

نمایه‌سازی ماشینی: اصول و کاربرد



اعظم مهری^۱

بود، به طور روزافزونی نیاز به نمایه‌سازی ماشینی را گسترش می‌داد و در این راستا طرح‌های موفق‌تری نیز ارائه گردید. نمونه‌ای از این طرح‌ها عبارتند از:

۱. طرح AIMS^۲ کتابخانه ملی پزشکی در آمریکا (سیستم مدیریت و نمایه‌سازی خودکار)؛
۲. طرح MIS و طرح Copcy توسط زیمنس؛
۳. MAI در مرکز دکوماتاسیون آمریکا؛ تحلیل عنوان و چکیده مدارک علمی برای نمایه‌سازی؛



۴. Passat توسط زیمنس، سیستم نمایه‌سازی برای زبان آلمانی و دیگر زبان‌ها؛ و
۵. TCS^۳ توسط گروه کارنگی (سیستم مقوله‌بندی متن). (۹، ص ۷)

تعریف نمایه‌سازی ماشینی

قبل از تعریف نمایه‌سازی ماشینی، لازم است به تعریف و توصیف نمایه‌سازی پرداخته شود؛ توجه به این که نمایه‌سازی چیست و چرا نمایه‌سازی می‌کنیم. برای نمایه‌سازی، تعریف‌های گوناگونی ذکر شده است که با ترکیب و تلفیق آن‌ها می‌توان به یک تعریف جامع از نمایه‌سازی رسید. نمایه‌سازی عبارتست از ثبت و ضبط

نشده است. مثلاً همه منتظرند زبان‌های طبیعی دنیا به صورت رایانه‌ای به یکدیگر ترجمه شوند ولی علی‌رغم تمام کوشش‌ها و سرمایه‌گذاری‌ها، هنوز مرز بین فرآیندهایی که می‌تواند رایانه‌ای شود و آن‌ها که امکان رایانه‌ای شدن ندارند به طور کامل مشخص نشده است. (۲، ص ۲۶۶)

تاریخچه نمایه‌سازی ماشینی

امروزه بسیاری از نمایه‌ها را به‌طور کامل با رایانه تدوین می‌کنند. نگاهی به تاریخچه و ظهور نمایه‌سازی ماشینی، حضور آن در سال ۱۹۶۰ گزارش شده است. در این سال علاقه و گرایش به استفاده از الگوهای برداری به جهت بازیابی اتوماتیک (خودکار) اطلاعات به وجود آمد. به طوری که براساس این الگوها در برابر سوالات مرجع، مدارک به کمک کلیدواژه‌گان قابل بازیابی شدند، همچنین طی سال‌های ۱۹۶۸-۱۹۶۹ تحقیقاتی در پروژه AIRIX توسط استیگ^۴ و فانگ میر^۵ انجام شده است. (۹، ص ۱)

از آغاز سال ۱۹۷۰ در آلمان، مطالعات ریخت‌شناسی سیستم‌های ترجمه با هدف بازیابی اطلاعات شروع شد. از جمله طرح‌های مهم این دهه، می‌توان به طرحی به نام Rassat که اساساً در حیطه مطالعه زبان‌های رایانه‌ای می‌باشد، اشاره کرد.

در حدود سال ۱۹۸۰، سیستم‌های هوشمندی که بسیاری از دانشمندان در جهت تحلیل محتوی به آن اعتماد داشته باشند، تا حدودی ایجاد شد. در سال‌های اخیر، تلاشی که در ۱۹۸۰ شروع شده

مقدمه

افزایش و رشد سریع اطلاعات در عصر حاضر و توسعه شبکه‌های متعدد کتابشناختی، فرصت‌های متعدد و چالش‌های بی‌سابقه‌ای را فرا روی حرفه کتابداری و اطلاع‌رسانی برای دستیابی و بررسی اطلاعات گذارده است.

نمایه، مهمترین ابزار برای بازیابی اطلاعات و کلید دسترسی فوری به محتوای منابع اطلاعاتی است. (۷، ص ۱۲)

حوزه فعالیت نمایه‌سازی، منحصر به مدارک علمی نیست؛ بلکه نشانه‌هایی از اختراعات ثبت شده و ترکیبات جدید شیمی و مسائل علم فیزیک و همچنین مشخصات محققین در حوزه فعالیت نمایه‌سازی ظاهر شده است.

با پیشرفت علم تکنولوژی در سال‌های اخیر، این فکر به ذهن انسان خطور کرد که با صرف وقت و هزینه برای توسعه و تحقیق، می‌توان



تمام کارها حتی نمایه‌سازی و چکیده‌نویسی را برعهده رایانه گذاشت. تاکنون سعی و تلاش‌هایی صورت گرفته است تا فرآیندهای ذهنی به صورت الگوریتمی ارائه شوند؛ ولی هنوز کاملاً عملی



محتوای اطلاعاتی مدارک با استفاده از روش‌های گوناگون به منظور سازمان‌دادن اطلاعات به‌قصد سهولت بازیابی و اما نمایه‌سازی رایانه‌ای؛ امروزه بسیاری از نمایه‌ها با رایانه یا با کمک رایانه تدوین می‌شوند. البته به نظر می‌رسد که رایانه می‌تواند کل نمایه یک متن را به طور خودکار و بدون دخالت انسان استخراج القیایی کند. (۵، ص ۳)

به هر حال، رایانه امکانات مفیدی را در نمایه‌سازی فراهم آورده است یکی از این امکانات، ثبت جای‌نمای واژگان تکراری است. این مسئله در نمایه‌ها، به‌ویژه اعلام و مکان‌ها بسیار مفید است. امروزه، نرم‌افزارها این قابلیت را دارند که پس از انتخاب اولین کلیدواژه، کلیدواژه‌های تکراری



بعدی را در صفحات دیگر نیز انتخاب کنند؛ به‌طوری‌که احتیاجی به دوباره انتخاب کردن آن نباشد. بدین ترتیب، صرفه‌جویی زیادی در وقت و زمان نمایه‌سازی می‌شود.

در نمایه‌سازی ماشینی، همه واژگان مفهومی محتوای مدرک، به زبان طبیعی انتخاب می‌شوند. نمایه‌سازی ماشینی تنها به واژگان استخراج‌شده محدود نمی‌شود؛ بلکه سرعنوان‌های موضوعی، یادداشت‌های توصیفی و مکنون‌ها را نیز ارائه می‌دهد.

برای استفاده از روش نمایه‌سازی ماشینی، باید داده‌ها به صورت ماشین‌خوان درآید و این که واژه‌ها در کدام محدوده قرار گیرد، بستگی به نرم‌افزار مورد استفاده دارد. شرکت زافتکس نوعی برنامه رایانه‌ای با عنوان IDX طراحی کرده است که به ارائه خدمات نمایه‌سازی ماشینی می‌پردازد. اصول این نرم‌افزار برای آزمایش بازیابی اطلاعات در مقاله‌ای توسط گرومان کاملاً تشریح شده است. این نرم‌افزار از روش استفاده از واژه‌نامه بهره می‌گیرد. تغییرات واژگان از طریق مطابقت‌دادن با واژه‌های ذخیره‌شده به صورت الکترونیکی اعمال می‌شود. همچنین این نرم‌افزار امکان تبدیل واژگان را به ریشه آن‌ها برای بازیابی‌های بعدی فراهم می‌کند. علامت‌گذاری و محدودکردن واژه‌های غیرمجاز و شکستن واژه‌های مرکب و مترادف‌سازی و تنظیم ساختار عبارت‌ها را نیز انجام می‌دهد.

روش‌های مختلف نمایه‌سازی

برای نمایه‌سازی ماشینی، روندهای مختلفی موجود است که به سه روش کلی انجام می‌گیرد:

۱. روش زبان‌شناختی؛
۲. روش سنجش احتمالات؛ و
۳. روش آماری.

در روش زبان‌شناختی، سعی می‌شود با کمتر ساختار نحوی و معنایی و تحلیل‌های شکل‌شناسی مدرک، توصیف‌گرها استخراج گردند. این عمل خود به سه روش تقسیم می‌شود:

۱. تجزیه و تحلیل ریخت‌شناسی؛
۲. تجزیه و تحلیل نحوی؛ و
۳. تجزیه و تحلیل معناشناسی.

در روش سنجش احتمالات، برای مدل‌سازی ریاضی مراحل بازیابی به کار گرفته می‌شود. برعکس توزیع آماری که اصطلاحات یک مدرک را مورد استفاده قرار می‌دهد و در روش آماری، مشخص می‌کند معنای هر مفهوم منفرد در مدرک، با حضور آن در جایگاه‌های مختلف مدرک رابطه تنگاتنگ دارد.

فاکتورهای ارزش‌گذاری اصطلاحات در

روش آماری:

۱. شمارش لغات
۲. وضعیت متن

باید توجه کرد که ارزش‌گذاری لغات مشخص شده در متن، زمانی صورت می‌گیرد که موقعیت لغات در متن با یک معیار مناسب سنجیده شود؛

۳. برای محاسبه مجموعه لغات از طریق روش‌های آماری می‌توان به روش «لون» که یکی از پیشگامان نمایه‌سازی ماشینی است، اشاره کنیم. وی روش خود را بر اساس محاسبه کثرت لغات، در سه مرحله بنا نهاده است. همچنین از روش‌های اسپارک جونز و روش دنیس و سالتون نیز استفاده می‌شود. (۳، ص)

استفاده از رایانه در نمایه‌سازی

در بحث نمایه‌سازی، چون تحلیل مدرک شامل مشخص ساختن محتوا یا پیام یک مدرک است، باید به این سوالات که آیا مدرک واقعاً درباره چیست و آیا موضوع آن به چه مبحثی می‌پردازد، پاسخ گوید. عمل تحلیل، فرایند مشکل و



پیچیده‌ای است که تنها انسان قادر به انجام آن است و ماشین به اندازه کافی صلاحیت و توانایی انجام کامل و همه‌جانبه آن را ندارد و به طور ساده می‌توان پرسید چگونه از رایانه انتظار می‌رود با در نظر گرفتن تمام محاسبات ذهنی بشر، معنی و مفهوم را به صورتی که موضوعات مهم مدرک کاملاً بیان شود، در چند واژه، جایگزین چند هزار کلمه سازد. در نظام‌های خودکار (ماشینی)، کلیدواژه‌ها معمولاً بر مبنای بسامدشان در مدارک یا مجموعه انتخاب می‌شوند. یعنی کلیدواژه‌هایی که بیشتر در مدرک تکرار شده باشند، به دفعات

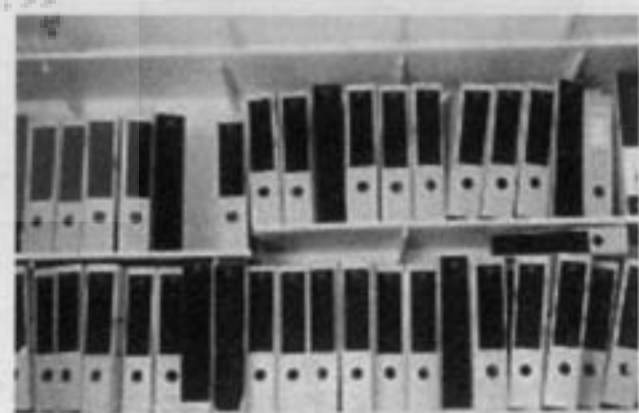


استخراج می شوند.

استفاده از رایانه در نمایه سازی، نتایجی را به صورت زیر به دنبال دارد. توصیف گرهای انتخاب می شوند که در هنگام بازیابی، مفاهیم قابل جستجو و دارای معنا باشند. در پایگاه های اطلاعاتی گسترده، در هنگام بازیابی، محدودیت کمتری وجود دارد. همچنین رشد روزافزون متقاضیان، مدارک مرتبط را به یکدیگر نزدیک تر می سازد؛ چرا که توصیف گرهای مرتبط بیشتری تولید می شوند.

از نظر جایگزینی رایانه به جای نیروی انسانی، اغلب سیستم های نمایه سازی خودکار واقعا خودکار نیستند بلکه با هدف کمک به نمایه سازی انسانی تولید شده اند. به طور کلی دو روش عمده برای نمایه سازی به کمک رایانه وجود دارد:

1. استفاده از رایانه برای تولید انواع نمایش های درون خطی و تهیه راهنمای کمکی برای نمایه ساز، به طوری که به صورت همزمان خطاهای نمایه ساز (مانند استفاده از ترکیب های نادرست، سرعنوان اصلی و سرعنوان فرعی و ...) را تشخیص و به اطلاع او برساند؛ و
2. استفاده از برنامه های رایانه ای برای خواندن متن و انتخاب اصطلاحات نمایه ای.



مزایای نمایه سازی ماشینی

برای بررسی معایب و مزایای نمایه سازی ماشینی می بایست به بررسی مقایسه ای نمایه سازی ماشینی و دستی پرداخته شود. اولین نکته مهم در بررسی و مقایسه نمایه سازی ماشینی و دستی، مسئله جامعیت و مانعیت آن ها است. نتایج بررسی های تحلیل محتوای نمایه سازی ماشینی و دستی نشان داده است که نمایه سازی ماشینی

نمایه سازی ماشینی دارای نقاط ضعف و نقاط قوتی است که نقاط قوت و مزایای آن عبارتند از:

1. صرفه جویی در وقت و زمان؛
2. استفاده از زبان طبیعی نمایه سازی؛
3. استفاده و پیروی از معیارهای اصولی روشن؛
4. پایین بودن مخارج نمایه سازی؛ و
5. برخورداری از امکانات بیشتر تکنولوژی نسبت به نمایه سازی دستی.

معایب نمایه سازی ماشینی

در نمایه سازی ماشینی، دیگر نمایه ساز مجرب وجود ندارد که توانایی امکان درک صحیح محتوای مدارک را داشته باشد. انتخاب توصیفگرها تخصصی نمی شود و به طور خلاصه باید گفت که نمایه سازی ماشینی نسبت به نمایه سازی دستی با توجه به رشد فزاینده اطلاعات که انتشار روزافزون مدارک را باعث می شود، مورد قبول بیشتری نسبت به دستیابی مدارک که نیاز به نمایه سازی در پایگاه های اطلاعاتی دارد، واقع می شود و آینده خوبی را برای نمایه سازی ماشینی می توان انتظار داشت.

نمایه سازی ماشینی بر خلاف نمایه سازی دستی نمی تواند محتوای صحیح یک مدرک را بفهمد و باید از معیارهای اصولی در متن پیروی کند. همچنین در روش های نمایه سازی ماشینی، تلاش بر این است که مراحل تحلیل محتوا به طور کامل ماشینی شوند تا از عملکرد نمایه سازها صرف نظر شود. نمایه سازی ماشینی اگرچه وسیله

ویرایش نشده می تواند به سطحی از جامعیت دست یابد که با سطح جامعیت به دست آمده از نمایه سازی دستی قابل مقایسه است؛ اما نسبت جامعیت آن بهتر از مانعیت حاصل از نمایه سازی انسانی است. (1- ص 382)

استخراج خودکار بر استخراج انسانی مزیتی آشکار دارد. استخراج ماشینی کاملاً یک پارچه و منسجم است و نمایه سازی دستی از این نوع استخراج برخوردار نیست بلکه از نوع نمایه سازی تخصصی است و انجام این نوع نمایه سازی از طریق رایانه مشکل است و امروزه دیگر نمایه سازی تخصصی جز در تولید نمایه های چاپی چندان مقبولیتی ندارد. سالی سال پیش، این روش بسیار استفاده و عمومیت داشت. به دلیل آن که در آن زمان ذخیره و پردازش رایانه ای هتون، در حجم زیاد بسیار پرهزینه بود و استفاده از هر روشی که بتواند کم هزینه و کم حجم باشد، قابل توجه بود.

از ویژگی های سیستم های نمایه سازی ماشینی پیشرفته می توان به نمایه سازی جزئی یک مدرک یا حداقل پیشنهاد اصطلاحات به نمایه ساز اشاره کرد. یکی از این سیستم ها، سیستم کین⁶ است که برای استفاده با پایگاه آگرپ⁷ توسعه یافته است. (آگرپ پایگاهی متعلق به جامعه متحد اروپاست که طرح های جاری در حوزه کشاورزی را در بر دارد). در نمایه سازی ماشینی مراحل نمایه سازی به جای نمایه ساز توسط رایانه انجام می پذیرد. انجام مراحل نمایه سازی بستگی به نوع نرم افزار مورد استفاده دارد.

Universitaet poatdam, Informationa
wissenschaft, Modellversuch. BETID,
1994.

پی نوشت ها:

۱- دانشجوی کارشناسی ارشد کتابداری و اطلاع‌رسانی
دانشگاه آزاد اسلامی واحد همدان

2. Gerhard Lastig
3. Fong Meyer
4. Automated Indexing
and Management
5. Text Categorization Shell
6. Kain
7. Agrp

نمی‌توان نمایه‌سازی را جایگزین فهرست‌های
موضوعی کرد بلکه آن را فقط می‌توان کامل کرد
و عاملی است که نمایه‌سازی دستی و ماشینی را
با هم پیوند می‌دهد.

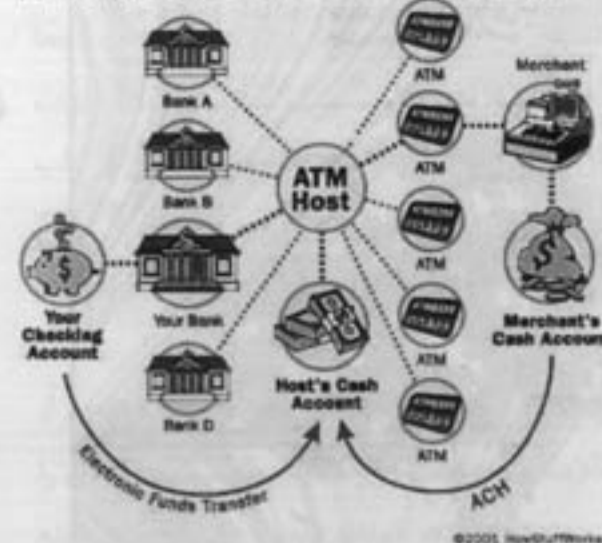
در حقیقت روش‌های نمایه‌سازی ماشینی،
نمایه‌سازی به کمک رایانه است. با استفاده از
نمایه‌سازی ماشینی، جستجوی بازیابی اطلاعات
سهل‌تر می‌شود؛ به‌خصوص در فهرست‌نویسی
مجلات و نشریات الکترونیکی. با استفاده از
کلیدواژه‌ها بهتر از سرعنوان‌های موضوعی
می‌توان به نتیجه مطلوب دست یافت و در نهایت
استفاده از نمایه‌سازی ماشینی باعث بالارفتن
جامعیت اما پایین آمدن مانعیت می‌شود.

مآخذ و منابع:

۱. لنکستر، دبلیو اف.، ترجمه عباس گیلوری
(نمایه‌سازی و چکیده‌نویسی: مبانی نظری و
عملی). تهران: چاپار، ۱۳۸۲.
۲. فاگن، پیتر، ترجمه علی مزینانی (تحلیل
موضوعی و نمایه‌سازی: مبانی نظری و توصیه‌های
عملی). تهران: کتابخانه ملی، ۱۳۷۴.
۳. نیاکان، شهرزاد (نمایه‌سازی ماشینی). تهران:
مرکز اطلاعات و مدارک علمی ایران، ۱۳۸۳.
۴. نیاکان، شهرزاد. "کاربرد نمایه‌سازی ماشینی
در کتابخانه‌ها"، اطلاع‌رسانی، ۱۶: ۶
۵. سینایی، علی. "راهنمای نمایه‌سازی"، اطلاع
رسانی.
۶. معتمدی، فاطمه. "نظام نمایه‌سازی و
ورداسمیت در شبکه کتاب‌شناختی OCLC"،
اطلاع‌رسانی، ۱۷، ۱۵.

7. Kaiser, Alexander.
Computer Unterstue Ztes Indexieren
inntelligence Retrieval systemen.
Doktrat Disserttion in Fachbereich
Informationswirt schaft, Wirtschafts
universitaet Wien, 1993.
8. Willitsm, John. Database Design
and Construction, 1992.
9. Knorz, Gerhard, und et al.
"Automatische Indexierung".
Lehramaterialen, N.3. Wissenser
praesentation und Information Retrieval.

کاملی برای استخراج مفهومی واژگان نیست،
اما م. ت. آن را به صورت قابل. تکمیل.



و مورد استفاده قرار داد. با کمک نمایه‌سازی
ماشینی نمی‌توان امکاناتی را برای استفاده از
مفاهیم مرتبط با موضوع‌های مدارک فراهم آورد.
نمایه‌سازی ماشینی عملکرد و نتایج جستجوها
را افزایش می‌دهد و ابزاری تکمیل‌کننده است
و بیشتر در کتابخانه‌های تخصصی و دانشگاهی
مورد استفاده و استقبال قرار گرفته است. در مورد
کیفیت کار نمایه‌سازی ماشینی، سه عملکرد را
باید در نظر گرفت:

۱. عملکرد نرم‌افزار نمایه‌سازی؛
۲. مدارک مورد نمایه‌سازی؛ و در نهایت
۳. کیفیت واژه‌نامه مورد استفاده.

نتیجه

مکانیزه کردن نمایه‌سازی و چکیده‌نویسی با
کیفیتی که انسان انجام می‌دهد، اقدامی است که
درصد موفقیت آن بسیار بالا نیست و مجبور است
یک سری دستورالعمل‌های از پیش تعریف‌شده
را به کار برد و از آنجا که ماهیت نمایه‌سازی و
چکیده‌نویسی با ترجمه مشابه هستند، به سادگی
می‌توان پیش‌بینی کرد که اگر بخواهیم این
فرآیند را با کیفیت بالا انجام دهیم، نمی‌توان آن‌را
کاملاً به صورت الگوریتمی و ماشینی انجام داد و
دلیل این امر عدم کیفیتی است که در این فرآیند
مستتر است.

مزیت تمام روش‌های نمایه‌سازی ماشینی
صرفه‌جویی در وقت و یک‌دستی در انتخاب
توصیف‌گرها است و در مقایسه با روش‌های
نمایه‌سازی دستی و معمول، عنوان‌های
مرتبط‌تری بازیابی نمی‌شود. اما به‌طور کل

