

بسمه تعالی

جزوه ی درس:

نمایه سازی

رشته:

فناوری اطلاعات

تهیه و تدوین:

لیلا سادات هاشمی

بهار ۹۴

مقدمه:

نمایه از مصدر نمودن به معنای نشان دادن و راهنمایی کردن مشتق شده است. معادل آن در زبان انگلیسی واژه Index است که خود از واژه لاتین dicare به معنای نمایاندن و دلالت کردن گرفته شده که با پیشوند In به معنای "اشاره‌گر از بیرون" به کار می‌رود.

در زبان فارسی و در غالب رشته‌های تخصصی، واژه "شاخص" معادل Index به کار می‌رود، ولی در رشته کتابداری و اطلاع‌رسانی واژه "نمایه" برای آن انتخاب شده است. نمایه فهرستی از نام‌ها، موضوع‌ها، و دیگر مطالب موجود در یک سند یا مجموعه اسناد و مدارک است که به شیوه‌ای (معمولاً الفبایی) مرتب شده و از طریق جای‌نماها کاربر را به محلی که این نام یا موضوع در آنجا آمده راهنمایی می‌کند. به فرایند "تحلیل محتوای یک سند و یا مجموعه اسناد و برگردان نتیجه این تحلیل به‌واژه‌هایی که در نمایه می‌آید"، نمایه‌سازی می‌گویند.

از آنجا که نمایه‌سازی شیوه‌های دستیابی به اطلاعات مجموعه‌ای از مدارک یا یک مدرک و یا بخش‌هایی از یک مدرک را، صرف‌نظر از محل نگهداری آن، فراهم می‌کند، می‌توان گفت نمایه‌سازی روش بنیادین سازمان‌دهی دانش است و فهرستنویسی و رده‌بندی مصادیق آن هستند.

تعریف:

طبق استاندارد نمایه‌سازی بریتانیا، نمایه سیاهه نظام یافته مدخل‌هایی است که به منظور کمک به استفاده کنندگان در جایابی اطلاعات یک مدرک ساخته می‌شود.

طبق استاندارد ایزو ۵۹۶۳، عمل توصیف یا شناسایی محتوای موضوعی یک مدرک را نمایه‌سازی گویند.

ثبت و ضبط محتوای اطلاعاتی مدارک با استفاده از روشهای گوناگون به منظور سازماندهی اطلاعات به قصد سهولت بازیابی را نمایه‌سازی گویند (نوروزی، ۱۳۸۰).

به عبارت دیگر، نمایه‌سازی یعنی **تخصیص واژه‌ها یا اصطلاحات به مدارک به منظور توصیف محتوای موضوعی آنها برای بازیابی در مراحل بعد**. هدف نمایه‌سازی اصولاً آماده کردن مدرک برای بازیابی است.

هدف، کارکرد، و ضرورت:

نمایه در زندگی روزمره کاربردهای بسیاری دارد. بدون وجود نمایه‌ها، تلفن کردن، زمان‌بندی قرارهای ملاقات، جایابی مکان‌ها، یافتن خیابان‌ها، یافتن مواد کتابخانه‌ای، و انجام بسیاری از امور دیگر که ما اکنون انجام می‌دهیم به سادگی میسر نیست. نمایه‌ها راهنماهای مفیدی هستند که نه تنها ما را به موضوعات مورد علاقه هدایت می‌کنند، بلکه تصویری یکپارچه و دیدگاهی جامع از یک حوزه موضوعی می‌سازند، کاری که سایر ابزارهای کتابشناختی قادر به تدارک آن نیستند.

هدف نمایه‌ها تسهیل بازیابی اطلاعات است، بنابراین، ارزش هر نمایه به میزان توفیق آن در کمک به جست‌وجوگر برای بازیابی اطلاعات موردنیاز بستگی دارد. ضرورت تهیه نمایه با اندازه، ناهمگونی، و پیچیدگی مدارک موضوع نمایه‌سازی ارتباط مستقیم دارد. یک کتاب ساده و کم حجم، یک شماره از یک نشریه با حجم متوسط، و یا جزوه‌ای شامل ده یا بیست قطعه شعر می‌تواند بدون داشتن نمایه به سادگی جست‌وجو شود؛ اما یک متن درسی و سایر منابع اطلاعاتی غیرداستانی، شماره‌های متعدد یک یا چند نشریه، و گزیده‌ای مفصل از اشعار یا مقالات و مجموعه‌های بزرگ‌تر همه نیازمند نمایه‌اند.

امروزه با رشد روزافزون انتشارات در حوزه‌های مختلف علمی، وجود نمایه‌ها ضروری‌تر شده و پژوهش بدون آنها اساساً دشوار و گاه غیرممکن است. نمایه‌ها با سازماندهی اطلاعات جست‌وجو را تسهیل می‌کنند. اما کار نمایه، بیش از کم کردن تلاش جست‌وجوگر است: نمایه‌ها سبب اطمینان جوینده از جامعیت و مانعیت جست‌وجو می‌شوند؛ بدین معنا که اگر کار نمایه‌سازی درست انجام شده باشد و توصیف‌گرها مناسب انتخاب شده و جای‌نماها دقیق باشند، و ابزارهای کنترل مانند یادداشته‌ها و ارجاعات روشن و کامل باشند، آنگاه پژوهشگر می‌تواند مطمئن باشد که حداکثر بازیابی را با حداکثر دقت، در حداقل زمان انجام داده است. از این‌رو، می‌توان گفت نمایه‌ها نقش مهمی در افزایش بهره‌وری علمی دارند.

ساختار نمایه

نمایه از تعدادی مدخل تشکیل می‌شود، هر مدخل در فایل رکوردهای اطلاعات کتابشناختی یا سایر اطلاعات نظیر نام‌ها، عناوین، موضوعات، و مانند آن یک نقطه بازیابی محسوب می‌شود. اجزای مدخل عبارتند از:

۱. **شناسه** یا توصیف‌گر: کلید واژه مرجحی است که نمایه‌ساز آن را به‌عنوان موضوع یا یکی از موضوعات مدرک انتخاب می‌کند و جزء اصلی مدخل را تشکیل می‌دهد. مدخل‌ها بر اساس شناسه‌ها مرتب می‌شوند. شناسه‌ها انواع متعددی دارند؛ شناسه نام، شناسه موضوع، و شناسه عنوان.

۲. **بیانگر**: کلمه یا عبارتی است که جنبه خاصی از شناسه را نشان می‌دهد. بیانگرها می‌توانند انواعی چون بیانگرهای زمانی، مکانی، موضوعی، و جز آن داشته باشند. بعضی از نمایه‌ها فاقد بیانگرند. نحوه آمدن بیانگر ذیل شناسه متفاوت است، گاهی بیانگرها به‌صورت الفبایی ذیل شناسه می‌آیند (برون‌بافتی)، و گاه نیز بر اساس توالی حضور در متن به‌دنبال شناسه اصلی می‌آیند (درون‌بافتی).

۳. **جای‌نما Locator**. عدد یا هر علامت دیگری است که در نمایه در مقابل شناسه‌ها یا بیانگرها می‌آید و جست‌وجوگر را به محل آمدن اطلاعات هدایت می‌کند. البته گاهی در بعضی نمایه‌های الکترونیکی، نمایه فاقد جای‌نمای قابل رؤیت است، اما در فایل اصلی این جای‌نماها وجود دارند. در این‌گونه موارد کاربر با انتخاب شناسه و یا بیانگر مستقیماً به اطلاعات کتابشناختی، چکیده، و جز آن هدایت می‌شود. اینکه جای نما جست‌وجوگر را به کدام واحد اطلاعاتی هدایت کند، به سیاست‌گذاری اولیه در زمینه نوع واحد سندی بستگی دارد.

"واحد سندی" می‌تواند یک یا چند صفحه از مدرک؛ یک یا چند پاراگراف از یک متن؛ اطلاعات کتابشناختی یا چکیده؛ و یا متن کامل باشد.

۴. **ارجاعات:** ارجاع بخشی از مدخل است که استفاده‌کننده را از یک شناسه یا بیانگر به شناسه دیگر راهنمایی می‌کند. ارجاعات در نمایه‌سازی نیز همانند فهرستنویسی انواع مختلفی دارد، از جمله ارجاع "نگاه کنید"، و "نیز نگاه کنید". در نمایه‌سازی از ارجاع "نگاه کنید ذیل" نیز برای هدایت کاربر از یک شناسه به اطلاعات ذیل شناسه دیگر استفاده می‌شود. البته باید از ارجاع کاربر به‌جایی که در آن اطلاعاتی وجود ندارد خودداری شود، این نوع ارجاعات، ارجاع کور نامیده می‌شود. بعضی نمایه‌ها فقط از شناسه و جای‌نما تشکیل شده‌اند و سایر اجزا را ندارند.

انواع نمایه از نظر نوع اطلاعات (محتوا و پوشش)

نمایه از نظر محتوا و نوع اطلاعات به دو گروه عمده تقسیم می‌شوند:

الف. نمایه نامها (names Index): به این نوع نمایه "نمایه اعلام" یا "نامنامه" نیز گفته می‌شود، سیاهه الفبایی نامهای خاص است که در متن منبع از آنها نام برده شده است. نامهای خاص شامل نامهای اشخاص، اماکن، پدیده‌های جغرافیایی، اشیاء، سازمانها و نظایر آن است که هر منبع به تناسب موضوع مورد بحث می‌تواند دربردارنده‌ی تمام یا برخی از آنها باشد. نمایه نامها دارای انواع مختلف است:

۱- **نمایه پدید آور:** (Authors Index): سیاهه‌ای از نام پدید آورندگان آثاری که در متن به آنها استناد شده است.

۲- **نمایه عناوین:** (Title Index): سیاهه‌ای از عنوان آثاری که در متن به آنها استناد شده است.

۳- **نمایه نام های جغرافیایی:** (Geographical Index) سیاهه الفبایی از نام اماکن، مناطق، کشورها و پدیده‌های جغرافیایی

۴- **نمایه نام سازمان ها و مؤسسات:** نام سازمان ها و موسسه‌ها در آثار گوناگون و یا در ادوار مختلف یا حتی در یک منبع واحد ممکن است به شکل‌های مختلفی آورده شود. در این مورد لازم است اسامی سازمانها و مؤسسات با استفاده از منبع معیار به شکلی هماهنگ و یکنواخت در نمایه‌سازی استفاده شود. چون برخی سازمان‌ها در طول زمان دچار تغییرات شده‌اند. (مانند سازمان برنامه‌سپس سازمان برنامه و بودجه سپس سازمان مدیریت و برنامه‌ریزی).

۵- **نمایه اشخاص:** عبارتند از سیاهه الفبایی از نام‌های اشخاصی است که در منبع ذکر شده است. یکی از اهداف نمایه‌سازی نام‌ها آن است که کلیه اطلاعات راجع به یک نام و یا محل و یا حضور آن نام در متن با یک شناسه واحد نشان داده شود تا بازیابی آن سریع و آسان شود.

ب. نمایه موضوعی (Subject index): سیاهه‌ای الفبایی از مفاهیم و موضوع‌های مهمی است که در منبع درباره آن بحث شده است. در این نوع نمایه‌سازی مفاهیم مطرح شده در متن به وسیله‌ی یک اصطلاح بیان می‌شوند. موضوع بحث نمایه‌سازی موضوعی پرداختن به چگونگی توصیف محتوای مدرک است.

انواع نمایه از نظر بیان محتوای اطلاعاتی:

برای توصیف محتوای مدرک دو رهیافت کلی وجود دارد: نمایه‌سازی واژه‌ای؛ و نمایه‌سازی مفهومی

۱- **نمایه‌سازی واژه‌ای.** در این نوع نمایه بر اساس سیاست‌گذاری‌های اولیه واژه‌هایی از تمام یا بخشی از متن انتخاب و موضوع مدرک تلقی می‌شود. واژه‌ها می‌توانند از جای‌جای متن (عنوان یا چکیده و یا سراسر متن) استخراج شوند. این نوع نمایه‌سازی را که در آن واژگان نمایه از متن مدارک استخراج می‌شود، "نمایه‌سازی استخراجی (اشتقاقی)" می‌گویند. در مقابل، شیوه‌ای را که در آن نمایه‌ساز پس از مطالعه مدرک و درک مفاهیم آن، خود واژه‌هایی را به‌عنوان موضوع به مدرک منتسب می‌کند، "نمایه‌سازی انتسابی" گویند.

روال اصلی کار در تهیه نمایه‌های واژه‌ای عنوان، گردش دادن واژه‌های عنوان به گونه‌ای است که هر بار یکی از واژه‌ها در محل الفبایی مناسب، مدخل قرار می‌گیرد. از این نظر به این نمایه‌ها، "نمایه‌های گردان" هم گفته می‌شود. پیش‌فرض تهیه این نوع نمایه‌ها این است که عنوان بیانگر موضوع متن است. برای تهیه رایانه‌ای نمایه‌های واژه‌ای عنوان، ابتدا فهرستی از کلمات غیرمجاز (سیاهه بازدارنده) شامل کلمات فاقد بار اطلاعاتی، حروف اضافه، کلمات ربط، افعال، صفات، و مانند آن، تهیه و به برنامه رایانه‌ای معرفی می‌شود. رایانه این کلمات را از مجموعه متون (مثلا در این مورد عنوان) حذف و بقیه واژه‌ها را موضوع تلقی می‌کند و در قالب‌های تعریف شده، مانند نمایه‌های کوئیک، کوواک، و نمایه جایگزینی ارائه می‌کند.

نمایه کوئیک Keyword in Context

از قدیمی‌ترین و موفق‌ترین نمایه‌های واژه‌ای عنوان نمایه کوئیک است. در این نمایه عناوین گزارش‌ها، مقالات، و مانند آن به‌صورت ماشین‌خوان درمی‌آید و به‌وسیله رایانه تقطیع می‌شود، برنامه رایانه‌ای واژه‌های معنادار را از میان واژه‌ها انتخاب می‌کند و سپس هر عنوان را به گونه‌ای چاپ می‌کند که کلیدواژه در موقعیتی برجسته (معمولا در ستونی در وسط عناوین) قرار گیرد و تمام عناوین بر اساس نظم الفبایی کلیدواژه‌ها و حروف بعد از آن مرتب شوند. این شناسه‌ها با روش‌هایی مانند نوشته شدن به‌صورت ایرانیک (ایتالیک) یا سیاه، و یا فاصله‌گذاری از کلمات قبل و بعد در متن عنوان برجسته می‌شوند تا چشم هنگام بازبایی خسته نشود. از نشانه‌گذاری هم برای مشخص کردن ابتدا و انتهای عنوان استفاده می‌شود. برای مثال مدخل‌های نمایه کوئیک برای یک عنوان مثل "واژه‌نامه‌ها و دایره‌المعارف‌های کتابداری و اطلاع‌رسانی" به‌شکل ذیل در می‌آید.

های علوم کتابداری و	اطلاع‌رسانی#	واژه‌نامه‌ها و دایره‌المعارف‌ها
واژه‌نامه‌ها و	دایره‌المعارف‌های	علوم کتابداری و اطلاع‌رسانی#
واژه‌نامه‌ها و دایره‌المعارف‌های	علوم	کتابداری و اطلاع‌رسانی
دایره‌المعارف‌های علوم	کتابداری	و اطلاع‌رسانی# واژه‌نامه‌ها
کتابداری و اطلاع‌رسانی#	واژه‌نامه‌ها	دایره‌المعارف‌های علوم

شکل ۱- نمونه‌ای از مدخل‌های یک نمایه کوئیک

نمایه کواک Keyword Out Of Context

نمایه کواک همان نمایه برون بافتی است. این نمایه برای رهایی از مسائل و مشکلات ناشی از ضرورت کوتاه کردن عنوان و نیز به منظور ساده ساختن خواندن مدخلها طراحی شده است. در این نمایه هر کلیدواژه به ترتیب از عنوان خود خارج شده و مقدم بر سایر اجزای عنوان قرار می‌گیرد. سپس عنوان مدرک به ترتیب طبیعی خود و به طور کامل در زیر این واژه یا به دنبال آن می‌آید. به این ترتیب برای هر واژه مهم یک مدخل ساخته می‌شود. در این نمایه نیز کلمات مجاز در متن عنوان وجود دارد و کلید واژه قرار نمی‌گیرد.

اطلاع‌رسانی

واژه‌نامه‌ها و دایره‌المعارف‌های علوم کتابداری و اطلاع‌رسانی

دایره‌المعارف‌ها

واژه‌نامه‌ها و دایره‌المعارف‌های علوم کتابداری و اطلاع‌رسانی

علوم

واژه‌نامه‌ها و دایره‌المعارف‌های علوم کتابداری و اطلاع‌رسانی

کتابداری

واژه‌نامه‌ها و دایره‌المعارف‌های علوم کتابداری و اطلاع‌رسانی

واژه‌نامه‌ها

واژه‌نامه‌ها و دایره‌المعارف‌های علوم کتابداری و اطلاع‌رسانی

شکل ۲ - نمونه‌ای از مدخل‌های یک نمایه کواک

از مزایای نمایه‌های واژه‌ای می‌توان به امکان تهیه سریع به کمک رایانه، عدم نیاز به تصمیم‌گیری موردی، یکدستی کار و سلیقه‌ای نبودن، ارزانی، طبیعی بودن زبان نمایه‌سازی، راحتی تهیه در هم‌کرد و وجود نوعی هم‌ارایی به دلیل حفظ بافت طبیعی عنوان اشاره کرد. عیب این نوع نمایه‌سازی آن است که همیشه عنوان، گویای موضوع مدرک نیست؛ عیب دیگر، نبود کنترل واژگانی و پراکندگی اطلاعات یک موضوع ذیل مدخل‌های مترادف و شبه مترادف، بدون وجود ارجاعات لازم است؛ و عیب سوم تلقی موضوعات هم املا به عنوان موضوع واحد است. به هر حال میزان مفید بودن این نمایه‌ها به واقع نما بودن و روشنی و معنا دار بودن واژه‌های عنوان بستگی دارد. نمایه‌سازی واژه‌ای معمولاً برای موادی نظیر کتاب و مقاله کاربرد دارد و به کار نمایه‌سازی مواد دیداری، اشیای هنری، و امثال آن نمی‌آید. نمایه‌سازی انواع اخیر معمولاً از نوع مفهومی است.

۲- **نمایه‌سازی مفهومی.** نمایه‌سازی مفهومی فرایندی پیچیده است و تشریح آن به سادگی میسر نیست. شیوه‌ای که توسط انسان به کار گرفته می‌شود تا موضوعات، ویژگی‌ها، معانی، و یا کاربردهای ممکن متون زبانی، تصاویر، و سایر

انواع پیام‌ها را درک کند و بشناسد، ریشه در فرایند شناختی ذهن و فکر انسان دارد. این فرایند به خوبی شناخته نشده و در عرصه‌هایی چون روان‌شناسی شناخت، علوم‌شناختی، و هوش مصنوعی موضوعی مهم به حساب می‌آید. همین اندازه می‌دانیم که این فرایند حداقل از دو مرحله تشکیل شده، شناخت موضوع و ویژگی‌های آن، و توصیف آنها در قالب واژگانی مناسب برای بازیابی. فرایند مذکور را می‌توان به مراحل جزئی‌تر ذیل تفکیک کرد:

- **تدوین سیاست‌ها و راهبردهای کلی.** قبل از شروع به کار، نمایه‌ساز باید با نگاه به تمامی مجموعه‌ای که باید نمایه شود، در مورد این مسائل تصمیم‌گیری کند: نوع نمایه (به‌طور مثال دستی یا ماشینی بودن یا نحوه انتشار و ارائه آن)، ساختار مدخل‌ها (مثلاً استفاده یا عدم استفاده از بیانگرها و شکل آنها)، میزان اطلاعاتی که قابلیت مدخل شدن را ایجاد می‌کند (عمق نمایه)، واحد سندی‌ای که جای‌نماها به آن رهنمون می‌شوند (پاراگراف، صفحه، کل مدرک، و جز آن)، نحوه کنترل واژگانی، نحوه همارایی، مسئله ارجاعات، و سایر سیاست‌ها که در تدوین نمایه ضروری است.

سه عامل مهم در همه این تصمیم‌گیری‌ها مؤثر است: اول ماهیت مجموعه یا نوع موادی که باید نمایه شود، به‌طور مثال تجانس یا عدم تجانس آنها، فنی یا غیرفنی بودن منابع، یک زبانه یا چندزبانه بودن آنها؛ دوم ویژگی‌های استفاده‌کنندگان است، نمایه‌ساز بداند که آیا استفاده‌کنندگان از مجموعه و نمایه، گروهی متجانس هستند یا نه، جست‌وجوگران موردی‌اند یا پژوهشگران حرفه‌ای، کم سوادند یا تحصیل‌کرده، متخصص هستند یا غیرمتخصص؛ و سوم شرایط و محیط فیزیکی است، مثلاً تعیین اینکه آیا نمایه به‌صورت مستقیم و به همراه منبع اصلی (مثل نمایه انتهای کتاب) و یا جداگانه منتشر می‌شود، و یا اینکه آیا نمایه در دست تهیه مستقل و خودکفاست یا بخشی از یک نمایه دنباله‌دار است که بعداً با سایر قسمت‌ها در هم‌مکرد می‌شود، و آیا به‌صورت دستی ذخیره و بازیابی می‌شود یا به‌صورت ماشینی. همچنین قبل از شروع باید درباره آنچه که نباید نمایه شود (مانند تبلیغات، یادداشت‌ها، و منابع و مآخذ)؛ یا آنچه که می‌تواند به‌صورت گذرا و سطحی نمایه شود (مانند نامه‌ها، یادداشت‌های کوتاه، پانویس‌ها، و مانند آن)؛ و آنچه که در مورد آن نیاز به استفاده از واژگان فنی تخصصی و یا واژگان سایر زبان‌ها وجود دارد، تصمیم‌گیری شود. پس از پاسخگویی به چنین سؤالاتی، طراحی کلی و معیارهای مهم ساخت نمایه تدوین شده و می‌توان کار تهیه نمایه را آغاز کرد. بهتر است اگر نمایه به‌صورت نمایش دادنی منتشر می‌شود، این سیاست‌ها در مقدمه نمایه بیاید.

- **آشنایی و شناخت.** نمایه‌ساز باید تصویر روشنی از محتوای اطلاعاتی مدرک به‌دست آورد. به این منظور مطالعه عنوان، رئوس مطالب، چکیده، نتیجه‌گیری، پیشنهادات، و یا تمام متن (بسته به عمق نمایه) ضروری است. از سوی دیگر نمایه‌ساز باید با نیازهای جامعه استفاده‌کننده (بالقوه و بالفعل) آشنا باشد تا بتواند تمامی راه‌های دستیابی موضوعی به مدرک را در نمایه پیش‌بینی کند. داشتن دید هدفمند اما بی‌طرف، به همراه توجه به نیازهای احتمالی جامعه، برای نمایه‌سازی ضروری است.

از این منظر نمایه‌ساز باید توجه کند که استفاده‌کننده از نمایه ممکن است با مخاطبانی که اصل اثر برای آنها نگاشته شده کاملاً متفاوت باشد، مثلاً اگر قطعاتی از یک متن فلسفی به‌عنوان گزیده نثر در یک کتاب آیین‌نگارش گنجانیده شده، در این حالت نمایه‌ساز متن را از بُعد آموزش متن نیز مورد توجه قرار می‌دهد و سعی در نمایش این ویژگی‌ها در نمایه دارد.

- **تحلیل محتوا و انتخاب اولیه مفاهیم.** ارائه تعریفی جامع از "مفهوم" در نمایه‌سازی مفهومی، مشکل است. در تحلیل محتوا شاید بتوان چکیده یک فکر یا خمیر مایه موجود در مدرک و یا روابط موجود میان اشیا یا رخدادها و سایر پدیده‌ها را "مفهوم" نامید و یا ممکن است به دسته‌ها و یا رده‌هایی که این پدیده‌ها به آنها تعلق دارند مفهوم اطلاق کرد. در این مرحله نمایه‌ساز باید در مورد موضوعاتی تصمیم‌گیری کند که از نظر نمایه‌سازی دارای اهمیت است. می‌توان اطلاعات هر مدرک را گروه‌بندی و هر گروه را در قالب یک عبارت کوتاه یادداشت کرد، تقریباً همه صاحب‌نظرانی که درباره تحلیل محتوای مدرک سخن گفته‌اند، بر اهمیت نقش تجربه در این امر تأکید دارند. طبیعی است که موضوعات اصلی باید مدخل قرار گیرند، اما در مورد موضوعات فرعی و اینکه تا چه حد در نمایه حضور یابند باید تصمیم‌گیری کرد. بعضی از مراکز نمایه‌سازی، سیاست خاصی در مورد نحوه انتخاب موضوع، تدوین و به نمایه‌سازان ارائه می‌کنند، مثلاً به نمایه‌سازان "چکیده‌نامه شیمی" گفته شده هر نوع روش اندازه‌گیری، ابزار کار، پیشنهاد و نظریه جدید، و تمام ترکیبات جدید شیمیایی که در متن آمده باید نمایه شود.

استخراج مفاهیم و واژه‌ها از یک متن و انتساب آن به مدرک بر مبنای اصول معینی صورت می‌گیرد: **الف) نیاز استفاده‌کننده،** یعنی آشنایی نمایه‌ساز با نیازهای اطلاعاتی استفاده‌کننده بالفعل و بالقوه و همچنین واژگان رایج در حوزه مورد نمایه‌سازی حائز اهمیت است؛ **ب) حفظ جامعیت نمایه،** بدین معنا که در نمایه‌سازی سعی بر آن است که تمام مفاهیم اصلی موجود در مدرک که دارای ارزش اطلاعاتی است به‌عنوان شناسه در نمایه گنجانده شود. انتخاب مفاهیم فرعی که در مدرک به‌طور جانبی به آن پرداخته شده، سبب حجیم شدن نمایه و انحراف کاربر از مفاهیم اصلی است. در این زمینه توجه به عمق نمایه که در سیاست‌گذاری‌های اولیه انجام می‌شود می‌تواند راهگشا باشد؛ و **ج) اخص بودن کلیدواژه،** بدین معنا که واژه‌هایی که به‌عنوان توصیفگر انتخاب می‌شوند باید اخص‌ترین واژه‌های ممکن برای توصیف موضوع باشد، مثلاً اگر مدرکی راجع به "مدارس ابتدایی" است، اختصاص توصیفگر کلی مدارس به آن خلاف این اصل است. البته با اخص شدن موضوع، میزان بازیابی کاهش می‌یابد، با این حال استانداردهای نمایه‌سازی به‌کارگیری واژه‌های اخص را به‌عنوان توصیفگر توصیه می‌کند، چرا که اگر کنترل واژگانی به خوبی صورت گرفته باشد، در صورت نیاز امکان گسترش و عمومی کردن جست‌وجو وجود دارد. در نظام‌های نمایه‌سازی واژه‌ای، مانند بیشتر شیوه‌های نمایه‌سازی خودکار که از کلیدواژه‌های متن استفاده می‌کنند، امکان انتخابی در این زمینه وجود ندارد، اما نظام‌های نمایه‌سازی که ضمن تحلیل موضوع، کلیدواژه به آن منتسب می‌کنند، باید این اصل را مدنظر داشته باشند.

- **ارزیابی و انتخاب نهایی مفاهیم.** در این مرحله نمایه‌ساز سعی می‌کند همه عناصری را که ممکن است استفاده‌کنندگان مایل باشند در جست‌وجو بازیابی کنند، از میان مفاهیم انتخاب شده در مرحله قبل گزینش کند و موارد بی‌ربط یا کم ربط از نظر اطلاعات را حذف کند. در این مرحله نمایه‌ساز تصمیم می‌گیرد که یک پاره اطلاع تا چه حد می‌تواند برای کاربر مفید باشد و حذف آن تا چه میزان از جامعیت نمایه می‌کاهد.

تصمیم‌گیری در این مورد غالباً کاری شاق است و ممکن است در سطوح مختلف کار بارها و بارها تکرار شود، نمایه‌ساز در این مرحله در سطحی خاص‌تر تصمیم‌گیری می‌کند، او ضمن حرکت از بخشی از مجموعه به بخشی دیگر، از موردی به مورد دیگر، و یا از پاراگرافی به پاراگراف دیگر باید تصمیم بگیرد که اصولاً مطلبی در نمایه بیاید یا نه و اگر بیاید در قالب چه توصیفگری.

مسئله عمق نمایه‌سازی، میزان ریز شدن و اخص بودن پیام‌هایی که باید در نمایه بیاید، نوع توصیفگر با توجه به سیاست‌گذاری‌های اولیه، و تعداد واژه‌هایی که هر مطلب باید با آن بیان شود، در اینجا مورد توجه قرار می‌گیرد. نمایه‌ساز برای تصمیم‌گیری صحیح باید همواره توجه کند که اولاً کاربر برای یافتن مطلبی خاص احتمالاً از چه واژه‌هایی استفاده خواهد کرد؛ ثانیاً آیا هنگامی که کاربر ذیل یک مدخل منتخب از نمایه به دنبال اطلاعات موردنظر می‌گردد، آیا اطلاعات موردنیاز برایش در نمایه گنجانده شده است. توجه به جدید و مفید بودن اطلاع برای گروه‌های بزرگی از استفاده‌کنندگان، و تخصصی یا عمومی بودن آن در حوزه مورد نمایه‌سازی، از ملاک‌های دیگر ارزیابی و تصمیم‌گیری است.

- **تبدیل مفاهیم انتخاب شده به اصطلاحات و توصیفگرهای نمایه و گنجاندن جای نماها.**
پس از انتخاب نهایی مفاهیم، برای تبدیل این مفاهیم به توصیفگرهای نمایه مسائل متعددی مطرح است که عمده‌ترین آنها عبارتند از: الف) نحوه کنترل واژگانی به‌منظور رعایت اصل وحدت [۲۶] در نمایه‌سازی؛ و ب) نحوه هم‌ارایی شناسه‌ها یا نظام‌های پیش‌همارا و پس‌همارا، که هر یک در ادامه مورد بحث قرار خواهد گرفت.

- **سایر مراحل.** در اینجا سه مرحله باید طی شود: الف) مرتب کردن مدخل‌های نمایه بر اساس نظام پیش‌بینی شده مثلاً به‌صورت الفبایی، تاریخی، رده‌ای، و یا هر نظم شناخته‌شده دیگر؛ ب) مقایسه و ویرایش مدخل‌ها و مشخص کردن روابط میان آنها؛ و ج) مرتب کردن اطلاعات توصیفی و برقراری ارتباط لازم بین بخش توصیفی و تحلیلی (در نمایه‌هایی که مستقل از متن اصلی تهیه می‌شوند).

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 مرجح 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 مرجح 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**: نمایه‌ای که شناسه‌ها در آن بر حسب حروف الفبا تنظیم شده‌اند. معمولی‌ترین روش تنظیم نمایه به دلیل راحتی و قابل درک بودن آن، الفبایی است. انواع ترتیب‌بندی نمایه الفبایی به روشهای کلمه به کلمه و حرف به حرف است.
- ۲- **نمایه رده‌ای 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 و 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 بار معنایی در مدرک به عنوان توصیفگر انتخاب شوند. روش های آماری عملاً برای بالابردن جامعیت به کار گرفته می شوند. در حالیکه روشهای زبان شناختی در جهت بهبود مانعیت کار دارند.

ج. روش های مبتنی بر احتمالات: در این روشها تئوری احتمالات برای مدل سازی ریاضی مراحل بازیابی به کار گرفته می شود. در حالیکه 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 بر اساس میزان ربط در هنگام بازیابی مورد ارزیابی قرار می گیرند.

مراحل نمایه سازی کامپیوتری :

ذخیره اطلاعات: در نمایه 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 بر حسب 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 قرار گرفته اند به عنوان توصیفگر انتخاب می شوند و اصطلاحات نامناسب حذف می شوند.

معرفی چند نرم افزار نمایه سازی:

۱- نرم افزار . RoboHelp HTML Edition

این نرم افزار از سوی شرکت Blue Sky software تولید شده است این نرم افزار جزء قدرتمندترین نرم افزارهای تولید help برای نرم افزارها و وبسایتها و شبکه های داخلی (اینترانتها) است.

۲- نرم افزار Macrex این نرم افزار از سوی شرکتی به همین نام تولید شده 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 میشود. آخرین ویرایش آن ویرایش یا سری هشت است که بر روی انواع سیستمهای عامل ویندوز، ویستا، و مکینتاش قابل نصب و اجراست.

۳- . نرم افزار Cindex این نرم افزار نمایه ساز ازسوی شرکت روچستر و از طریق نشانی <http://www.indexres.com> قابل دسترسی میباشد. از این نرم افزار بیشتر برای نمایه سازی کتابها، روزنامه ها، و سایر نشریات ادواری استفاده می شود. این نرم افزار به عنوان یک ابزار حرفه ای برای تولید انواع نمایه است. به وسیله این نرم افزار می توان واژه نامه، نمایه های موضوعی، نویسندگان، و مانند آن را براساس نمایه تولید شده اولیه ساخت. این نرم افزار قابلیت نمایش نمایه تولید شده در محیط های پردازشگر کلمه مانند ورد را داراست که قابلیت غلط گیری و رفع اشکال از نمایه تولید شده را میسر میسازد. در حقیقت این نرم افزار مانند یک نرم افزار پردازشگر کلمه کار میکند. 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.

۴- نرم افزار Retriever دراصل این نرم افزار یکی از ابزارهای نمایه سازی تصاویر است. این نرم افزار از سوی شرکتی آلمانی با نام Dirk Djuge تولید می شود، این نرم افزار در عین کارآمدی و وجود امکانات بسیار، دارای محیطی بسیار ساده است.

۵- INDEXING RESEARCH محصول شرکت CINDEX

۶- INDEX AID محصول شرکت Santa Barbara Software Products

۷- INQUIRY محصول شرکت Indexer assistant

نمایه سازی در محیط وب:

وب به عنوان یکی از جذاب ترین بخش های اینترنت کاربردهای فراوانی دارد و مجموعه ای است از صفحات به هم پیوسته که حاوی اطلاعات مفیدی در زمینه های موضوعی متفاوت است. همه دلایل مربوط به چرایی سازماندهی اطلاعات در محیط چایی با شدت بیشتری در محیط الکترونیک صادق است. چرا که امروزه میلیونها نفر نیازهای خود را از طریق وب مرتفع می سازند. در حال حاضر موتورهای کاوش، وب را تحت ضابطه درآورنده اند و با نمایه سازی صفحات، پاسخی برای پرس وجوی کاربران فراهم می آورند. نحوه سازماندهی اطلاعات بازتاب مستقیمی در نحوه بازیابی (رتبه بندی و نمایش) نتایج دارد. موتورهای کاوش هر یک با الگوریتمها و سیاستهای متفاوتی به مقوله نمایه سازی می نگرند. تفاوت در نتایج بازیابی شده در موتورهای کاوش مختلف نشان بارزی از وجود تفاوت در الگوریتمهای نمایه سازی آنهاست. سازماندهی و سپس ذخیره مناسب اطلاعات به بازیابی نتایج مرتبط و مناسب خواهد انجامید.

جمع آوری اطلاعات در اینترنت به دو گروه اصلی تقسیم میشود:

۱- **موتورهای جستجوگر یا پیمایشی (Search Engine):** این موتورها برای فهرست میلیونها و گاهی بلیونها

صفحه ی وب از نرم افزاری به نام عنکبوت ها استفاده میکنند.

اجزاء و شیوه کار موتورهای جستجوگر

- **عنکبوت یا روبات خزنده** و یا اصطلاحاً روبات نرم افزاری: بخشی از موتور جستجو است که از صفحات وب بازدید می کند، آنها را بررسی می کند و سپس پیوند های آن صفحه را تعقیب می کند. در نتیجه به واسطه همین تعقیب پیوندها، صفحات وب توسط موتور های جستجو شناسایی می شوند. از طرفی صفحاتی که هیچ پیوند ورودی نداشته باشند توسط موتور های جستجو شناسایی نمی شوند و بخشی از وب نامرئی (Invisible Web) را تشکیل می دهند.

- **نمایه یا پایگاه اطلاعاتی:** نمایه مانند کتاب بزرگی است که کپی کاملی از تمام صفحاتی که توسط عنکبوت یافت می شود در آن ذخیره می شود. در فرایند نمایه سازی، برخی موتورهای جستجو نظام نمایه سازی تمام متن دارند و هر واژه موجود در متن به جز واژه های فاقد بار اطلاعاتی مانند حروف تعریف، ربط، و اضافه را نمایه می کنند. برخی دیگر بخشی از منبع یا مدرک را نمایه می کنند. بعضی از موتور های جستجو، سر عنوان ها، عناوین فرعی، فرآیندها به سایت های دیگر را به همراه ۲۰ خط ابتدایی متن و ۱۰۰ کلمه ای که از بسامد بالایی برخوردار است، نمایه می کنند. در نتیجه قابلیت های هر موتور جستجو بسته به نوع نمایه سازی آن موتور جستجو متفاوت است.

- **نرم افزار جستجو:** نرم افزار جستجو برنامه ای است که در خواست یک جستجو را پردازش می کند تا یک جستجو از میان داده های ذخیره شده (در نمایه) بازیابی شود و اطلاعات یافته شده را گزارش نماید.

عوامل مورد استفاده در موتورهای جستجو برای رتبه بندی سایت ها

هنگامی که برای جستجوی موضوعی خاص، از موتورهای جستجو استفاده می کنید، تمام موتورهای جستجو سعی می کنند تا لیستی از صفحات وب که بیشترین مطابقت را با جستجوی شما دارد، ارائه دهند. لیستی که موتورهای جستجو ارائه می دهند به ترتیب از بیشترین ارتباط به کمترین ارتباط با واژه ی مورد جستجو مرتب می گردد. البته موتورهای جستجو همیشه به کاربران جواب های کاملا مرتبط را نمی دهند. گاهی اتفاق می افتد که صفحات نامرتب در لیست نتایج دیده می شود، این امر باعث می شود که وقت کاربر در مرور نتایج مرتبط تلف شود. هر موتور جستجویی که میزان صفحات مرتبط بیشتر و نامرتب کمتری را بازایی کند بهتر است، برای دستیابی به این منظور و تعیین ربط وب سایت ها، موتورهای جستجو از مجموعه ای از قواعد بهره می گیرند که به آن الگوریتم می گویند. هر موتور جستجو الگوریتم خاص خود را دارد. در نتیجه یک سایت ممکن است در موتور های جستجوی مختلف، رتبه ی متفاوتی کسب کند.

برخی از مهمترین معیارها که در رتبه بندی اغلب موتور های جستجو مشترک است :

مکان کلید واژه

این که کلمه یا عبارتی که مورد جستجو قرار می گیرد در کجای صفحه ی وب قرار داشته باشد یکی از مهمترین معیارها در رتبه بندی سایت است. هر کدام از مکان های صفحه ی وب ارزش وزنی بخصوصی دارند. برای مثال کلمه ای که در عنوان واقع شده است از ارزش بیشتری نسبت به کلمه ای که در متن است، دارد.

بسامد واژگان

فاکتور مهم دیگر در تعیین رتبه سایت تعداد تکرار واژگان (بسامد واژگان) در صفحه وب است. در نتیجه هر چه بسامد واژه یا عبارات مورد جستجو بیشتر باشد رتبه آن صفحه در هنگام بازیابی بالاتر است.

چند نمونه از موتورهای پیمایشی که ترافیک اصلی جست و جو بر روی آن ها قرار دارد:

Google (www.Google.com)

All the Web (www.alltheweb.com)

AltaVista (www.altavista.com)

Yahoo (www.Yahoo.Com)

MSN (www.msn.Com)

Ask Jeeves (www.Askjeeves.Com)

۲- موتورهای راهنما یا Web Directory

این موتورها تنها سرفصلها و عناوین موضوعات را جستجو میکند مانند یاهو. این جستجو، شبیه جستجو در فهرست یک کتاب است.

موتورهای دایرکتوری یک تفاوت اساسی با موتورهای جستجوی پیمایشی دارند و آن به کارگیری عنصر انسانی بجای نرم افزارهای عنکبوت، در جمع آوری، ذخیره و نگهداری اطلاعات می باشد. راهنماها توسط افراد متخصص خلق و نگهداری میشوند و در حالیکه موتورهای جستجو نمایه سازی را به صورت خودکار و توسط نرم افزارهای روبات یا همان عنکبوت انجام میدهند.

(LookSmart (<http://www.looksmart.Com>)

(Open Directory Project (<http://dmoz.Org>)

(Yahoo! (<http://www.yahoo.Com>)

(Google Directory (www.google.Com)

به طور کلی همه موتورهای کاوش، فرایند جستجو و بازیابی اطلاعات را از طریق برقراری ارتباط میان سه جزء اصلی تشکیل دهنده خود یعنی رباتها (عنکبوت ها)، پایگاه اطلاعاتی و نرم افزار بازیابی اطلاعات که کاربر از طریق رابط کاربری به آن مرتبط می شود انجام می دهند.

موتورهای جستجو جهت کاهش زمان جستجو، بخش های بسیاری از کار خود را از قبل انجام می دهند و پایگاه داده خود را آماده می کنند، تا در لحظه ای که کاربر عبارتی را جست و جو می کند، عمل جست و جو را که شامل مقایسه و تطبیق دادن عبارت جست و جو شده می باشد، فقط از میان داده های موجود در پایگاه داده خود انجام دهند نه از میان میلیاردها وب سایت.

ارزیابی نمایه

به طور کلی هر نظامی جهت بهبود و ارتقاء احتیاج به ارزیابی دارد. هدف از ارزیابی یک سیستم نمایه سازی نیز بررسی میزان اثربخشی و کارایی آن است.

معیارهای ارزیابی نمایه از نظر Cleveland کلوند:

کلوند نمایه ها را در سه سطح ارزیابی کرده است:

۱- **فنی:** در این سطح نمایه باید دارای زبان مناسب و شکل قابل درک باشد و به سادگی بتوان از آن استفاده کرد.

۲- **معنایی:** در این سطح واژه ها باید معانی را بدون ابهام منتقل کنند.

۳- **میزان کارایی:** در این سطح نمایه باید اطلاعات مرتبط را به درستی شناسایی کند و در بازیابی اطلاعات موثر

باشد

دو شاخص مهم جهت ارزیابی نمایه ها ارزیابی جامعیت و مانعیت یک سیستم است .

جامعیت (Recall) : توانایی نظام در بازیابی مدارک مرتبط. جامعیت شاخصی است که نشان می دهد یک نظام تا چه اندازه قدرت دارد اسناد مرتبط بیشتری را بازیابی کند. بازیافت یا جامعیت عددی است بین صفر و یک و هرچقدر مقدار آن به یک نزدیک تر باشد عملکرد نظام بهتر است.

آنچه به دست آمده و سودمند تصور می شود (یافته ها) = جامعیت

کل مدارک سودمند در فایل (یافته ها + نا یافته ها)

دقت یا مانعیت (Precision) : به معنای توانایی نظام در کنار گذاشتن مدارک نامرتب است. شاخصی است که توانایی یک نظام را در ممانعت از بازیابی شدن اسناد غیر مرتبط نشان می دهد. دقت یا مانعیت نیز عددی است میان صفر و یک و هر چه به یک نزدیک تر باشد، نشانه عملکرد بهتر نظام است.

تعداد مدارک بازیابی شده مرتبط = مانعیت

تعداد کل مدارک بازیابی شده

از طریق این دو شاخص ، توانایی نظام در بازیابی آنچه می خواهیم و کنار گذاشتن آنچه با نیاز ما سازگار نیست سنجیده می شود.

ریزش کاذب (False drop یا Fall out) : ریزش کاذب در واقع ناتوانی نظام در بازیابی نکردن موارد غیر مرتبط است. به عبارت دیگر ریزش کاذب نماد احتمال بازیابی اطلاعات ناخواسته است. ریزش کاذب نیز عددی است میان صفر و یک . اما بر خلاف شاخص های قبلی، مقدار ریزش کاذب هر چه قدر به صفر نزدیک تر باشد (یعنی هر چه قدر کمتر باشد) نشان دهنده عملکرد بهتر نظام است.

ارتباط ریزش کاذب با جامعیت و مانعیت

معمولاً ریزش کاذب درصد کمی از مدارک بازیابی شده را شامل می شود؛ چرا که برخی از خطاهای بازیابی گریز ناپذیرند، بویژه هنگامی که به منظور دستیابی به جامعیت بالا، کاوش را گسترده کنیم . یعنی هر چه کاوش را گسترده تر کنیم وبخواهیم جامعیت را بالا ببریم، در مانعیت دچار مشکل میشویم و رکوردهای ناخواسته بازیابی شده زیاد می شود که همان ریزش کاذب است . در نتیجه، هر چه میزان جامعیت بیش تر باشد، میزان ریزش کاذب نیز زیاد می شود و هر چه میزان مانعیت زیاد باشد، میزان ریزش کاذب کاهش می یابد..

کارایی (Efficiency)

کارایی شاخصی است که به صورت یکجا عملکرد سیستم را در بازیابی هر چه بیشتر اقلام مرتبط و بازیابی نکردن هر چه بیشتر اقلام غیر مرتبط نشان می دهد. کارایی نیز عددی است میان صفر و یک و هر چه بیشتر باشد نشانه عملکرد بهتر نظام است.

- ✓ به طور کلی کیفیت یک نظام نمایه سازی علاوه بر جامع بودن و مانع بودن یافته ها به تازگی (بهنگام بودن) و صحت داده ها نیز بستگی دارد.
- ✓ یکی از عواملی که می تواند بازیابی موفق با جامعیت و مانعیت بالا را در پی داشته باشد، استفاده از عملگرهای بولی (AND-OR-NOT)، مجاورت کلمات، جستجوی کلمات هم ریشه، عملگرهای رابطه ای و رعایت ترتیب ارائه واژه ها و عملگرها می باشد.

اصطلاحنامه Thesaurus

گنجواژه یا اصطلاحنامه، مجموعه اصطلاحات یک رشته است که میان آنها روابط معنایی، رده‌ای، و سلسله مراتبی برقرار شده و توانایی آن را دارد که موضوع آن رشته را با همه جنبه‌های اصلی و فرعی و وابسته، به گونه‌ای نظام‌یافته و به منظور ذخیره و بازیابی اطلاعات ارائه دهد.

اصطلاحنامه معادل فارسی واژه انگلیسی Thesaurus است که ریشه در زبان لاتینی باستان دارد و به معنای گنجینه، ذخیره، و مجموعه به کار می‌رود.

اصطلاحنامه از نظر وظیفه و کارکرد، ابزار کنترل واژه‌ها به منظور برگرداندن زبان طبیعی مدارک به زبان مقید است و از نظر ساختار، واژگان کنترل شده و پویای زمینه‌ای خاص از دانش بشری است که برای ذخیره و بازیابی اطلاعات آن حوزه به کار می‌رود.

اصطلاحنامه دارای هدف‌های اساسی زیر است:

- (۱) نمایاندن ساختار زمینه معینی از دانش چنان‌که هم نمایه‌ساز و هم جست‌وجوگر بتوانند از گستره آن زمینه و ارتباط میان مفاهیم آن با اندیشه‌های مرتبط آگاهی یابند
- (۲) ارائه اصطلاحات استاندارد در زمینه‌ای معین.
- (۳) برقراری نظام ارجاعات میان اصطلاحات و رده‌بندی اصطلاحات به صورت سلسله مراتبی
- (۴) تأکید بر توجه به نیازهای اطلاعاتی استفاده‌کنندگان
- (۵) تعیین اصطلاحات مجاز و مشخص کردن حدود معانی اصطلاحات به منظور ایجاد هماهنگی در نمایه‌سازی

روابط میان اصطلاحات

باید توجه داشت که ویژگی ذاتی اصطلاحنامه، توانایی تعیین و نمایش روابط معنایی میان واژه‌هاست، این روابط ممکن است یکی از این سه نوع باشد:

الف) رابطه هم ارزی : **Equivalence Relation**

میان اصطلاح پذیرفته‌شده و اصطلاح پذیرفته‌نشده برقرار می‌شود

مثال: اصطلاح جامعه‌شناسی به جای علم‌الاجتماع

ب) رابطه سلسله مراتبی **Heirachial Relation**

بیان‌کننده رابطه اعم و اخص میان مفاهیم است که در واقع، اصطلاحنامه‌ها را از واژه‌نامه‌های متداول متمایز می‌کند.

ج) رابطه همبسته یا همایند: **Associative Relation**

رابطه میان دو اصطلاح که به دلیل وابستگی معنایی، وجود یکی دیگری را نیز به ذهن متبادر می‌کند. مثال: دو اصطلاح اسب و سوارکاری

ساختار اصطلاحنامه

اصطلاح اعم (ا ع) : **Brother Term** : اصطلاح عام‌تر

مثال: انگور

ا. ع. میوه

اصطلاح اخص (ا خ) : **Narrow Term** : اصطلاح خاص‌تر

مثال: میوه

ا. خ. انار

انگور

پرتقال

اصطلاح وابسته (ا و) : **Related Term** : اصطلاحی که به اصطلاح برگزیده به نوعی ارتباط دارد.

مثال: انگور

ا. و. شیر

اصطلاح به کار ببرید (بک) : **Use** : ارجاع از اصطلاح ناگزیده (نامرجح) به برگزیده (مرجح)

مثال: تعلیم و تربیت

بک. آموزش و پرورش

اصطلاح به کار ببرید به جای (بچ) : **Used for**: ارجاع از اصطلاح گزیده (مرجح) به اصطلاح ناگزیده (نا مرجح) که البته مترادف است.

مثال: آموزش و پرورش

بچ. تعلیم و تربیت

یاد داشت دامنه (ی د) **Scope Note** :: یادداشتی برای توضیح مدخل که دارای ابهامی و یا دارای چند معنا است و یا معنی خیلی کلی دارد استفاده می شود .

مثال: فرهنگ (آداب و رسوم)

کنفرانس ها

بچ. گردهماییها

او. محاورات

سمینارها

مباحثات

مقالات کنفرانسها

اصطلاحنامه ها نقش مهمی در نظامهای ذخیره و بازیابی اطلاعات دارند. ظهور وب، همراه با توسعه و پیشرفتهای اخیر در ر کاربرد اصطلاحنامه ها به عنوان ابزارهای بازیابی اطلاعات باعث تولد نسل جدید اصطلاحنامه ها شده است.

اصطلاحنامه های وب محور، راه خود را به محیطهای بازیابی و سازماندهی اطلاعات وب محور باز نموده و در تهیه ابر داده ها، نمایه سازی صفحات، سایتهای وب، پایگاههای داده و موتورهای جستجو استفاده می شوند.

ساختارهای معنایی موجود در اصطلاحنامه ها می توانند هم در سازماندهی و هم در بازیابی اطلاعات وب و منابع دانش نقش داشته باشند.

معرفی چند اصطلاحنامه فارسی و لاتین

اصطلاحنامه های فارسی

اصطلاحنامه اصف

اصطلاحنامه نما

اصطلاحنامه توسعه فرهنگی یونسکو

اصطلاحنامه آنلاین جامع علوم اسلامی islamicdoc.com

اصطلاحنامه های لاتین

اصطلاحنامه آنلاین techterms.com

اصطلاحنامه آنلاین کشاورزی agclass.nal.usda.gov

اصطلاحنامه اریک <http://eric.ed.gov>

وب معنایی (Semantic web)

وب معنایی نسبت به وب لغوی یک انقلاب محسوب می شود که در آن اطلاعات قابل خواندن و تجزیه و تحلیل توسط ماشین است. درحالی که صفحات وب کنونی را فقط انسان می تواند بخواند وب معنایی این اجازه را به مرورگرها و دیگر نرم افزارها می دهد تا اطلاعات را خوانده به راحتی تجزیه و تحلیل کنند. فضایی جهانی از جنس محاسبات هوشمند ماشینی را می توان تصور کرد که در آن تمامی پایگاه های دانش به صورت معنا گرا و با توانایی درک مفهومی همدیگر در کنار هم قرار خواهند گرفت. در آینده وب نه فقط توسط انسانها قابل فهم است بلکه توسط ماشین ها نیز قابل درک و پردازش است.

در زیر سه تعریف مختلف از وب معنایی ارائه شده است:

پروژه ای با هدف ایجاد رسانه ای جهانی برای رد و بدل کردن اطلاعات بصورتی که برای کامپیوتر قابل فهم و پردازش باشد.

وب معنایی، شبکه ای از اطلاعات در مقیاس جهانی است به نحوی است که پردازش آنها توسط ماشین ها به سادگی امکان پذیر است.

وب معنایی شامل داده های هوشمند وب است که توسط ماشین ها قابل پردازش است.

آنتولوژی (Ontology)

آنتولوژی در لغت از دو کلمه **onto** به معنی هستی و **logy** به معنی مطالعه تشکیل شده است. ریشه آنتولوژی در فلسفه است و به ارسطو نسبت داده می شود. در علوم کامپیوتر و در حوزه وب معنایی، آنتولوژی مفهوم کلمات و ارتباطات بین آنها را در حوزه ای که مورد استفاده قرار میگیرند، نشان می دهد.

اطلاعات در وب به صورت فزاینده ای رو به رشد هستند و استفاده کنندگان وب نیازمند یک درک مشترک از آنها هستند.

آنتولوژی یک مدل مفهومی است که موجودیتهای واقعی در یک دامنه خاص و روابط بین آنها را به صورت صریح و رسمی مدلسازی می کند.

آنتولوژی در وب

فرض کنید که می خواهید در مورد موضوعی با کسی صحبت نمایید. برای اینکه طرف مقابل، حرف شما را کامل و درست متوجه شود، احتیاج است که حوزه بحث کاملاً مشخص باشد. واژه های مختلف در حوزه های گوناگون، معانی یا تعبیرهای متفاوتی دارند و حتی گاهی ممکن است با وجود مشخص بودن حوزه بحث، یک کلمه خاص در ذهن افراد مختلف دارای تفاوتی اندکی باشد. واضح است که تنها راه رهایی از چنین وضعیتی، یک مجموعه واژگان مشترک بین افراد است.

آنتولوژی در وب معنایی دقیقاً چنین کاربردی را دارد.

در هر آنتولوژی، تمام موجودیتهای یک حوزه به صورت کامل و با ذکر تمام ویژگیها فهرست می شوند، سپس باید ارتباطات بین آنها بیان شود و مرحله بعدی این است که تمام اطلاعات فوق با یک فرمت خاص درون مستندات اینترنتی قرار داده شود و اطلاعات موجود در آن مستند به آنتولوژی اتصال داده شود.

یک آنتولوژی، لغات و مفاهیمی را که در تعریف و نمایش محدود ه ای از دانش به کار میروند، تعیین کرده و بنابراین معانی را استاندارد می کند.

بطور کلی، به وسیله آنتولوژی دو کار صورت می پذیرد:

- آنتولوژی یک فهم مشترک از یک مفهوم واحد را ارائه میدهد. زیرا بعضی وقتها ما برای یک مفهوم از دو لغت متمایز استفاده میکنیم. یا بالعکس از یک لغت برای دو مفهوم متفاوت استفاده میکنیم
- آنتولوژی بین مفاهیم در وب و دنیای واقعی ارتباط برقرار می کند

عناصر اصلی تشکیل دهنده آنتولوژی عبارتند از:

۱. مفاهیم
۲. ارتباط بین مفاهیم
۳. خصوصیات آنها

به عبارت دیگر آنتولوژی ارتباط بین مفاهیم در اسناد وب و دنیای واقعی را مشخص میکند که با این کار اسناد مربوطه توسط ماشینها قابل پردازش و فهم می شوند و اشتراک گذاری بین عاملها را تسهیل مینماید.

کنسرسیوم وب جهانی در نوامبر ۲۰۰۲، زبان OWL را به عنوان زبان نشانه گذاری معنایی به منظور انتشار هستی شناسی های وب پیشنهاد کرد. این زبان بر مبنای RDF است. OWL با تعریف کلاسها ، نمونه ها و روابط به طور واضح و رسمی در توسعه و ساخت هستی شناسی ها به کار می رود. OWL نسبت به RDF، XML و RDFS امکانات بیشتری برای بیان مفاهیم و معانی دارد.

موفق باشید