارشناس ارشد کتابداری و اطلاع‌رسانی پزشکی و مدرس گروه کتابداری و اطلاع‌رسانی دانشگاه پیام نور ارومیه a.doulani@gmail.com

۲. دانشجوی دوره دکتری کتابداری و اطلاع‌رسانی دانشگاه آزاد تهران واحد علوم و تحقیقات

ماشین‌خوان و چه به صورت دستی)، متخصصان اطلاع‌رسانی ناگزیرند از رایانه و برنامه‌های رایانه‌ای بهره‌گیرند. حجم اطلاعات موجود، تنوع ساختار اطلاعات در قالب‌هایی چون اصل مقاله، خلاصه مقالات، چکیده، ضرورت یکسان‌سازی قالب آنها، و ارائه درخواست‌های اطلاعاتی بهره‌گیران به صورت اصطلاح یا واژه‌های منفرد باعث شده است تا ایجاد نمایه‌ها و انطباق درخواست‌ها با شیوه‌های دستی بسیار دشوار گردد.

در طول چند دهه گذشته و با تلاش‌های متخصصان اطلاع‌رسانی نمایه‌سازی خودکار توانسته است به عنوان ابزاری برای ساماندهی و در دسترس قرار دادن حجم عظیم اطلاعات موجود، یا به عبارت بهتر مقابله با انفجار اطلاعات مطرح گردد. هر چند نمایه‌سازی خودکار نیز دارای کاستی‌ها و اشکالاتی است که ابراز چنین ادعاهایی را مورد تردید قرار می‌دهد. در بسیاری از برنامه‌های بازیابی اطلاعات می‌توان فرایندهای الگوریتمی رایانه‌ای را برای پردازش‌های اطلاعاتی و به عنوان جایگزینی برای پردازش‌های فکری بشری به کار گرفت. بی‌شک سهم علم آمار را در ایجاد چنین ساختارهایی نمی‌توان انکار کرد. در فرایندهایی چون تولید نمایه‌های خودکار، طبقه‌بندی مدارک و اصطلاحات، تدوین راهبردهای کاوش، و یا ایجاد ارتباط بین اصطلاحات وابسته می‌توان از رایانه بهره گرفت. تاکنون از رایانه در اغلب فعالیت‌های بازیابی همچون استخراج پیام از متن مدارک، تدوین اصطلاحنامه‌ها، فهم نیازهای اطلاعاتی بهره‌گیران از طریق درخواست‌های آنها، و

مانند آن استفاده شده است، اما تاکنون رایانه نتوانسته است در هیچ موضوعی به اندازه حوزه نمایه‌سازی خودکار به صورت مستقل انجام وظیفه نماید (۳: ۳۴).

نمایه‌سازی خودکار

تعاریف چندی برای نمایه‌سازی خودکار ارائه شده است: «فرایند استخراج مجموعه‌ای از مدخل‌های نمایه‌ای که بیانگر موضوع متن هستند توسط رایانه از متن ماشین‌خوان را نمایه‌سازی خودکار می‌نامند» (۱۵)؛ «اگر نمایه‌سازی را فرایندی دو مرحله‌ای فرض کنیم، شامل: انتخاب و استخدام واژگان که بازنمون مناسبی از اطلاعات باشند و همچنین جست‌وجوپذیر ساختن واژگان انتخاب شده؛ هر نوع نمایه‌سازی را که هر دوی این مراحل و یا حداقل یکی از مراحل فوق را به صورت خودکار و ماشینی ارائه دهد، نمایه‌سازی خودکار می‌نامیم (۱: ۲۳۵)؛ «نوعی از نمایه‌سازی که از الگوریتم‌های اجرایی در محیط‌های الکترونیکی پیروی می‌کند. این الگوریتم‌ها در پایگاه‌های داده با اطلاعات کتابشناختی و حتی متن کامل اجرا می‌شوند. همچنین پایگاه‌های غیرمتنی مانند پایگاه‌های تصویری و یا موسیقی (صوتی) نیز قابل نمایه‌سازی با این الگوریتم‌ها می‌باشند (۱۰: ۷۵-۷۹). اصولاً نمایه‌ها دارای کارکردهای زیر می‌باشند:

۱. محتوای اطلاعاتی مدارک را فشرده می‌سازند،
۲. به عنوان واسطه‌ای برای تطبیق و یکسان‌سازی زبان مدرک و زبان کاوش به کار می‌روند، و

۳. به عنوان ابزاری کارآمد بر شیوه تدوین راهبردهای کاوش در جست‌وجوهای اطلاعاتی نظارت دارند (۸: ۲۵۵-۲۷۷).

نمایه‌سازی پایگاه‌های اطلاعاتی و نمایه‌های ماشینی به روش‌های زیر ممکن است انجام گیرد:

۱. استفاده از رایانه برای انجام امور دفتری نمایه‌سازی مثل ورود اطلاعات در پایگاه‌های اطلاعاتی،

۲. بهره‌گیری از رایانه برای کنترل کیفیت نمایه‌های تولیدی مثل بررسی این مسئله که آیا همه اصطلاحات نمایه در اصطلاحنامه وجود دارند یا خیر؟

۳. بهره‌گیری فکری از رایانه مثل استفاده از رایانه. برای مثال استفاده از رایانه برای وزن‌دهی و انتخاب اصطلاحات نمایه‌ای، و

۴. نمایه‌سازی کامل خودکار به کمک رایانه (۹: ۹۱-۱۶۳).

با توجه به مطالب فوق می‌توان چنین بیان داشت که در نمایه‌سازی خودکار از چهار نوع رویکرد استفاده می‌شود:

۱. رویکرد آماری که در آن بر مبنای شمارش واژه‌ها، فنون آماری مربوط به مجموعه که از دسته‌بندی واژه‌های هم‌وزن و مشابه بهره می‌جوید، عمل می‌کند؛

۲. رویکرد نحوی، که در آن بر دستور زبان و قسمت‌های دستوری جمله و مفاهیم تعیین شده در ترکیبات دستوری طراحی شده مانند عبارات تأکید دارد؛

۳. رویکرد معنایی که به متن و موقعیتی

که واژه در متن قرار گرفته است می‌پردازد؛

۴. رویکرد مبتنی بر دانش که فراتر از اصطلاحنامه و روش‌های هم‌ارزش سعی دارد ارتباط بین واژه‌ها را درک کند (۲).

در پایگاه‌های داده متنی، الگوریتم‌ها را می‌توان در فرایند نمایه‌سازی بر پایه جست‌وجوهای رشته‌ای^۳ به کار برد، که در آن یک سری مدخل نمایه چرخشی، از طریق اصطلاحات نمایه‌ای نشان داده می‌شوند، به طوری که یک رشته را تشکیل دهند، و هدف از آن ارائه نقطه مدخلی به استفاده‌کننده برای همه اصطلاحات نمایه و نشان دادن ارتباط آنها با هم است (۲).

ولی اغلب موارد جست‌وجوها در بیشتر پایگاه‌ها به صورت کلیدواژه‌ای می‌باشند و بدیهی است تولید فایل‌های مقلوب^۴ در چنین پایگاه‌هایی لازم و ضروری است. در جست‌وجوی کلیدواژه‌ای بیشتر سلسله‌مراتب بین موضوعات اصلی و فرعی مدنظر است و یا اینکه الگوریتم‌های نمایه‌سازی واژه‌های مشترک را در بین منابع و ارجاعات مدرک جست‌وجو و با یکدیگر تطبیق می‌دهند. از سوی دیگر، برخی از مراحل نمایه‌سازی دستی کاملاً از الگوریتم‌های از پیش تعیین شده نمایه‌سازی پیروی می‌کنند، همانند آنچه که در نمایه‌سازی ماشینی اتفاق می‌افتد. برای مثال، در نظر بگیریم یک نمایه‌ساز می‌خواهد بر اساس کلمات عنوان نمایه‌سازی کند. او در مرحله‌ای مجبور است کلمات موجود در عنوان مدرک را با یک فهرست کنترلی^۵ تطبیق

3. String searching

4. Inverted files

5. Control list

دهد، این دقیقاً رویدادی است که ماشین نیز انجام می‌دهد (۷: ۲۳۱-۲۵۴). همچنین می‌توان این پرسش را مطرح کرد که آیا روش نمایه‌سازی ماشینی و خودکار قادر به آنالیز و تجزیهٔ چهریزه‌های یک موضوع است؟ در نمایه‌سازی تخصصی^۶ و نمایه‌سازی استخراجی^۷، روش‌های به‌کاررفته شده توسط ماشین و انسان به طرق متفاوت و شکل‌های مختلف استفاده می‌شوند، ولی عملاً یکسان هستند (۱۱). اندرسون و پرز^۸ اظهار می‌دارند اطلاعات و دانش ما از نمایه‌سازی ماشینی به مراتب بیشتر از نمایه‌سازی توسط انسان است، چراکه نمایه‌سازی ماشینی براساس دستورالعمل‌ها با جزئیات آنهاست که باید قبلاً وارد رایانه گردند و این مستلزم تدوین اصول نمایه‌سازی به‌طور کامل است. ولی در نمایه‌سازی انسانی در موقع عمل چنین نظمی وجود ندارد و اغلب نمایه‌سازان در نمایه کردن اطلاعات واحد بر مبنای دانش و نظرات شخصی خود به‌صورت متفاوت عمل می‌کنند و بیشتر مواقع از اظهارنظر در مورد چنین مطالبی پرهیز می‌کنند (۴: ۱۵-۲۵). از طرفی با گذشت زمان و افزایش اطلاعات و مهارت‌های نمایه‌سازان، روند نمایه‌سازی آنها نیز به مرور تغییر می‌کند و احتمال نمایه‌سازی متفاوت متنی واحد از سوی یک نمایه‌ساز غیرقابل اجتناب است.

می‌توان گفت روش‌های نمایه‌سازی مانند کوئیک، کوواک، و کووک^۹ جزء

ابتدایی‌ترین روش‌های نمایه‌سازی خودکار هستند که فرایند نمایه‌سازی از روی کلمات عنوان است. از نمایه‌های خودکار پیشرفته‌تر می‌توان نمایه‌سازی چرخشی^{۱۰} و نمایه‌سازی کلیدواژه‌های مضاعف^{۱۱} (نمایه‌سازی کلیدواژه‌های مضاعف بر پایهٔ تخصیص واژه‌های نمایه‌ای از عنوان‌هایی هستند که به آنها استناد شده است) را نام برد. چنین روش‌هایی بسیار مورد استفاده قرار می‌گیرد. نمونهٔ بارز استفاده از این روش‌ها در بنیاد اطلاعات علمی آی.اس.آی. است. از سوی دیگر، زمانی که پایگاه‌های تمام‌متن نمایه‌سازی می‌شوند، تمامی کلیدواژه‌های موجود در متن، به غیر از واژه‌های ممنوعه، در یک نمایهٔ الفبایی قرار می‌گیرد. ساخت چنین نمایه‌هایی (جدا از کیفیت آنها) توسط رایانه بسیار ساده و توسط انسان بسیار زمان‌بر و طاقت‌فرساست. این نوع نمایه‌ها بسیار ابتدایی هستند و از اصول خاصی پیروی نمی‌کنند، ولی با نمایه‌های دیگر مقایسه می‌شوند. چنین نمایه‌هایی در عین سادگی مزیت‌های قابل ملاحظه‌ای نسبت به انواع دیگر نمایه‌ها دارند. برای مثال نمایه‌سازی کلیدواژه‌ای تمام‌متن را برای پاسخگویی به نیازهای اطلاعاتی مناسب‌تر می‌داند. چنانچه اغلب کاربران خواهان جست‌وجوی مدارک و صفحات مختلف، کلمات و عبارات خاصی هستند که احتمال چنین جست‌وجویی در نمایهٔ کلیدواژه‌ای تمام‌متن بسیار بیشتر است.

6. Assigned indexing
7. Extractive indexing
8. Anderson & Perez

9. KWIC, KWAC and KWOC
10. Rotated subject index
11. Keyword plus index

در چنین نمایه‌هایی مشکل درج کلمات جدید (در علوم مختلف) تا حد زیادی رفع می‌شود، چرا که مدت زمانی طول می‌کشد که واژه‌های جدید وارد اصطلاحنامه‌ها و واژه‌نامه‌ها (به‌عنوان واژگان مهارشده سیستم) شوند. ولی به محض ورود مدرک به پایگاه، تمام متن از جمله واژه‌های جدید نمایه شده و قابل جست‌وجو می‌باشند. می‌توان چنین بیان داشت که استفاده از زبان طبیعی در نمایه‌سازی خودکار قابلیت استفاده از روش‌های پردازش آماری را افزایش می‌دهد. در عوض، استفاده از سایر روش‌های نمایه‌سازی انسانی باعث رفع مشکل نمایه‌سازی واژه‌ها و کلمات دوپهلوی و مبهم می‌شود، که در روش‌های خودکار قابل کنترل نیست (۱۲: ۱۷۵-۱۹۳).

در کل، نگاه معنایی و یا وزن‌دهی به کلمات در نمایه‌سازی خودکار با نمایه‌سازی‌های دستی و انسانی متفاوت است. برای مثال، نگاه معنایی نمایه خودکار به کلمات مترادف می‌تواند بر پایه میزان بسامد کاربرد آنها در یک مدرک باشد، که خود این روش نیز نگاه آماری صرف به موضوع است و برای فرایند حلاجی کلمات از نظر مفهومی و معنای کنایه‌ای آنها دخالت عاملی هوشمند و تحلیگر را می‌طلبد.

از سوی دیگر، اگر به تاریخچه نمایه‌سازی خودکار نظری بیفکنیم، دو مدل غالب در این نوع نمایه‌سازی دیده می‌شود: (۱) نمایه‌سازی فضا برداری^{۱۲}، و (۲) مدل احتمالی. سالتون

و مک‌گیل^{۱۳} اظهار می‌دارند در بین انواع نمایه‌سازی غیراحتمالی، مدل فضا برداری بیشترین کاربرد را داشته است. از سوی دیگر، گاهی اوقات تشخیص تفاوت بین این دو نوع نمایه‌سازی بسیار مشکل است، چراکه برخی از مراحل نمایه‌سازی در هر دو روش از روش دیگری الگوبرداری می‌شوند و گاهی عیناً به‌کار می‌رود (۱۹: ۸۷-۱۷۹). در کل، مقایسه‌های زیادی بین روش‌های مدل‌ها و فنون مختلف نمایه‌سازی صورت گرفته است. برای مثال اسپارک و والکر و رابرتسون^{۱۴} مقایسه‌ای بین مدل پیشنهادی خود و سایر روش‌های نمایه‌سازی اعم از اصول مدل فضایی برداری، مدل‌های احتمالی، روش شبکه‌ای، و نمایه‌سازی پس‌گرا^{۱۵} انجام دادند. نتیجه تحقیق آنها این بود که مدل برداری فضایی رایج‌ترین مدل نمایه‌سازی غیراحتمالی است (۲۲).

جورلند^{۱۶} دلیل تفاوت‌ها و شباهت‌های موجود در به‌کارگیری فنون نمایه‌سازی را که توسط کتابداران و اطلاع‌رسانان به‌کار می‌روند در تفاوت متون علوم مختلف می‌داند. برای مثال او بیان می‌دارد، طبقه‌بندی و انتخاب کلمات نمایه‌ای و در کل استعمال روش‌های مختلف براساس نیازها و هنجارهای علوم مختلف مانند علوم تجربی، فلسفه، و تاریخ است. او در مقاله خود با عنوان «نظریه نمایه‌سازی» سؤالی را مطرح می‌کند که به‌کارگیری روش‌های فوق را در

12. Vector- space model

13. Salton & McGill

14. Sparck, Walker & Robertson

15. Regression indexing

16. Hjørland

نمایه‌سازی خودکار، به چالش می‌کشد: «آیا استعمال و یا اعمال روش‌های نمایه‌سازی انسانی در نمایه‌سازی خودکار صحیح است؟ و آیا اصولاً استفاده از این روش‌های منعطف به‌صورت رایانه‌ای امکان‌پذیر است؟ (۱۴: ۱-۲۱).

لیدی^{۱۷} که در زمینه سطوح مختلف نمایه‌سازی رایانه‌ای به تحقیق پرداخته است بیان می‌دارد، اصول پردازش زبان طبیعی که آنها را زبان‌های سطح پایین می‌گوید (اصوات، کلمات، و عبارات اسمی) بیشتر با نمایه‌سازی خودکار عجین‌اند، درحالی‌که زبان‌های سطح بالا که شامل معانی^{۱۸}، زبان‌شناسی^{۱۹}، واقع‌گرایی^{۲۰}، و تعامل^{۲۱} می‌باشد؛ بیشتر با روح نمایه‌سازی انسانی سازگار است (۱۶: ۲۵۷-۲۷۰). بی‌شک به‌کارگیری روش‌های ذکر شده در روند سیر تکاملی نمایه‌سازی جهت رفع نقایص فرایندهای نمایه‌سازی و غلبه بر مشکلات مالی و زمان صرف شده آن است.

از سوی دیگر، اسونیوز^{۲۲} اظهار می‌دارد تشخیص خودکار موضوع یک مدرک بر پایه اصل اثبات‌گرایی^{۲۳} محض صورت می‌گیرد. او بیان داشت که در نمایه‌سازی خودکار، موضوع و یا موضوعات یک مدرک براساس بسامد و تکرار واژگان (البته غیر از واژگان ممنوعه) و یا محل قرار گرفتن آنها (مثلاً در عنوان) و یا براساس الگوریتم‌های

خوشه‌ای تعیین می‌شود. در روش خوشه‌ای که مبتکر آن گافمن است برونداد حاصل از بازیابی فهرستی است که بر پایه میزان ارتباط احتمالی مدارک با پرسش موردنظر درجه‌بندی می‌شود (۱: ۲۸۹). برای مثال، اگر مدرک الف در مورد موضوع ب باشد و از سوی دیگر، مدرک ج نیز مشابه مدرک الف باشد، پس می‌توان نتیجه گرفت که موضوع مدرک ج نیز موضوع ب است (۲۱).

در کل، روش‌های نمایه‌سازی در رده‌بندی‌های مختلفی گنجانده می‌شوند که در آن مسائل انسانی، رشته‌ای، و مانند آن نقش اساسی دارند. برای مثال به یک نمونه از رده‌بندی روش‌های نمایه‌سازی اشاره می‌کنیم:

۱. روش‌های استقرایی (جزء به کل^{۲۴}) که شامل روش‌های زیر است:

الف) روش‌های نمایه‌سازی کلاسیک،
 ب) روش‌هایی که براساس بسامد واژگان و نوع مدارک است (TF - idf)^{۲۵}، و
 ج) سایر روش‌های نمایه‌سازی اخذ شده از علوم کتابسنجی.

۲. روش‌های منطقی و استنتاجی - قیاسی (کل به جزء^{۲۶}):

الف) روش‌های نمایه‌سازی براساس نظریات رایج جهانی درباره زبان و اندیشه،
 ب) روش‌های ابتدایی و بر پایه معناگرایی،
 و

17. Liddy

18. Meaning

19. Semantic

20. Pragmatic

21. discourse

22. Svenonius

23. Positivism

24. Inductive

25. Term Frequency- invert document frequency

26. Deductive

ج) روش‌های ترکیبی و پیشرفته معناگرایی.
 ۳. روش‌های تاریخی:
 الف) روش‌های برگرفته از علوم کتابسنجی،
 ب) روش‌های مورد استفاده در زبان‌های فرعی، و
 ج) روش‌های تجزیه‌ای (تجزیه و تحلیل مدرک).
 ۴. روش‌های اثبات‌گرایی (روش‌های مبتنی بر اهداف، گرایش‌ها، و الگوها):

الف) شامل روش‌های علوم کتابسنجی،
 و
 ب) بر پایه مدارک نمونه و یا الگو^{۲۷} (۱۸).

طی دهه اخیر، بنیاد اطلاعات علمی آی.اس.آی. پروژه‌های مختلفی در باب نمایه‌سازی خودکار بر پایه اصول رده‌بندی چهریزه‌ای رانگاناتان و تئوری بهاتاچاریا^{۲۸} با عنوان «ساختارهای عمیق موضوعی زبان‌های نمایه‌سازی» انجام داده است، که از آن جمله می‌توان به پروژه‌هایی چون پاپسی^{۲۹} (نمایه‌سازی علوم بر پایه قوانین استنباطی در زبان‌شناسی)؛ پرومتئوس^{۳۰} (نمایه‌سازی بر پایه تجزیه کلمات پرمعنی و عنوان‌ها و استخراج عبارت‌های اسمی^{۳۱} است و در مجموع مدلی معنایی برای نمایه‌سازی را به دست می‌دهد)؛ پروژه ویاسا^{۳۲} (نمایه‌سازی بر پایه تجزیه و

ترکیب واژه‌های موجود در متن) (۶). پروژه‌های قابل توجه دیگری نیز از سوی ا.سی.ال. جهت بهبود بیشتر فرایند نمایه‌سازی ابداع شد که از مهم‌ترین آنها می‌توان به پروژه ورود اسمیت^{۳۳} اشاره نمود. در این روش که از نرم‌افزار نت‌اوول^{۳۴} استفاده می‌شود، هدف تشخیص خودکار عبارت‌های درون مدارک و سازماندهی هوشمندانه آنها در قالب نمایه‌های مفید قابل جست‌وجوست (۴: ۱۵-۲۵).

نمایه‌سازی خودکار در مقابل نمایه‌سازی انسانی

مارتین تولیک^{۳۵} در بیان پیچیدگی و معناگرایی فرایند نمایه‌سازی اظهار می‌دارد: «یکی از دلایلی که رایانه نمی‌تواند به‌طور خودکار به نمایه‌سازی (به معنای واقعی کلمه) بهینه پردازد این است که فرایند نمایه‌سازی و چکیده‌نویسی چیزی بیش از الفبایی کردن واژگان مدارک است. برای مثال چکیده‌نویسی حاصل فرایند کاملاً عقلانی و شهودی^{۳۶} می‌باشد که بر پایه داوری در مورد موضوع یک مدرک و محتوای مدرک است». او در ادامه بیان می‌دارد: «رایانه ابزار مناسبی برای فرایند الفبایی کردن و اعمال مشابه می‌باشد، ولی ابزار خوبی برای تولید محصولات فکری و انتزاعی^{۳۷} مانند چکیده‌نویسی نیست» (۲۳: ۷۹-۹۳). دلیل دیگر از نظر سالتون^{۳۸} این است

27. Exemplary document
 28. Bhattacharyya
 29. POPSI
 30. PROMETHEUS
 31. Noun phrases
 32. VYASA

33. WordSmith
 34. NetOWL
 35. Martin Tulic
 36. Intellectual
 37. Inexplicable
 38. Salton

که واژه‌های نمایه‌ای در یک نمایه رایانه‌ای، که به احتمال زیاد از واژه‌های خود مدرک انتخاب شده است، ممکن است هیچ ارتباطی با واژگان مورد استفاده جست‌وجوگران نداشته باشد و یا مدارک مشابه و هم موضوع دیگر را در زمان بازیابی پشتیبانی نکند. برای مثال، در واژه‌های پزشکی، اغلب موارد تکیه بر واژگان مدرک مورد نظر کافی نیست و نیاز به مرتبط ساختن واژگان مدرک به مترادف‌ها، موضوعات کلی‌تر، و حتی ذکر محل‌های جغرافیایی که دارو و یا ماده شیمیایی در آنجا استعمال می‌شود، لازم و ضروری می‌باشد (۲۰). دلیل دیگر ضعف رایانه در نمایه‌سازی از نظر پری، لینفورد، و ریچ^{۳۹} که تحقیقات زیادی در مورد نمایه‌سازی علوم بین‌رشته‌ای کرده‌اند، نمایه کردن مدارکی است که در مورد موضوعات مختلف بحث می‌کنند و به عبارتی دیگر جزء آثار کلی هستند. واژه‌های زیادی وجود دارند که در متن آورده می‌شوند، ولی به اندازه کافی در متن به آنها پرداخته نشده است و ارزش نمایه کردن ندارند و رایانه با توجه به عدم تشخیص این نکته، تنها به دلیل اینکه این گروه از واژگان در فهرست ممنوعه نیستند، اقدام به نمایه‌سازی آنها می‌کند (۱۷). البته موتورهای امروزی قابلیت نمایه‌سازی معنامحور (در مقابل محتوامحور یا براساس کلیدی واژه) را نیز دارند که جهت رفع مشکل جست‌وجوی کلیدواژه‌ای، از راهکارهایی چون استفاده از فهرست مترادف‌ها و بهره‌گیری از جست‌وجوی فازی بهره می‌برند. یکی از انواع این راهکارها،

نمایه‌سازی معنایی پنهان است. در این روش، کاربر می‌تواند جست‌وجوی خود را با توجه به مفاهیم انجام دهد و از تکیه صرف بر کلیدواژه‌ها رها گردد. البته این روش‌ها در ابتدای مسیر هستند و هنوز مشکلات اساسی بر سر راه وجود دارند (۲). دلیل دیگر بر عدم کارآیی رایانه در نمایه‌سازی را می‌توان به این مطلب نسبت داد که موضوعات اصلی و فرعی انتخاب شده در نمایه باید براساس نیاز و سطح معلومات خوانندگان باشند، چرا که برخی از کاربران در مورد موضوع مورد مطالعه دارای اطلاعات و پیش‌زمینه قبلی هستند و برخی دیگر چنین نیستند و با توجه به اینکه اغلب نمایه‌ها در مؤسسات تخصصی و در ارتباط با رشته‌های خاص تولید می‌شوند، رعایت نکردن این اصل، در نمایه‌های رایانه‌ای نمی‌تواند قابل توجه باشد. حال آنکه تولید نمایه‌های تخصصی توسط انسان با توجه به شناخت نمایه‌ساز از کاربران و اصولاً تعریف اولیه سطح نمایه‌سازی، هدفمندتر دنبال می‌گردد (۱۷). از سوی دیگر، اندرسون و پرز دریافتند که نمایه‌سازی توسط انسان بیشتر گرایش به نمایه‌سازی تخصصی (در رشته‌ای خاص) دارد. این در حالی است که نتیجه کار، هزینه‌ها، و زمان صرف شده که اغلب بیشتر از نمایه‌سازی خودکار می‌باشد، چالش دیگری را پیش می‌کشد. آنها به تفاوت‌های نمایه‌سازی انسان و رایانه نیز پرداختند که حاصل آن به قرار زیر است:

۱. مسئله اصلی در نمایه‌سازی این است که اصولاً چه کسی صلاحیت نمایه

کردن یک مدرک را دارد؟ آیا خود نویسنده در انتخاب واژه‌ها و عبارات نمایه‌ای بهترین گزینه نمی‌باشد؟ و آیا در رایانه می‌توان چنین اموری را مورد بحث قرار داد؟

۲. رایانه می‌تواند تمامی کلیدواژه‌های متن و حتی در یک پایگاه داده را براساس الگوریتم‌های مختلف آماری که قابلیت کارکرد با فنون بازیابی اطلاعات می‌باشند، مرتب نماید (مانند TF-idf). در حالی که در نمایه‌سازی انسانی چنین امکانی وجود ندارد؛ و

۳. در نمایه‌سازی انسانی می‌توان براساس معانی متون و حتی در برخی موارد معانی ضمنی و مستتر^{۴۰} واژه‌های مدرک را به فایل نمایه‌ای افزود، که نمونه بارز آن تشخیص معانی مختلف واژه‌های هم‌نویسه است، در حالی که انجام چنین اموری در رایانه کار ساده‌ای نیست (۵).

نرم‌افزارهای نمایه‌سازی خودکار

اندیشه ساختن نمایه برای مدارک الکترونیکی و فرامتن چندی است که مورد بحث قرار گرفته است. شرکت‌ها و ناشران رسانه‌های الکترونیکی، دارندگان صفحات وب، و سرویس‌دهندگان شبکه‌های اینترنتی در پی نمایه‌سازی مطلوب صفحات فرامتن خود هستند. بی‌شک تکیه صرف به نمایه‌سازی انسانی نمی‌تواند جوابگوی کمبود منابع مالی و مهم‌تر از آن کمبود وقت باشد، زیرا اطلاعات (خصوصاً در فرمت‌های

الکترونیکی) به سرعت در حال افزایش است. طبیعی است استفاده بهینه از اطلاعات، فرایند جست‌وجوپذیر ساختن آن را غیرقابل اجتناب می‌کند. با توجه به محدودیت‌های فوق‌الذکر تمایل و توافق بر سر نمایه‌سازی خودکار بیش از پیش نمایان است. (۱: ۲۳۵). بنابراین، امروزه، با توجه به نیاز روزافزون به نمایه‌سازی فرامتن، ابزارهای مختلفی برای این امر طراحی شده‌اند.

برای انتخاب یک نرم‌افزار نمایه‌ساز ابتدا باید برای سؤالات زیر پاسخی پیدا کنیم: (۱) چه نوع مدارکی و با چه فرمت و قالبی باید نمایه‌سازی شود؟ (۲) چه نوع نرم‌افزاری برای نمایه کردن مدارک موردنظر مناسب‌تر است؟ (۳) مزایا و برتری‌های نرم‌افزار انتخاب شده از میان سایر نرم‌افزارها چیست؟

با توجه به جدید بودن استفاده از چنین فناوری‌هایی، برخی از نمایه‌گران از نرم‌افزارهایی استفاده می‌کنند که کار نمایه‌سازی ساده و سطحی انجام می‌دهند. برای مثال یکی از آنها نرم‌افزارهای هوم‌سایت^{۴۱} است که از طریق نشانی زیر قابل دریافت است:

<http://www.allairc.com/products/homesite/40/hstour/index.cfm>

و یا مورد دیگر نرم‌افزار فرانت پیج^{۴۲} می‌باشد که از طریق نشانی زیر قابل دسترسی است:

<http://www.microsoft.com/catalog/display.asp?site=768subid=228pg=1>.

40. Tacit means

41. Homesite

42. Front page

نرم افزارهای فوق کار نمایه سازی را به صورت ساده و سریع انجام می دهند. تنها مسئله ای که در مورد این نوع نرم افزارها قابل ذکر است، وجود برخی مشکلات در فرایند نمایه سازی است. برای مثال، در نرم افزارهای فوق برای ایجاد یک تورفتگی ساده باید کد اچ.تی.ام.ال. آن را با دقت وارد کنید. در حالی که ایجاد چنین اموری در برنامه های پردازشگر دیگر و حتی نرم افزارهای پیشرفته تر با اعمال تنظیمات مختصری که اغلب با ماوس صورت می گیرد انجام می شود. مشکل دیگر، دشوار بودن تصحیح و غلطیابی در نمایه تولید شده می باشد. با توجه به اینکه کلیه مراحل با درج کدهای اچ.تی.ام.ال. انجام می شود و فرمان های زیادی در آن ثبت شده است پیدا کردن مشکل و یا افتادگی فرمانی مانند یافتن سوزن در انبار گاه است. این در حالی است که یافتن اشتباهی در نمایه دستی (انسانی) به مراتب آسان تر می باشد.

نسل بعدی نرم افزارهای نمایه سازی جهت نمایه کردن وبسایت ها تولید شدند. شاخصه اصلی آنها نداشتن محدودیت های درج کدهای اچ.تی.ام.ال. است. این ابزارها از جهت کار نمایه سازی بسیار ساده تر و در عین حال پیشرفته تر می باشند. از قابلیت های کلی این نرم افزارها تولید برچسب های اچ.تی.ام.ال. به صورت خودکار و همچنین دریافت کدهای اچ.تی.ام.ال. از ناشران آنها به صورت پیوسته است (۱۳). حال به معرفی تعدادی از این سری از نرم افزارها می پردازیم:

نرم افزار HTML/Prep

این نرم افزار از سوی شرکت لوریج تکنولوژی^{۴۳} تولید می شود و از طریق نشانی زیر قابل دسترسی می باشد:
<http://www.levetechic.com/htmlprep.htm>
 قیمت تقریبی آن ۶۵ دلار است. این نرم افزار تحت داس بوده و اساساً کار تبدیل برچسب های ASCII که در یک فایل نمایه ای قرار دارند را به اچ.تی.ام.ال. انجام می دهند. البته نمایه تولید شده به طور مستقیم محصول خود نرم افزار نیست، بلکه برای تولید نهایی باید اصل نمایه در یک نرم افزار پردازشگر دیگر مانند Cindex, Sky index, و Macrex تولید شوند. در حقیقت همان طور که گفته شد HTML/Prep یک نرم افزار تبدیل کننده است که در محیط داس فعال می باشد، در نتیجه، خروجی دیداری آن برای استفاده کاربران مطلوب نخواهد بود.

مراحل نمایه سازی در این نرم افزار به طور خلاصه عبارت است از:

۱. همان طور که گفته شد فرایند نمایه سازی در یکی از نرم افزارهای دیگر نمایه سازی و یا پردازشگرهای دیگر مانند ورد صورت می پذیرد. توجه داشته باشیم که در هنگام ثبت نشانی سایت هایی که برای نمایه سازی انتخاب کرده اید به جای شماره سایت و یا آی.پی. از نشانی نوشتاری استفاده کنیم. این کار سرعت و دقت انجام کار را افزایش خواهد داد.

۲. بعد از نمایه سازی، مدرک را با فرمت تی.اکس.تی. ذخیره کنید. البته اگر نمایه سازی

را در داخل نرم‌افزار Cindex انجام می‌دهید، می‌توانید از فرمت اچ.تی.پی. نیز استفاده کنید (کار با این نرم‌افزار در ادامه خواهد آمد).

۳. پس از تبدیل شدن به اچ.تی.ام.ال. می‌توانید به ویرایش آن از قبیل تغییر رنگ پس‌زمینه و حتی نشان و آرم سایت پردازید.

همچنین، با توجه به اینکه هر حرف در نمایه تولید شده در این نرم‌افزار به صورت پیوند به اشکال مختلف (بزرگ، کوچک، و چسبیده) دارد، می‌توان به راحتی عمل ویرایش را انجام داد. به وسیله این قابلیت می‌توان بدون جست‌وجوی کل صفحه از طریق نوار پیمایش^۴، با کلیک راست، به انتخاب حرف مورد نظر مبادرت کرد. درست مانند تصحیح غلط‌های املایی در نرم‌افزار ورد که با کلیک راست بر روی لغت مورد نظر صورت می‌پذیرد.

۴. می‌توان برای کل نمایه تولید شده نمایه حرفی - الفبایی ایجاد کرد، که کاربر با کلیک کردن بر روی هر یک از حروف به قسمت الفبایی رهنمون شود. مشابه این وضعیت در لغتنامه‌های پیوسته دیده می‌شود که حروف در بالای صفحه به صورت الفبایی درج شده‌اند و با کلیک کردن بر روی آنها می‌توان لغات مورد نظر را مشاهده نمود.

۵. می‌توان به تولید نمایه‌ای از کلمات مهم (برای مثال کلماتی که جزء سرعنوان‌های موضوعی رشته مورد نظر می‌باشد) پرداخت تا کاربران با کلیک کردن بر روی آنها علاوه بر مشاهده واژه‌های مورد نظر تمامی

زیرشاخه‌های آن و یا به عبارتی موضوعات فرعی آنها را نیز به صورت سلسله‌مراتبی مشاهده کنند. این کار با ایجاد پیوند بین سرعنوان‌های اصلی و فرعی انجام می‌شود.

۶. نحوه ارجاع نیز به طرق مختلف صورت می‌پذیرد. برای مثال می‌توان از خط‌دار کردن کلمات و یا حروف ارجاع شده و یا ایجاد رنگ خاص و حتی نشان‌دار کردن آنها با اعداد استفاده کرد. توجه داشته باشید که با توجه به اینکه نمایه‌سازی اصولاً در نرم‌افزار دیگری انجام می‌شود، تا زمانی که نمایه را به نرم‌افزار HTML/Prep تبدیل نکرده‌اید، به راحتی قابل ویرایش است. بدیهی است ویرایش و یا افزودن و حذف آیتمی از نمایه تبدیل شده نهایی مستلزم نمایه‌سازی دوباره تمامی فایل خواهد بود.

نرم‌افزار HTML indexer

این نرم‌افزار نمایه‌سازی توسط شرکت بروان^۵ تولید می‌شود و از طریق نشانی زیر قابل دسترسی است:

<http://www.html-indexer.com/index.htm>

این نرم‌افزار، برخلاف نرم‌افزار HTML/Prep، به تنهایی و بدون نیاز به نرم‌افزار پردازشگر دیگر، کار نمایه‌سازی را انجام می‌دهد.

مراحل نمایه‌سازی در این نرم‌افزار عبارت است از:

۱. ابتدا باید فایل‌ها و مدارکی که می‌خواهیم نمایه‌سازی شوند به نرم‌افزار

معرفی و یا به عبارتی وارد نماییم. این نرم‌افزار هوشمند به جست‌وجو و پردازش فایل‌های موردنظر می‌پردازد. نوع پردازش و جست‌وجو را نیز می‌توانید معین کنید. برای مثال، می‌توان تعیین کرد که واژگان نمایه‌ای مستخرج براساس سلسله‌مراتب موضوعی (البته باید سرعنوان‌های موضوعی رشته مورد نظر از قبل وارد نرم‌افزار شود) و یا براساس بسامد واژگان (در این مورد نیز براساس فرمان نمایه‌گر میزان بسامد نیز تعیین می‌گردد) مرتب شوند. تقریباً به نوعی همان فرایندی که در قسمت جست‌وجوی فایل ویندوز رخ می‌دهد.

۲. اگر علاوه بر نمایه‌واژگان و عبارت‌ها، می‌خواهید ارجاع یا پیوندی نیز از مداخل نمایه‌ای مثلاً به پاراگرافی خاص در متن ایجاد نمایید (ارجاع از واژگان نمایه شده به متن)، باید برای آن مدخل نشان دیگری، علاوه بر نشانی‌ای که نرم‌افزار به آن به‌عنوان واژه نمایه‌ای می‌دهد، اختصاص دهید.

۳. برای تعیین محل مدخل‌ها، نرم‌افزار می‌تواند به هر یک از قسمت‌هایی که در نقشه وب‌سایت موجود است مانند About, News, Event، و غیره آن پیوند داده و محل آن را نیز دقیقاً مشخص کند.

۴. همچنین این امکان وجود دارد که بعد از تولید نمایه در قالب فایل‌های مختلف به اضافه کردن فایل‌های دیگر برای نمایه‌سازی مبادرت کنیم. نرم‌افزار به‌طور خودکار کار تطبیق فایل‌های جدید و ورود واژگان نمایه‌ای آنها را به نمایه قبلی انجام می‌دهد. مثال ساده

آن را می‌توان به افزودن پانویس جدید در میان پانویس‌های قبلی در نرم‌افزار ورد شبیه دانست که به‌طور خودکار کار تطبیق شماره‌ها انجام می‌شود.

۵. از قابلیت‌های دیگر این نرم‌افزار، ایجاد نمایه الفبایی (مانند نرم‌افزار HTML/Prep) در قالب حروف درج شده در بالای صفحه وب‌سایت است.

۶. علاوه بر ایجاد نمایه‌ای از کلمات به‌صورت سلسله‌مراتبی و با تورفتگی (برای موضوعات فرعی) می‌توان یک نمایه به‌صورت ساده و خطی و یا به عبارتی افزوده^۶ تولید کنیم.

۷. ایجاد نمایه به‌صورت تک‌ستونی و دوستونی.

۸. با فشردن کلیدهای کنترل و شیفت می‌توان به‌طور همزمان برای کلمات نمایه‌ای در فایل‌های مختلف ارجاع مشترک ایجاد کنیم.

۹. یکی از مزیت‌های عمده این نرم‌افزار، که در نرم‌افزار قبلی وجود نداشت، این است که با توجه به اینکه تولید نمایه در داخل همان فایل‌هایی که به نرم‌افزار وارد شده است صورت می‌گیرد، در مواقعی که دارندگان صفحات وب، وب‌سایت نمایه‌سازی شده را در راهنماهای موضوعی و یا وب سرورهای دیگر با نشانی متفاوت اینترنتی قرار دادند، می‌توان به‌راحتی با گشودن دوباره همان فایل‌ها در نرم‌افزار و انتخاب گزینه Update به روزآمدسازی نمایه پرداخت و اصولاً نیازی به تغییر یک به یک نشانی‌های اینترنتی

مدخل‌ها نیست. این امکان بسامد اشتباهات نگارشی را توسط انسان به حداقل می‌رساند و مهم‌تر از آن از ایجاد پیوندهای کور جلوگیری می‌کند.

۱۰. قابلیت دیگر این نرم‌افزار امکان تولید گزینه `help` است که در اغلب نرم‌افزارها و وبسایت‌ها با همین نام وجود دارد و در حقیقت راهنمایی برای نمایه محسوب می‌شود. البته ناگفته نماند که برای تولید `help` به نرم‌افزار دیگری به نام `html help` نیازمندیم که این نرم‌افزار از سوی شرکت مایکروسافت، <http://www.microsoft.com>.

به صورت رایگان قابل دسترسی است.

از محدودیت‌های این نرم‌افزار عدم امکان ویرایش نمایه قبل از تکمیل تمامی آن است. در کل، با به کارگیری این نرم‌افزار می‌توانید تا حد زیادی در وقت و هزینه‌ها صرفه‌جویی کنید.

نرم‌افزار RoboHelp HTML Edition

این نرم‌افزار از سوی شرکت `Blue Sky software` تولید شده و از طریق نشانی زیر قابل دسترسی است:

<http://www.blue-sky.com>

این نرم‌افزار جزء قدرتمندترین نرم‌افزارهای تولید `help` برای نرم‌افزارها و وبسایت‌ها و شبکه‌های داخلی (اینترانت‌ها) است.

مراحل اصلی در نمایه‌سازی این نرم‌افزار عبارت است از:

۱. فایل‌هایی که می‌خواهیم مطالب آنها در گزینه `help` درج شود در قالب اچ.تی.ام.ال.

وارد نرم‌افزار می‌کنیم.

۲. کلیدهای واژه‌هایی که می‌خواهیم برای آنها مطالبی در گزینه `help` قرار دهیم انتخاب می‌کنیم (توجه داشته باشید که این قسمت با توجه به سلیقه نمایه‌ساز باید صورت بگیرد). روش انتخاب به دو صورت می‌باشد. نخست اینکه، خود نرم‌افزار تمامی کلمات را به ترتیب حروف الفبا ارائه می‌دهد که با کلیک بر روی آنها و انتخاب گزینه `Add` می‌توان آنها را به فهرست نمایه اضافه کرد. دوم اینکه، واژه‌های مورد نظر خود را در کادری نوشته و گزینه `Add` را انتخاب کنیم.

۳. در هر مرحله می‌توانید به ویرایش نمایه تولید شده با انتخاب گزینه `back` اقدام کنید. امکانی که در نرم‌افزارهای قبلی وجود نداشت.

۴. از مزیت‌های دیگر این نرم‌افزار تولید گزینه `help` به صورت پیوسته است. همچنین در کادر جست‌وجوی این گزینه می‌توان از خاصیت گسترش‌پذیری نیز استفاده کرد (توضیح اینکه خاصیت گسترش‌پذیری یعنی با تایپ حرف و حروف اول یک کلمه تمامی کلمات موجودی که با این حروف آغاز می‌شوند در پایین کادر ظاهر و قابل انتخاب باشند).

از محدودیت‌های آن نیز عدم تولید نمایه‌های حرفی الفبایی و همچنین زمان‌بر بودن بازگذاری آن، به علت تولید نمایه نهایی در فرمت `Java applet` است که تا اندازه‌ای کاربران را در استفاده از این گزینه دلسرد می‌کند.

نرم افزار Macrex

این نرم افزار از سوی شرکتی به همین نام تولید می شود، و از نشانی <http://www.macrex.com> قابل دسترسی است. این نرم افزار در طی ۲۷ سال فعالیت خود به عنوان یک نرم افزار قدرتمند در نمایه سازی محسوب شده و به طور مرتب روزآمدسازی می شود. آخرین ویرایش آن ویرایش یا سری هشت است که بر روی انواع سیستم های عامل ویندوز، ویستا، و مکینتاش قابل نصب و اجراست.

کار با نرم افزار Macrex کمی با بقیه نرم افزارهای نمایه ساز متفاوت می باشد. برای ایجاد و یا درج واژه های نمایه ای در این نرم افزار آشنایی با برخی فرمان ها، که از طریق کلیدهای میان بر صفحه کلید صورت می گیرد، لازم و ضروری است. برای مثال با فشار دادن کلید E یک سند جدید نمایه ای در هنگام ثبت اولین کلیدواژه ایجاد می شود. در این نرم افزار، قابلیت وجود دارد که نمایه ساز می تواند به صورت دستی موضوعات فرعی (هر تعداد که لازم باشد) را به موضوع و واژه اصلی با قرار دادن ویرگول اضافه کند. همچنین، می توان تقسیمات فرعی موضوعی را در مقابل موضوع دیگر کپی کرده و حتی با اعمال تغییرات و پس و پیش کردن آنها این کار را انجام داد. قابلیت دیگر این نرم افزار نمایه ساز این است که نمایه های تولید شده در انواع برنامه های دیگر را می توان به آسانی به عنوان درون داد به نرم افزار وارد کرد و سپس به تنظیم همان نمایه برای مثال به صورت الفبایی و موضوعی پرداخت. همان طور که اشاره شد تمامی این اعمال

با فشردن کلیدهای موجود در صفحه کلید رایانه صورت می پذیرد.

یکی از مزیت های قابل توجه این برنامه این است که پس از اینکه یک واژه نمایه ای و موضوعی اصلی در کادر مربوط درج شد، تمامی موضوعات فرعی آن نیز بلافاصله نمایش داده می شود (البته توجه داشته باشید که قبلاً باید سرعنوان های موضوعی و یا اصطلاحنامه مورد نظر به برنامه معرفی شود). دقت این برنامه نیز آن را جزء برنامه های قدرتمند کرده است. برای مثال، خاصیت هشدار این نرم افزار باعث کاهش قابل توجه در میزان اشتباهات از سوی نمایه ساز می شود. برای مثال هرگاه برای واژه ای نمایه ای محل و یا آدرس ارجاع و دسترسی درج نشده باشد، اختطاری مبنی بر پر کردن فیلد مربوط بر روی صفحه ظاهر می شود. حتی می توان با فعال کردن گزینه ای خاص غلط های نگارشی و اشتباهات در نقطه گذاری را به صورت خودکار توسط برنامه اصلاح کرد. همچنین این برنامه با ارائه اعداد و یا نشان هایی به جای کلمات و عباراتی که امکان دارد در نمایه بارها تکرار شود، از ثبت مکرر و دستی برخی واژه ها و عبارات نمایه ای جلوگیری کرده و با فشردن کلیدی خاص و تعریف شده می توان آنها را در نمایه وارد نمود. علاوه بر این خود کاربر نیز می تواند اعداد و نشان های جدیدی را تعریف کند (مانند ساختن کلیدهای میان بر در واژه پرداز ورد).

این نرم افزار در به کارگیری انواع نشان ها و کاراکترها محدودیتی ندارد. برای مثال کدهای استاندارد (Unicode) را، که اغلب ناشران

در اروپا به کار می‌برند، پشتیبانی می‌کند.

نرم افزار Cindex

این نرم‌افزار نمایه‌ساز از سوی شرکت روچستر^{۴۷} و از طریق نشانی <http://www.indexres.com> قابل دسترسی می‌باشد. از این نرم‌افزار بیشتر برای نمایه‌سازی کتاب‌ها، روزنامه‌ها، و سایر نشریات ادواری استفاده می‌شود. این نرم‌افزار به‌عنوان یک ابزار حرفه‌ای برای تولید انواع نمایه است. به‌وسیله این نرم‌افزار می‌توان واژه‌نامه، نمایه‌های موضوعی، نویسندگان، و مانند آن را براساس نمایه تولید شده اولیه ساخت. این نرم‌افزار قابلیت نمایش نمایه تولید شده در محیط‌های پردازشگر کلمه مانند ورد را داراست که قابلیت غلط‌گیری و رفع اشکال از نمایه تولید شده را میسر می‌سازد. در حقیقت این نرم‌افزار مانند یک نرم‌افزار پردازشگر کلمه کار می‌کند. برای مثال می‌توانیم مدخل‌های جدید را به نمایه اضافه کنیم و حتی با توجه به پیش‌بینی‌های قبلی برای مدارکی که بعداً وارد فایل خواهند شد، مدخل‌های موردنظر را در نظر گرفت.

در ابتدای امر تمامی مدخل‌هایی که توسط نمایه‌ساز وارد نرم‌افزار می‌شوند، به‌صورت الفبایی مرتب می‌گردد (پیش‌فرض آن الفبایی است). این امر، جست‌وجو و بازبینی را آسان می‌سازد. حتی زمانی که نمایه‌ساز تصمیم می‌گیرد مدخل‌هایی را حذف کند و یا نحوه مرتب نمودن نمایه (مثلاً به‌صورت موضوعی) را تغییر دهد، تنظیم نمایه نسخه الفبایی در نرم‌افزار محفوظ می‌ماند. می‌توان

نمایه را به‌صورت الفبایی کلمه به کلمه، حرف به حرف، و یا به‌صورت نزولی و صعودی موضوعی تنظیم و براساس تاریخ مرتب نمود. برای شماره‌گذاری نیز می‌توان از اعداد و یا نشان‌های عربی و لاتین استفاده نمود. در این نرم‌افزار می‌توان با تعریف واژه‌های ممنوعه باعث عدم نمایه شدن آنها شد. از قابلیت‌های دیگر این نرم‌افزار، نقطه‌گذاری، قالب‌بندی، و حذف مدخل‌های تکراری به‌صورت خودکار است. برای قالب‌بندی نهایی نمایه‌ساز می‌تواند از قالب‌بندی‌های پیش‌فرض خود نرم‌افزار انتخاب کند و یا اینکه به‌صورت دستی به قالب‌بندی آن پردازد (همان‌طور که ذکر شد این نرم‌افزار مانند یک نرم‌افزار پردازشگر کلمه نیز کاربرد دارد). نمایه نهایی را می‌توان در فرمت‌های مختلف مانند اچ.تی.ام.ال. و اس.جی.ام.ال. براساس نیاز سفارش‌دهنده تولید کرد. در وارد کردن برخی مدخل‌ها که به‌صورت مکرر در نمایه تکرار می‌شوند، می‌توان برای آنها کلیدهای میان‌بر و یا از علائم اختصاری (مانند نرم‌افزار ورد) تولید کرد، که با وارد کردن آنها در نمایه، مدخل مورد نظر در نمایه ظاهر خواهند شد. نحوه جست‌وجوی نمایه نیز بسیار کاربرپسند و دارای امکانات زیادی است. یک نمونه از جست‌وجوهای قابل توجه این است که می‌توان جست‌وجوها را براساس مترادف‌ها و متضادها انجام داد. یعنی با جست‌وجوی کلمه‌ای خاص مترادف‌ها و یا متضادهای آنها نیز بازایی شوند. همچنین می‌توان از امکان جست‌وجوی کلمات مشابه در املا و نگارش

استفاده کرد. یعنی جست‌وجوی کلماتی که با حرف و یا حروف معینی شروع می‌شوند امکان‌پذیر است. این امکان برای انتخاب صحیح مدخل‌ها توسط نمایه‌ساز بسیار مهم است. از دیگر امکانات این نرم‌افزار قابلیت تصحیح اشتباهات نگارشی از سوی نمایه‌ساز در هنگام وارد کردن مدخل‌ها به صورت دستی است.

گاهی اوقات نمایه‌ساز مشغول اصلاح و یا وارد کردن مدخل‌ها به بخش خاصی از نمایه است، در این صورت می‌توان تمامی جست‌وجوها و سایر اعمال نرم‌افزار اعم از جست‌وجوی واژه و تصحیح مدخل‌ها را منحصر به همان بخش کرد. این قابلیت سرعت پردازش اطلاعات را نرم‌افزار و رایانه به میزان قابل توجهی افزایش می‌دهد. همچنین Cindex توانایی تولید انواع ارجاعات (مانند نگاه کنید به و نیز نگاه کنید) را داراست و به محض درج اولین موضوع فرعی بعد از موضوع اصلی به تولید چنین ارجاعاتی مبادرت می‌ورزد (البته با تنظیمات اولیه که توسط نمایه‌ساز اعمال شده است). حتی شکل ظاهری موضوعات اصلی و فرعی مانند درشت‌نویسی و یا ریزنویسی و مورب^{۴۸} نویسی را می‌توان از ابتدا برای نرم‌افزار مشخص نمود. قابلیت مهم دیگر این نرم‌افزار توانایی تشخیص مدخل‌های تکراری و حتی اشتباه است که این نکات را طی برچسب‌هایی خاص به اطلاع نمایه‌ساز می‌رساند.

نرم‌افزار Retriever

در اصل این نرم‌افزار یکی از ابزارهای نمایه‌سازی تصاویر است. این نرم‌افزار از سوی شرکتی آلمانی با نام Dirk Djuge تولید می‌شود، و از نشانی <http://www.djuge.net/retriever> قابل بازیابی است. این نرم‌افزار در عین کارآمدی و وجود امکانات بسیار، دارای محیطی بسیار ساده است. از جمله امکاناتی که در آن پیش‌بینی شده است می‌توان به موارد زیر اشاره نمود:

۱) تولید انواع نمایه‌های فرعی در قالب فصل‌ها و بخش‌های مختلف با توجه به نیاز کاربر،^{۴۹} (۲) تولید کاربرگه‌های دسترسی سریع، (۳) ذخیره خودکار کاربرگه‌های دسترسی در قابل تصویر برای شناسایی دیداری، (۴) استفاده از تصاویر و قالب‌های موجود در نمایه در تولید یک وب‌سایت (البته براساس فرمت‌های پیش‌فرض اما قابل ویرایش)، (۵) قابلیت جست‌وجو و حذف تصاویر مشابه و تکراری،^{۵۰} (۶) مدیریت و سازماندهی تصاویر در یک پایگاه داده قابل جست‌وجو، (۷) قابلیت نمایش تصاویر در اندازه واقعی خود و به عبارتی نمایش شفافه^{۵۱}، (۸) تولید و تنظیم تصاویر به صورت سری و پشت سر هم^{۵۱}، (۹) گروه‌بندی تصاویر در بخش‌های جداگانه براساس کیفیت تصاویر (براساس پیکسل)،^{۱۰} (۱۰) جست‌وجوی تصاویر براساس رنگ‌های مختلف و فایلی که در آن قرار گرفته‌اند.

این نرم‌افزار قابلیت ذخیره و سازماندهی

48. Italic

49. Contact sheet

50. Slid show

51. Mosaic

تصاویر در قالب‌های جی.پی.جی. و ویندوز بی.ام.پی.، و همچنین فرمت‌های دبلیو.ام.اف.، تی.آی.اف.، جی.آی.اف.، و پی.ان.جی. را داراست. از سوی دیگر این نرم‌افزار قابلیت تبدیل فرمت‌های فوق به یکدیگر را داراست. البته با توجه به امکان فوق نباید این نرم‌افزار را یک تبدیل‌کننده^{۵۲} به حساب آورد. از دیگر امکانات آن، می‌توان به امکان کپی، انتقال، حذف، و بازیافت تصاویر پس از حذف آنها اشاره کرد. این در حالی است که برای بازیافت تصاویر حذف شده نیازی به نرم‌افزار جانبی بازیافت^{۵۳} نیست. در بازیابی تصاویر نیز علاوه بر جست‌وجوی تصاویر براساس رنگ و کیفیت می‌توان از کاربرگه‌های جست‌وجو استفاده کرد. خود کاربرگه‌های جست‌وجو را نیز می‌توان براساس ظرفیت و رنگ تصاویر ذخیره شده در آنها جست‌وجو نمود. نحوه تنظیم و مرتب‌سازی^{۵۴} کاربرگه‌ها نیز به صورت خودکار براساس موارد فوق (ظرفیت، رنگ‌ها، و مانند آن) و همچنین به صورت دستی (drag and drop) صورت می‌پذیرد. می‌توان تصاویر داخل هر کاربرگه جست‌وجو را که همانند فایل‌ی برای ذخیره تصاویر می‌باشند، به قطعه‌های کوچک‌تر تقسیم کرد و از به هم چسباندن آنها تصویری جامع برای نمایش دیداری خود کاربرگه استفاده کرد (مانند تغییر ظاهر آیکن‌ها و فایل‌ها در ویندوز). این کار در بازیابی تصاویر موردنظر می‌تواند مطلوب باشد. از

امکانات کاربرپسند دیگر این نرم‌افزار قابلیت نمایش تصاویر (پشت سر هم با فاصله زمانی مشخص) به‌عنوان تصویر صفحه نمایش^{۵۵} و همچنین به‌عنوان محافظ صفحه نمایش^{۵۶} رایانه است.

بحث و نتیجه‌گیری

نمایه‌سازی یکی از سریع‌ترین روش‌های بازیابی محتوای اطلاعاتی مدارک است. از سوی دیگر، استفاده از هر روش و ترفندی جهت جست‌وجوی اطلاعات، به ماهیت خود اطلاعات و تغییرات مستمر در زمینه‌های تولید آن بستگی دارد. ابداع روش‌های مختلف نمایه‌سازی دستی، تدوین اصطلاحنامه‌ها، و فهرست‌های مختلف واژه‌ها و عبارات مهار شده و گذر از نمایه‌سازی دستی به ماشینی و خودکار، همگی ناظر بر وقوع تغییرات در زمینه تولید و اشاعه اطلاعات می‌باشند. در این بین، نمایه‌سازی خودکار، با توجه به فلسفه وجودی خلق خود، یعنی میزان فزاینده انتشارات الکترونیکی و تمایل کاربران به جست‌وجوهای تمام‌متن، شتاب مضاعفی به خود گرفته است. چنانچه شروع به کار این نوع نمایه‌سازی، صرفاً با به‌کارگیری روش‌های شمارشی و براساس بسامد واژگان بود، تا به امروز که انواع مختلف و قدرتمندی از نرم‌افزارهای نمایه‌ساز خودکار، که توانسته‌اند تا حد امکان تعادل و تعاملی بین قابلیت‌های انسانی و رایانه‌ای در فرایند نمایه‌سازی برقرار نمایند؛ بنابه

55. Desktop

56. Screen saver

52. Converter

53. Recovery

54. Sorting



نیازهای تخصصی مؤسسات سفارش دهنده و پایگاه‌های داده اطلاعاتی به نمایه‌سازی مدارک در فرمت‌های مختلف می‌پردازند. از سوی دیگر امروزه تولید بهترین نمایه مدنظر می‌باشد نه بزرگ‌ترین آن. اگر نگاهی دقیق به کارکرد نرم‌افزارهای نمایه‌سازی (در متن مقاله و سایر انواع آن) کنیم، مشاهده می‌شود که هر یک با راه و روشی خاص و با رعایت جامعیت و مانعیت مطلوب و نمایه‌سازی انواع فرمت‌های اطلاعات مانند متن، صوت و تصویر اقبال مشتریان و کاربران را به خود جلب می‌کنند. همچنین نرم‌افزارهای موجود در موتورهای کاوش در نمایه‌های خود به گزینش سایت‌ها و صفحات وب روی می‌آورند و بسامد روزآمدسازی سایت‌ها را افزایش می‌دهند، چراکه تولید نمایه‌های جامع بدون توجه به کیفیت اطلاعات نمایه شده در آنها، توجیه علمی و اقتصادی نخواهد داشت (۲۰). از طرفی نرخ فزاینده تولید اطلاعات الکترونیکی، ادامه سرمایه‌گذاری‌ها را به طرف نمایه‌سازی دستی، دچار تردید کرده است. چنانچه بانیان پایگاه‌های اطلاعاتی جهت حضور هر چه قدرتمندتر در بازار اشاعه اطلاعات، ناگزیر از به کار بستن برنامه‌های رایانه، برای نمایه کردن اطلاعات خود می‌باشند. از این روست که طراحان نرم‌افزارهای نمایه‌سازی خودکار در پی افزایش توانمندی‌های برنامه‌های تولیدی خود هستند، به نوعی که روش‌های خاص نمایه‌سازی دستی را نیز در برنامه‌های خود اعمال نمایند (۱۷). گواه این مدعا اجرای مطالعات بی‌شماری است که در رابطه با سیستم‌های متن آزاد شکل می‌گیرد،

مطالعاتی با عنوان‌هایی چون تحلیل هوشمندانه متن، فرارسانه‌ها، تحلیل سیستم‌های اطلاعاتی و... از جمله این نمونه‌ها می‌باشند. از سوی دیگر کاربران و نمایه‌سازانی که نگران از دست دادن فرصت بازیابی‌های مفهومی و معنایی هستند، در مواجهه با محدودیت‌های ایجاد شده (زمان، هزینه و...) ناشی از قرار گرفتن در ترافیک‌های بالای اطلاعاتی عصر حاضر، در راستای متعادل نمودن آرمان‌های خود، ناچار به پذیرفتن راهکارهایی چون نمایه‌سازی به کمک رایانه شده‌اند. در این روش، فرایند نمایه‌سازی توسط انسان و رایانه توأمان صورت می‌پذیرد. وانگهی در این روش نمایه‌ساز (عامل انسانی) می‌تواند با انتخاب توصیفگرهایی که توسط رایانه از متن مورد نظر استخراج می‌شود، به نقش‌آفرینی در فرایند مذکور بپردازد. چنانچه الکساندر کایزر^{۵۷} (در مقاله خالویی) نمایه‌سازی به کمک رایانه را راهی برای پیوند دادن نمایه‌سازی دستی و ماشینی معرفی می‌کند، و این فرایند را به سه بخش: ۱) تحلیل متن، ۲) استفاده از روش‌های نمایه‌سازی ماشینی، و ۳) کنترل به وسیله یک نمایه‌ساز متخصص تقسیم می‌نماید (۲). گذشته از مباحث فوق توجه به منابع اطلاعاتی موجود در مراکز مختلف و همچنین نیازهای اطلاعاتی کاربران آنها، در تصمیم‌گیری‌های کلان در امر بازیابی اطلاعات بی‌اثر و در عین حال مخمل نیست. برای مثال در کتابخانه‌های عمومی با توجه به کمبود منابع مالی گرایش به نمایه‌سازی ماشینی برای پرهیز از به کارگیری واژگان کنترل شده (مانند سرعنوان‌های موضوعی) در نمایه‌سازی

57. Alexander Kaiser

که هزینه‌های نمایه‌سازی را به‌طور فزاینده‌ای افزایش می‌دهند، وجود دارد. از سوی دیگر عاقلانه آن است که برای به‌دست آوردن اطلاعات تخصصی در یک کتابخانه عمومی که در طیف وسیع موضوعی به مجموعه‌سازی می‌پردازد، از سرعنوان‌های موضوعی استفاده کرد. این بحث در مورد کتابخانه‌های تخصصی با مجموعه‌سازی موضوعی بسیار متمرکز و خاص و لزوم انجام بازیابی‌هایی با جامعیت بالا، نرخ استفاده از نمایه‌سازی ماشینی و تمایل کاربران برای استفاده از توصیفگرهای خود متن (نمایه‌سازی به زبان مدرک) را متذکر می‌شود. همچنین است در مورد بازیابی مقالات و اطلاعات الکترونیکی پایگاه‌های اطلاعاتی مختلف، که در آن با توجه به تولد و به‌کارگیری واژگان جدید و به‌روز، و در نتیجه استفاده از کلیدواژه‌های خود متن، کفه ترازو را به سود نمایه‌سازی خودکار متمایل می‌سازد. در پایان، گرچه تکرار این مطلب باعث ملال است، لیکن با شرایط کنونی تولید و گردش اطلاعات از یکسو، و نیازهای اطلاعاتی کاوشگران و همچنین لزوم توجه به هزینه‌ها و زمان انجام فرایندهای نمایه‌سازی از سوی اشاعه‌دهندگان اطلاعات، انتخاب روش‌های مقرون به صرفه نمایه‌سازی مانند نمایه‌سازی خودکار و تلفیق قابلیت‌های انسان و رایانه در نمایه‌سازی به کمک رایانه، به شرایط و کمال مطلوب قربت بیشتری دارند.

منابع

۱. پائو، میراندا لی. مفاهیم بازیابی اطلاعات. ترجمه اسدالله آزاد، رحمت‌الله فتاحی. مشهد: دانشگاه

- فردوسی مشهد، مؤسسه چاپ و انتشارات، ۱۳۷۸.
۲. خالوئی، مرضیه. «نمایه‌سازی ماشینی». نما، دوره ششم، ۳ (بهمن ۱۳۸۵).
۳. کلیوند، دونالد. درآمدی بر نمایه‌سازی و چکیده‌نویسی. ترجمه مهدی حسینی. تهران: چاپار، ۱۳۸۵.
۴. گیلوری، عباس. «نمایه‌سازی خودکار: گذشته، حال، آینده». پیام کتابخانه، دوره دهم، ۴ (زمستان ۱۳۷۹): ۱۵-۲۵.
۵. معتمدی، فاطمه. «نظام نمایه‌سازی «وورداسمیت» در شبکه کتابشناختی (اوس.ال.سی)». اطلاع‌رسانی، دوره هفدهم، ۳ و ۴ (پاییز و زمستان ۱۳۷۵).
6. American Society for Indexing (ASI). [on-line]. Available: [http:// www.asindexing.org](http://www.asindexing.org)
7. Anderson, J. D.; Pérez-Carballo, J. "The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing". *Information Processing & Management*, Vol.37, No.2 (2001): 231 - 254.
8. Ibid. "The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort". *Information Processing & Management*, Vol.37, No.2 (2001): 255 - 277.
9. Bar- Hillel, Y. "The present status of automatic translation of languages". *Advances in Computers*, No.1 (1960): 91 - 163.
10. Baxendale, Phyllis B. "Autoindexing by automatic processes". *Special Libraries*, Vol.56. No.10 (1965): 75 - 79.

and Information Science. New York: Marcel Dekker, 2003.

19. Parry, A. A.; Linford, R. G.; Rich, J. I. "Computer literature searches- a comparison of the performance of Two commercial systems in an interdisciplinary subject". *Information Scientist*, No.8 (1974): 179 - 187.

20. Rehm, G. Towards. "Automatic Web genre identification". Published in the Proceedings of the Hawai'i International Conference on System Sciences, 7-10 January, 2002. [on-line]. Available: <http://www.uni-giessen.de/~g91063/pdf/HICSS35-rehm.pdf>

21. Salton, G. "Automatic indexing using bibliographic citations". *Journal of Documentation*, No.27 (1971): 98. [on-line]. Available: [http://www.garfield.library.upenn.edu/papers/jdoc31\(3\)p216y1975.html](http://www.garfield.library.upenn.edu/papers/jdoc31(3)p216y1975.html)

22. Ibid. *Automatic text processing: Transformation, analysis, and retrieval of information by computer reading*. Mass.: Addison- Wesley, 1989.

23. Sparck-Jones, K. "Statistically-based document indexing". *Skrifter om Anvendt og Matematisk Lingvistik (SAML)*, No. 6 (1980): 79 - 93.

24. Svenonius, E. *The intellectual foundations of information organization*. Cambridge, MA: MIT Press, 2000.

25. Tulic, M. "Book indexing, about indexing & automatic indexing". 2005. [on-line]. Available: <http://www.anindexer.com/about/auto/autoindex.html>

11. Browne, Glenda. "Automatic indexing and abstracting". *Journal of the American Society for Information Society*. 2002. [on-line]. Available: http://www.autoindexing\automatic_indexing_and_abstracting.htm

12. Chen, H. ... [et al]. "Automatic thesaurus generation for an electronic community system". *Journal of the American Society for Information Society*, No.46 (1995): 175 - 193.

13. Collison, Robert L. *Indexing and abstracting*. London: Ernest Benn, 1972.

14. Faraj, N.; Godin, R.; Missaoui, R. "Analysis of an automatic-indexing method based on syntactic analysis of text". *Canadian Journal of Information and Library Science- Revue Canadienne des Sciences de l'Information et de Bibliotheconomie*, Vol.21, No.1 (1996): 1 - 21.

15. Fidel, Raya ... [et al], (edited). *Challenges in indexing electronic text and images*. Medford: Learned Information, 1994.

16. Hjørland, B. "Epistemology and the socio- cognitive perspective in information Science". *Journal of the American Society for Information Science and Technology*, Vol.53, No.4 (2002): 257 - 270.

17. Golub, K. "Automated subject classification of textual web pages, for browsing". Lund: Lund University, Department of Information Technology, 2005. [on-line] Available.: <http://www.it.lth.se/koraljka/Lund/publ/LicE.pdf>

18. Liddy, E. D. "Natural language processing". In *Encyclopedia of Library*