

همارایی شناسه‌ها در نمایه‌سازی

یکی از مباحث مهم در نمایه‌سازی در مرحله تبدیل مفاهیم مدارک به اصطلاحات نمایه، بحث همارایی شناسه‌هاست. منظور از همارایی نحوه و زمان ایجاد پیوند نحوی بین اصطلاحات موجود در شناسه‌هایی است که از چند اصطلاح تشکیل شده‌اند. نظام‌های نمایه‌سازی از نظر قواعد آرایش شناسه‌ها، زمان و عامل هم‌آراکننده به دو دسته کلی پیش‌همارا و پس‌همارا تقسیم می‌شوند.

۱. نظام‌های نمایه‌سازی پیش‌همارا. در این نظام‌ها برقراری رابطه نحوی بین اصطلاحات، قبل از ذخیره‌سازی و توسط نمایه‌ساز صورت می‌گیرد. تعیین موضوع در قالب رشته‌ای از کلمات که با ترکیب نحوی معینی پشت سر هم قرار گرفته، بیان می‌شود و اگر جست‌وجوگر در همان قالب به دنبال موضوع بگردد قادر به بازیابی خواهد بود. از آنجا که زمینه‌های عدم تطبیق بین الگوی پیش‌همارای نمایه‌ساز از موضوع و الگوی کاربر برای یافتن اطلاعات بسیار زیاد است، باید اولاً توصیفگر مرّجّح به‌گونه‌ای انتخاب شود که مورد استفاده تعداد بیشتری از استفاده‌کنندگان باشد، ثانیاً در مواقعی که مفهوم مدرک از متن در قالب چند کلمه به صورت یک عبارت بیان می‌شود، طبعاً فقط یکی از این چند کلمه در ابتدا قرار می‌گیرد و مدخل نمایه می‌شود و سایر کلمات که بعد از آن می‌آیند (= مفاهیم ثانویه) باید امکان مدخل شدن بیابند. ایجاد ارجاعات از مفاهیم ثانویه قاعداً باید دارای الگو باشد، و در آن اصل صرفه‌جویی و جلوگیری از حجیم شدن نمایه رعایت شود.

این ارجاعات به لحاظ ماهیت با ارجاعاتی که از مفاهیم مترادف و وابسته به شناسه مورد نظر صورت می‌گیرد متفاوت است. بنابراین در نمایه‌های پیش‌همارا با دو نوع ارجاع از نظر ماهیت روبه‌رو هستیم: ارجاع از مفاهیم مترادف و وابسته، و ارجاع از اصطلاحات ثانویه در یک شناسه مرکب به آن شناسه. از مهم‌ترین اشکال نمایه‌های پیش‌همارا می‌توان به نمایه‌های انتهای کتاب اشاره کرد.

۲. نظام‌های نمایه‌سازی پس‌همارا. به دلیل مشکلات ذکر شده در نظام‌های نمایه‌سازی پیش‌همارا و مخالفت‌هایی که با همارایی اصطلاحات به دست نمایه‌ساز صورت گرفت، نظام‌های دیگری شکل گرفت که در آن برقرار کردن پیوند بین واژه‌های یک شناسه مرکب، در مرحله بازیابی و توسط جست‌وجوگر صورت می‌گیرد. به علاوه، فراهم شدن امکان استفاده بیشتر از رایانه در ذخیره و بازیابی، موجب رواج بیشتر نظام‌های پس‌همارا شد.

در این نظام موضوع مدرک در قالب تک‌واژه‌ها یا اصطلاحات تک مفهومی بر روی برگه‌ها یا در فایل‌های رایانه‌ای ثبت می‌شود و در مقابل آن جای‌نما می‌آید. جست‌وجوگر می‌تواند موضوع مورد نظر خود را به تک‌واژه‌ها تفکیک و هر واژه را جداگانه از نمایه استخراج کند، سپس جای‌نمای مدارک استخراج شده را با هم مقایسه و مدارک دارای جای‌نمای مشترک را به عنوان نتیجه، بازیابی کند. از

عمده‌ترین مشکلات این نظام‌ها ملاحظات نحوی و معناشناختی است که یکی از مهم‌ترین آنها مسئله تقدم و تأخر اصطلاحات در هنگام همارایی است. مثلاً به هنگام جست‌وجوی مدرکی در مورد "فلسفه علم"، ممکن است همه مدارک مربوط به "علم و فلسفه" و "علم فلسفه" هم بازیابی شود که مرتبط با موضوع نیست. البته برای رفع این مشکل پیشنهادهایی در مورد کدگذاری روی تک‌واژه‌ها و یا تعیین نقش و ربط آنها ارائه شده است. بر اساس این پیشنهادات باید دستورالعملی خاص تهیه شود و در اختیار نمایه‌ساز قرار گیرد که در سرتاسر کار به‌طور یکدست از آن استفاده کند. از مهم‌ترین انواع قدیمی نظام‌های پس‌همارا، نظام‌های پس‌همارای مبتنی بر برگه (کارت) است که خود به دو دسته بزرگ تقسیم می‌شوند. در یک دسته در بالای هر کارت یا برگه یک توصیفگر و زیر آن جای‌نماها (= شماره بازیابی مدارک نمایه شده) می‌آید. این نوع برگه‌ها را می‌توان کارت اصطلاح" نامید که گونه‌های رایج آن برگه‌های ده ستونی و نظام‌های تطابق نوری هستند. در دسته دوم کل اطلاعات یک مدرک بر روی کارت نوشته می‌شود، مانند برگه‌های لبه‌منگنه. این نوع برگه‌ها را می‌توان "کارت مدرک" نامید.

بیشتر موتورهای جست‌وجوی وب بر نمایه‌های واژه‌ای از نوع پس‌همارا متکی هستند.

منطق جست‌وجو در نظام‌های پس‌همارا

از آنجاکه در نمایه‌های پس‌همارا ترکیب کردن واژه‌ها در مرحله جست‌وجو انجام می‌شود، به‌منظور صورت‌بندی دقیق‌تر عبارت جست‌وجو، از الگوهایی برای همارایی توصیفگرها استفاده می‌شود که به "منطق جست‌وجو در نظام‌های پس‌همارا" معروفند. منطق جست‌وجو در اکثر نمایه‌های پس‌همارا منطق ریاضی جورج بول ریاضی‌دان انگلیسی است. عملگرهای منطق بولی عبارتند از: AND (و)، OR (یا)، و NOT (نه). از عملگر AND وقتی برای همارایی استفاده می‌شود که قصد یافتن مدارکی باشد که در آنها همه توصیفگرها با هم آمده باشد، این عملگر با‌اخص کردن جست‌وجو سبب محدود شدن مدارک بازیابی شده می‌شود. از عملگر OR زمانی استفاده می‌شود که قصد یافتن مدارکی باشد که در آنها یکی از توصیفگرهای موردنظر وجود داشته باشد، استفاده از این عملگر سبب گسترش دامنه بازیابی می‌شود. از عملگر NOT وقتی استفاده می‌شود که هدف بازیابی مدارکی است که در آن، توصیفگر موردنظر وجود دارد، اما توصیفگری که به‌دنبال NOT می‌آید وجود ندارد. از منطق جست‌وجوی بولی، خصوصاً در نظام‌های رایانه‌ای بازیابی اطلاعات به‌صورت مستقیم و یا تلویحی استفاده می‌شود. گزینه "دربدارنده همه این کلمات" در برنامه‌های جست‌وجوی رایانه‌ای معادل عملگر AND و گزینه "هر یک از این کلمات"، معادل عملگر OR و گزینه "نباید دربرداشته باشد" معادل عملگر NOT است.

کنترل واژگان و زبان‌های نمایه‌سازی

همان‌گونه که در بحث فرایند نمایه‌سازی مفهومی گفته شد، پس از استخراج مفاهیم مدرک، این مفاهیم باید در قالب واژه‌هایی بیان شود که با پرسش کاوشگر منطبق باشد. مسئله انطباق واژه‌های انتخاب شده توسط نمایه‌ساز با واژه‌های کاربر هنگام جست‌وجوی اطلاعات در نمایه، امری مهم است. پژوهش‌ها نشان می‌دهد که نحوه استفاده از زبان توسط نمایه‌سازان و استفاده‌کنندگان بسیار متفاوت است، به نحوی که میزان توافق در مورد به‌کارگیری واژه‌های ناظر به یک مفهوم را بین نمایه‌سازان و کاربران تنها ۲۵ درصد می‌دانند. برخی از این تفاوت‌ها ناشی از تفاوت در انتخاب واژه‌ها و بقیه ناشی از اختلاف در درک موضوع یا مفهوم‌سازی است. یک بررسی نشان داده است که برای برآورده کردن نیاز اطلاعاتی ۸۰ درصد کاوشگران پیرامون یک موضوع، گنجاندن ۱۵ کلیدواژه متفاوت ناظر به آن موضوع در نمایه ضروری است. قواعد مربوط به برگرداندن مفاهیم به واژگان قابل استفاده در نمایه را می‌توان زبان نمایه‌سازی نامید.

زبان های نمایه سازی

زبان‌های نمایه‌سازی را می‌توان در سه گروه بررسی کرد:

۱. **نمایه‌سازی به زبان آزاد**. در این نظام، نمایه‌ساز از توصیفگرهایی که خود مناسب می‌داند برای توصیف مفاهیم موجود در مدارک استفاده می‌کند، خواه این توصیفگرها در متن آمده باشد یا نه. اصل انتخاب توصیفگر اخص و اصل وحدت در نمایه‌سازی باید در این زمینه مورد توجه قرار گیرد. منظور از اصل وحدت آن است که در نمایه‌سازی به زبان آزاد باید تدابیری اندیشیده شود که از پراکندگی یک مفهوم ذیل توصیفگرهای متعدد اجتناب شود. در نمایه‌سازی به زبان آزاد ممکن است یک مفهوم با توصیفگرهای متعدد بیان شود که مترادف معنایی واژه‌ها، قدیمی یا جدید بودن آنها، بومی یا وارداتی بودن، علمی یا عمومی بودن، اختصاری یا کامل بودن، مفرد یا جمع بودن آنها، و یا وجود چند اصطلاح ناظر به یک مفهوم می‌تواند از جمله علت‌های آن باشد. در نمایه‌سازی به زبان آزاد سعی می‌شود با اتخاذ تدابیری از جمله انتخاب یک توصیفگر مرجح و گنجاندن ارجاعات از موارد مورد اشاره، مثلاً از مترادفات به توصیفگر مرجح، از پراکنده شدن اطلاعات مربوط به یک مفهوم ذیل مدخل‌های متعدد جلوگیری شود. این کار در واقع اعمال نوعی کنترل واژگانی است که در جریان کار و توسط نمایه‌ساز، بدون استفاده از اصطلاحنامه انجام می‌شود.

۲. **نمایه‌سازی به زبان طبیعی**. در این نظام از واژه‌های به‌کار رفته در خود مدرک به‌عنوان توصیفگر استفاده می‌شود. یعنی مفاهیم استخراج شده فقط با استفاده از واژه‌های به‌کار رفته در متن به توصیفگر تبدیل می‌شوند. این نوع نمایه‌سازی را می‌توان از مصادیق نمایه‌سازی به زبان آزاد هم

تلقى کرد. بنابراین اصول پیش‌گفته در مورد نمایه‌سازی به زبان آزاد در اینجا هم کاربرد دارد، مگر هدف تهیه نمایه‌های واژه‌ای باشد. نمایه‌های کوئیک و کووک از انواع نمایه‌های واژه‌ای به زبان طبیعی هستند.

۳. **نمایه‌سازی به زبان کنترل شده.** در این نوع نمایه‌سازی مفاهیم استخراج شده از متن به منظور انتخاب توصیفگر با فهرست‌های معیار مانند سرعنوان‌های موضوعی و یا اصطلاحنامه‌ها تطبیق داده می‌شود.

تفاوت عمده اصطلاحنامه‌ها با سرعنوان‌های موضوعی در آن است که روابط موجود در اصطلاحنامه از نوع سلسله مراتبی (اعم و اخص)، وابستگی، و یا مترادف است، هر چند در سرعنوان‌های موضوعی هم علاوه بر ارجاعات "نگاه کنید" و "نیز نگاه کنید" واژه‌های اعم و اخص مشخص می‌شوند، اما این موارد ضرورتاً با نگاهی دقیق به روابط صحیح معنایی گنجانده نشده و بیشتر مواردی را شامل می‌شوند که گنجاندن آنها برای کاربر مفید است. ارجاع "نیز نگاه کنید" در این فهرست‌ها می‌آید، اما این ارجاعات از نوع سلسله مراتبی یا وابسته نیستند. تفاوت دیگر این دو ابزار کنترل واژگانی در آنهاست؛ فهرست‌های سرعنوان موضوعی اغلب فهرستی از توصیفگرهای پیش‌همارا هستند، ولی اصطلاحنامه‌ها دارای توصیفگرهای تک مفهومی هستند که می‌تواند در مرحله نمایه‌سازی توسط نمایه‌ساز انتخاب و همارا شود و مدخل قرار گیرد و یا در مرحله بازیابی توسط کاربر، بازیابی و همارا شود.

مزیت استفاده از زبان کنترل شده در نمایه‌سازی، جلوگیری از حذف اطلاعات به دلیل پراکندگی ذیل توصیفگرهای متعدد است. همچنین در این شیوه استفاده‌کننده نیازی به مراجعه به مدخل‌های متعدد برای یافتن اطلاعات مربوط به یک مفهوم واحد را ندارد. نمایه‌سازی واژه‌ای و عدم کنترل واژگان در شیوه‌های نمایه‌سازی خودکار از جمله در بانک‌های اطلاعاتی موتورهای جست‌وجو سبب آشفتگی کاربران در بازیابی اطلاعات می‌شود. بازیابی هزاران مدرک غیرمرتبط و عدم بازیابی مدارک مرتبطی که مفهوم مورد جست‌وجو را با توصیفگر متفاوت ذخیره کرده، نیاز بیشتر به نمایه‌سازی مفهومی و کنترل واژگانی را در این محیط یادآور می‌شود.

با رشد استفاده از فنون نمایه‌سازی ماشینی و افزایش نیاز به جست‌وجوی پایگاه‌های اطلاعاتی متعدد که در آنها زبان‌های کنترل شده متفاوتی به کار گرفته می‌شود، دانشمندان علوم اطلاع‌رسانی تلاش‌هایی را به منظور تهیه اصطلاحنامه‌های جست‌وجو (در مقابل اصطلاحنامه‌های نمایه‌سازی) آغاز کرده‌اند، این نوع اصطلاحنامه‌ها دیگر از فهرست اصطلاحات مرّجّح که در نمایه‌سازی به کار می‌روند تشکیل نمی‌شوند، بلکه بیشتر سعی می‌کنند نمایی از واژگان یک حوزه موضوعی ترسیم کنند و پیوندی بین واژه‌های مختلف ناظر به یک مفهوم و همچنین واژه‌های اعم و اخص و وابسته برقرار

کنند. هدف این تلاش به جای محدود کردن یا کنترل واژگان نمایه‌سازی تسهیل جست‌وجو، فارغ از واژگان به کار رفته در نمایه است.

در دوران معاصر، همچنین، علاقه به استفاده از "انتولوژی" (هستی‌شناسی) ها برای مدیریت کنترل واژگان و همچنین جابه‌جایی مفهومی در عرصه جست‌وجوی الکترونیکی و به‌کارگیری رایانه افزایش یافته است.

آنتولوژی شاخه‌ای مهم از فلسفه است که به شناخت هستی، چیستی/ماهیت می‌پردازد. گزاره انتولوژیک حکمی است که چیستی یک پدیده با آن بیان می‌شود. این واژه توسط پژوهشگران حوزه هوش مصنوعی و مهندسی دانش از فلسفه به عاریت گرفته شده تا تلاش‌هایی را نشان دهد که به‌طور بسیار منظم و با استفاده از رایانه به‌منظور توصیف ویژگی‌های واقعیت و یا "واقعیت جهان" در یک قلمرو کوچک انجام می‌شود. انتولوژی‌ها ویژگی‌های اصطلاحنامه‌ای رده‌بندی شده را دارند. تفاوت عمده آنها با اصطلاحنامه‌ها در این است که اصطلاحنامه‌ها بیشتر برای استفاده مستقیم توسط انسان که نیازمند کمک در زمینه واژگان و دستیابی به اطلاعات است طراحی شده، حال آنکه انتولوژی‌ها برای گنجانده شدن در عملیات نرم‌افزاری طراحی می‌شوند، از این‌رو، قواعد تدوین مقولات و سلسله مراتب در انتولوژی‌ها باید بسیار دقیق‌تر باشد.

انواع نمایه از نظر نحوه تنظیم

- ۱- **نمایه الفبایی Alphabetical index**: نمایه ای که شناسه‌ها در آن بر حسب حروف الفبا تنظیم شده اند. معمولی ترین روش تنظیم نمایه به دلیل راحتی و قابل درک بودن آن، الفبایی است. انواع ترتیب بندی نمایه الفبایی به روشهای کلمه به کلمه و حرف به حرف است.
- ۲- **نمایه رده ای Classified index**: در این نوع تنظیم مدخل‌ها بر اساس رده‌ها یا سرعنوان‌های موضوعی مرتب می‌شوند.
- ۳- **نمایه تاریخی**: این نوع تنظیم بیشتر برای مدارک تاریخی کاربرد دارد. که بر اساس زمان از قدیم تا حال مرتب می‌شوند.

انواع نمایه از نظر شیوه تهیه

۱- نمایه سازی دستی

هنگامی که کلیه فرآیند های نمایه سازی مانند انتخاب توصیفگرها ، تنظیم آنها و غیره توسط نمایه ساز و به صورت دستی انجام می شود نمایه سازی دستی گفته می شود.

معایب:

- وقت زیادی به خود اختصاص می دهد
- از یکدستی برخوردار نیست
- تخصص زیاد شخص نمایه ساز را می طلبد
- مخارج زیاد جهت استفاده از نیروی انسانی متخصص

مزایا:

- انتخاب صحیح توصیفگرها چون یک نمایه ساز متخصص و مجرب به دلیل توانایی عقلانی اش امکان درک صحیح محتوای مدرک را دارد.
- احتیاج به امکانات کم

نکته:

نکته ای که در اینجا باید به آن اشاره داشت اینکه در امر نمایه سازی خودکار باید بین دو اصطلاح نمایه سازی به کمک رایانه و نمایه سازی خودکار تفاوت قائل شد. گرچه در بسیاری از متون تخصصی بین این دو اصطلاح تفاوت قائل نشده اند اما باید گفت که در نمایه سازی خودکار پس از تدوین برنامه رایانه ای و تدارک الگوریتم های خاص ، نیروی انسانی در فرآیند تولید نمایه هیچ دخالتی ندارد. اما در نمایه سازی به کمک رایانه ، نمایه ساز از رایانه برای انجام امور نمایه سازی استفاده می کند. (اولین نمایه های رایانه ای مانند کوئیک و کووک از نوع نمایه سازی به کمک رایانه بوده اند.

به طور کلی کاربرد کامپیوتر در نمایه سازی را به دو دسته تقسیم می کنند: نمایه سازی رایانه ای و نمایه سازی با کمک رایانه.

۲- نمایه سازی ماشینی یا رایانه ای

نمایه سازی ماشینی نوعی نمایه سازی است که در آن با استفاده از الگوریتم رایانه ای، واژه های کلیدی یک مدرک از عنوان یا متن استخراج، و سپس در قالب مدخل های نمایه مرتب و سازماندهی می شوند. نمایه سازی کامپیوتری، رایانه ای ، ماشینی و خودکار همگی با هم مترادفند و بجای یکدیگر به کار می روند.

فرایند کار:

در این نوع نمایه سازی همه امور از انتخاب کلید واژه، شماره گذاری، ترتیب بندی و غیره توسط کامپیوتر انجام می گیرد. کامپیوتر مفاهیم مهم را که بارها تکرار شده اند و جزو کلمات غیر موضوعی زبان نیستند، به علاوه اسمهای اشخاص، مکانها و غیره را به عنوان کلید واژه در نظر می گیرد. کامپیوتر مفاهیم مهم را توسط یک نرم افزار به نام نرم افزار بسامدی استخراج می کند. مثلاً یک مفهوم که در ذیل یک بخش از متن بارها تکرار شده باشند، محتملاً کلید واژه است، مگر آن که جزو واژگان غیر موضوعی باشد. واژگان غیر موضوعی مانند (است، که، را و...) توسط یک فهرست به نام فهرست ایستا یا فهرست کلمات غیر مجاز مشخص و نادیده گرفته می شود. قسمت دیگر نرم افزار یک دادگان حاوی اسمهای خاص است. نرم افزار با استفاده از این دادگان، اسمهای خاص متن را تشخیص می دهد و به عنوان کلید واژه در نظر می گیرد. نمایه سازی که با اصول تحلیل و طراحی سیستم و برنامه سازی کاربردی آشنایی دارند، می توانند در زمینه کامپیوتری سازی نمایه سازی موثر باشند.

نمایه سازی رایانه ای عمدتاً به دوشیوه زیر انجام می شود:

الف) حفظ اصطلاح:

این شیوه توسط دکتر سوزان آرتانندی در سال ۱۹۶۳ به کار گرفته شد. در نمایه سازی به شیوه حفظ اصطلاح یک تعریف عملیاتی از اصطلاحات نمایه ای لازم و ضروری است تا رایانه بتواند راهنمایی برای انتخاب موضوعهای مدرک تهیه کند، علاوه بر این باید برنامه یا مجموعه برنامه های سازگار با این شیوه ساخته شود. در نظام حفظ اصطلاح سیاهه یا فهرستی به رایانه داده می شود که حاوی کلمات کلیدی است و به نظام گفته می شود، هر یک از کلمات این لیست را که در متن بود، حفظ کند و آنرا به عنوان اصطلاح نمایه ای انتخاب کند. در این روش استفاده از اصطلاحات از پیش انتخاب شده راهنمای خوب و مفیدی است و امکان انتخاب واژه های استفاده شده به وسیله مولف مدرک را فراهم می کند. نقطه ضعف این روش این است که تضمینی وجود ندارد تا تمامی مفاهیم موضوعات جدید و مهم نمایه سازی شوند.

ب) حذف اصطلاح:

در شیوه حذف اصطلاح یک سیاهه بازدارنده به رایانه داده می شود که کلمات فهرست را هر کجا که بود حذف کند و بقیه را به عنوان کلید واژه انتخاب نماید. نمایه سازی کوئیک و کواک از نوع حذف اصطلاح هستند. در این شیوه آنچه را که نباید کلیدواژه تلقی شود، حذف می شود.

نمایه سازی با کمک رایانه

در این نظام نخست نمایه ساز با دقت متن را می خواند، شناسه های موضوعی را انتخاب و علامت گذاری می کند و نمایه دستنویس تهیه می کند. آن گاه اپراتور مدخلها را براساس نمایه دستنویس از طریق صفحه کلید وارد رایانه می کند. بعد از ورود ، رایانه مدخلها را بر اساس نظم خاصی مرتب می کند، توصیفگرها را زیر شنا سه ها می برد، جاینامهای هرمدخل را با هم ادغام می کند و بعد یک نمایه الفبایی ارائه می دهد. در این برنامه هر اصطلاح نمایه ای همراه جاینامهای آن وارد می شود اگر توصیفگر داشته باشد، پس از شناسه این توصیفگر با یک ممیز آورده می شود. رایانه نیز پس از بررسی مدخل ها بر اساس برنامه از پیش طراحی شده آن مداخل را تنظیم می کند. این نوع نمایه سازی با دوروش متفاوت انجام می شود:

۱) نمایه سازی در محیط گرافیکی (نمایه سازی درون کاشتی)

امروزه بیشتر نرم افزارها در محیط جی. یو. آی یعنی با رابط گرافیکی کار می کنند. Graphic User Interface (GUI)

کار با این روش به صورت زیر است:

نمایه سازی ویراستار، پس از اتمام صفحه آرایی و مشخص شدن شماره صفحه های کتاب ، صفحه ها را از صفحه نمایش کامپیوتر مرور می کند و کلیدواژه ها یا کلمات و عبارتهای مهم را انتخاب می کند. این عبارت انتخاب شده از نظر سیستم یک کلیدواژه محسوب می گردد. نمایه ساز همزمان با انتخاب کردن این کلید واژه ها باید یک پنجره واژه پرداز مستقل دیگر را باز می کند و مدخلهای ارجاعی را که در مورد برخی مدخلها پیش می آیند به طور دستی تحریر می نماید. نتیجه این کار یک فایل فرعی نیز به دست می آید. نمایه ساز فایل اصلی مدخلها و فایل فرعی را ادغام و توسط یک برنامه ترتیب بندی مناسب الفبایی می کند. نمایه ساز این مداخل الفبایی شده را مطالعه می کند و هرگاه به فرایند تبدیل با کما نیاز باشد فرایند به طور دستی در فایل انجام می دهد سپس برنامه الفبایی کننده شماره صفحه ها را به طور صعودی مرتب می کند. .

۲) نمایه سازی در محیط دستوری (نمایه سازی فرمانی)

این نرم افزار در محیط سی. ال. آی یعنی با رابط دستوری (فرمانی) کار می کنند: Command Line Interface

کار با این روش چنین است:

نمایه ساز پس از اتمام صفحه آرای و مشخص شدن شماره های صفحه های کتاب ، صفحه ها را در صفحه نمایش کامپیوتر مطالعه می کند و کلیدواژه ها را با دونماد کاملا استثنایی چپ و راست که در مشابه آن در متن وجود نداشته باشد، علامت گذاری می کند (مثلا کلیدواژه درون سه گروه قرار می گیرد یا...) که این هم نوعی انتخاب کلیدواژه است. سپس نمایه ساز در همان فایل ارجاعات مورد نظر تحریر می کند. پس از اتمام علامت گذاری با گروه ها و ارجاعات و تبدیلهای یک ابزار نرم افزار می تواند کل نمایه را تهیه و الفبایی کند. اگر نرم افزار دارای امکانات نمایه سازی نباشد ویراستار می تواند یک برنامه بسیار ساده (مثلا به زبان بیسیک بنویسد) که فهرست کلیدواژه های علامت گذاری شده و شماره صفحه های آنها را استخراج کند و یک برنامه یا ابزار ساده دیگر نمایه را ترتیب بندی و مرتب کند. بنابراین نمایه کتاب نمی تواند به صورت کاملا رایانه ای ساخته شود و رایانه فقط قادر است مدخلها را ذخیره و به منظور اصلاحات به سرعت بازبازی کند. پس رایانه نمیتواند جایگزین نمایه ساز شود. برنامه رایانه ای مدخلها را الفبایی می کند و مانع دوباره کاری مدخلها می شود و ارجاعهای دوسویه را کنترل و نمایه را چاپ می کند. روشهای نمایه سازی کتاب با کمک رایانه دارای مزایای سرعت ورود اصطلاحات نمایه ای، ظرفیت ذخیره سازی بالا ، سهولت اصلاح ، حروفچینی و تایپ، ویرایش و چاپ نمایه هستند و از لحاظ اقتصادی مقرون به صرفه می باشند. (۷)

تجهیزات لازم برای نمایه سازی ماشینی:

تمامی روشهای نمایه سازی ماشینی می کوشند مراحل فراهم آوری اطلاعات و سازماندهی توصیفگرهای یک مدرک را به صورت کامل خودکار سازی نمایند. نمایه سازی ماشینی به دو صورت انجام می پذیرد و هر کدام تجهیزات خاص خود را می طلبد:

۱) تجهیزات لازم برای نمایه سازی ماشینی مدرک. برای آنکه توصیفگرها از درون متن استخراج شوند به عبارت دیگر وارونه سازی متن انجام میگیرد. (روش استخراجی) هنگام وارونه سازی متن از میان مدارک فقط واژگان سیاهه بازدارنده استخراج می شوند و کنار گذاشته می شوند. مابقی به عنوان واژه های نشان دهنده محتوای مدرک بصورت مقلوب ذخیره می شوند. بنابراین مراحل لازم برای وارونه سازی به این شرح می باشد که اولاً تولید داده های وارونه شده بوسیله استخراج همه مفاهیم از مدرک با اشاره به شماره مدرک صورت می گیرد و ثانیاً مفاهیمی از داده ها نمایه که در سیاهه بازدارنده نیز وجود دارد، حذف می شوند.

۲) تجهیزات مربوط به مفاهیم . در این روش یک بانک واژگان وجود دارد که نوعی واژه نامه می باشد. (روش استنادی)

روش های نمایه سازی ماشینی

الف. روش های زبان شناختی: این روشها می کوشند با کمک تحلیل های شکل شناسی و ساختار نحوی و معنایی مدرک توصیفگرها را استخراج نمایند.

ب. روش های آماری: مشخص می کند که معنی هر مفهوم منفرد در مدرک با حضور آن در جایگاههای مختلف مدرک ارتباط تنگاتنگ دارد. بنابراین لغات درون متن شمارش می شوند و ارتباط آنها ارزش گذاری می شود. هدف آماری از اطلاعات آن است که لغات دارای بارمعنایی در مدرک به عنوان توصیفگر انتخاب شوند. روش های آماری عملاً برای بالابردن جامعیت به کار گرفته می شوند. درحالیکه روشهای زبان شناختی در جهت بهبود مانعیت کاردارند.

ج. روش های مبتنی بر احتمالات: در این روشها تئوری احتمالات برای مدل سازی ریاضی مراحل بازیابی به کارگرفته می شود. درحالیکه در توزیع آماری اصطلاحات یک مدرک مورد استفاده قرار می گیرد. این روش با عملیات ریاضی مفروضات ساده و مطمئنی را ارائه می دهد. فرض بر آن است که مدارک براساس میزان ربط درهنگام بازیابی مورد ارزیابی قرار می گیرند.

مراحل نمایه سازی کامپیوتری :

ذخیره اطلاعات: درنمایه سازی ماشینی اساس کار ذخیره اطلاعات به صورت الکترونیکی و تشخیص اطلاعات بسته به توانایی سیستم جهت این تشخیص و سپس تفکیک به پاره های اطلاعاتی توسط کامپیوتر می باشد. منظور از ذخیره اطلاعات به صورت الکترونیکی ، ذخیره اطلاعات بخشی از مدرک است که جهت نمایه سازی از آن استفاده می شود و شامل عنوان یا چکیده یا فهرست مندرجات یا حتی کل مدرک می باشد.

تشخیص اطلاعات: منظور از تشخیص اطلاعات، تشخیص کلمات و جملاتی است که به صورت الکترونیکی ذخیره شده است و بستگی به توانایی سیستم از لحاظ سخت افزاری و نرم افزاری دارد.

تفکیک اطلاعات به پاره های اطلاعاتی: منظور از تفکیک اطلاعات به پاره های اطلاعاتی این است که کل اطلاعات یک مدرک که ذخیره شده است به جملات یا کلمات شکسته شود.

مقایسه پاره های اطلاعاتی: منظور از مقایسه پاره های اطلاعاتی همان مقایسه جملات یا کلماتی است که در مراحل قبل شکسته شده است.

طبقه بندی: منظور از طبقه بندی پاره های اطلاعاتی ، اولویت بندی پاره های اطلاعاتی برحسب نیاز و با توجه به برنامه ای است که قبلاً به سیستم داده شده است.

انتخاب: انتخاب آخرین مرحله است که پس از اولویت بندی پاره های اطلاعاتی آن دسته از اصطلاحاتی که در اولویت قرارگرفته اند به عنوان توصیفگرانتخاب می شوند و اصطلاحات نامناسب حذف می شوند.