



Comprehensive Pattern of Designing and Implementing Automatic Indexing Software

Document Type: Research Paper

Mohsen Haji Zeinolabedini

*Corresponding author: Assistant Professor, Department of Knowledge and Information Science, Shahid Beheshti University, Tehran, Iran. E-mail: zabedini@gmail.com

Fatemeh Nemati

MSc, Department of Knowledge and Information Science, Shahid Beheshti University, Tehran, Iran. E-mail: nematif80@gmail.com

Amirreza Asnafi

Associate Professor, Department of Knowledge and Information Science, Shahid Beheshti University, Tehran, Iran. Email: aasnafi@gmail.com

Abstract

Purpose: The purpose of the research is to provide a model for designing indexing software that will automatically store information and index resources.

Methodology: : The research method was comparative evaluation and system design. The data collection tool was a researcher-made checklist that was validated by experts. The research community was considered 9 software's that were consistent with the research conditions. The checklist was completed in person by the analysis and development or support departments of the desired software.

Findings: The findings of the research indicated that for the design of automatic indexing software, it is necessary that the sections of cataloging and documentation, digital resources, indexing, management, search and general capabilities must be considered in the software. Also, practical and side parts such as entering standard information, correcting data caused by OCR, using artificial intelligence, using machine learning and the possibility of interoperability with other systems should be considered.

Conclusion: A review of the various software available in Iran indicated that none of these tools have fully implemented the capabilities required for automatic indexing, but the necessary infrastructure for designing and developing an independent automatic indexing software in Persian or adding a profile section Automatic adaptation to existing systems is provided. The final proposed template of the research includes the parts and conditions necessary for computer and information science professionals to design automatic indexing firmware.

Keywords: Automatic Indexing, Keyword Extraction, Software Template, Persian Automatic Indexing.

Citation: Haji Zeinolabedini, M., & Nemati, F., & Asnafi, A. (2022). Comprehensive Pattern of Designing and Implementing Automatic Indexing Software. *Management of Information Resources & Services*, 8(4), (31-44). (In Persian)

(DOI): 10.30473/MRS.2022.65043.1530

Quarterly Journal of Knowledge and Information Management

Vol 8, No 4, (Series 32), winter 2022, (31-44)

Received: (2022/08/09)

Accepted: (2022/12/01)

Copyrights

© 2022 by the authors. Licensee PNU, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)





الگوی جامع طراحی و پیاده‌سازی نرم‌افزار نمایه‌سازی خودکار

نوع مقاله: پژوهشی

محسن حاجی زین‌العابدینی

*نویسنده مسئول: استادیار، گروه علم اطلاعات و دانش‌شناسی، دانشگاه شهید بهشتی، تهران، ایران.

E-mail: zabedini@gmail.com

فاطمه نعمتی

E-mail: nematif80@gmail.com

کارشناس ارشد، گروه علم اطلاعات و دانش‌شناسی، دانشگاه شهید بهشتی، تهران، ایران.

امیر رضا اصنافی

E-mail: aasnafi@gmail.com

دانشیار، گروه علم اطلاعات و دانش‌شناسی، دانشگاه شهید بهشتی، تهران، ایران.

چکیده

هدف پژوهش: هدف پژوهش ارائه الگویی است جهت طراحی نرم‌افزار نمایه‌سازی که عمل ذخیره‌سازی اطلاعات و نمایه منابع را به شکل خودکار انجام دهد.

روش کار پژوهش: روش پژوهش ارزیابی تطبیقی و طراحی سیستم بوده است. ابزار گردآوری اطلاعات سیاهه واریسی محقق ساخته بوده که توسط متخصصان اعتباریابی شد. جامعه پژوهش ۹ نرم‌افزار که با شرایط پژوهش همخوانی داشتند در نظر گرفته شد. سیاهه واریسی به صورت حضوری و از طریق بخش‌های تحلیل و توسعه یا پشتیبانی نرم‌افزارهای مورد نظر تکمیل شد.

یافته‌ها: یافته‌های پژوهش نشان داد که برای طراحی نرم‌افزار نمایه‌سازی خودکار لازم است که بخش‌های فهرست‌نویسی و مستندسازی، منابع دیجیتال، نمایه‌سازی، مدیریت، جستجو و قابلیت‌های عمومی به‌طور اجباری در نرم‌افزار مدنظر قرار بگیرند. همچنین، بخش‌های کاربردی و جانبی نظیر ورود اطلاعات استاندارد، اصلاح داده‌های ناشی از اوس سی آر، بهره‌گیری از هوش مصنوعی، استفاده از یادگیری ماشین و امکان میان‌کنش‌پذیری با سایر سامانه‌ها در نظر گرفته شوند.

بحث و نتیجه‌گیری: بررسی نرم‌افزارهای مختلف موجود در ایران، نشان داد که هیچ‌کدام از این ابزارها، قابلیت‌های لازم برای نمایه‌سازی خودکار را به صورت کامل پیاده‌سازی نکرده‌اند، اما زیرساخت‌های لازم برای طراحی و توسعه نرم‌افزار مستقل نمایه‌سازی خودکار به زبان فارسی یا افزودن بخش نمایه‌سازی خودکار به سامانه‌های موجود فراهم است. الگوی نهایی پیشنهادی پژوهش، بخش‌ها و شرایط لازم برای متخصصان علم اطلاعات و رایانه برای طراحی نرم‌افزار نمایه‌سازی خودکار را در بر گرفته است.

واژه‌های کلیدی: نمایه‌سازی خودکار، استخراج کلمات کلیدی، قالب نرم‌افزار، نمایه‌سازی خودکار فارسی.

استناد: حاجی زین‌العابدینی، محسن و نعمتی، فاطمه و اصنافی، امیررضا (۱۴۰۰). الگوی جامع طراحی و پیاده‌سازی نرم‌افزار نمایه‌سازی خودکار. مدیریت منابع و خدمات اطلاعاتی، ۸(۴)، (۴۴-۳۱).

(DOI): 10.30473/MRS.2022.65043.1530

فصلنامه مدیریت اطلاعات و دانش‌شناسی

دوره ۸، شماره ۴ (پیاپی ۳۲)، زمستان ۱۴۰۰، (۴۴-۳۱)

تاریخ دریافت: (۱۴۰۱/۰۵/۱۸)

تاریخ پذیرش: (۱۴۰۱/۰۹/۱۰)

Copyrights

© 2022 by the authors. Licensee PNU, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



مقدمه

با پیشرفت علم و تکنولوژی و ایراداتی که در نمایه‌سازی دستی وجود داشت متخصصان به این نتیجه رسیدند که برای کم کردن هزینه و زمان و افزایش دقت و کارآمدی، از راه‌های دیگری برای نمایه‌سازی استفاده کنند. راهی که متخصصان علم اطلاعات و علوم کامپیوتر مطرح کردند استفاده از فناوری رایانه‌ای برای نمایه‌سازی خودکار بود. در نمایه‌سازی خودکار، رایانه کلیدواژه‌های مهم را از متن استخراج می‌کند و آن‌ها را براساس روش‌های استاندارد طبقه‌بندی می‌کند (مونتجو راز، ۲۰۰۲).

نظام نمایه‌سازی، فرایندی است که طی آن، محتوای موضوعی مدارک براساس قواعد خاصی در قالب توصیف‌گرها یا سرعنوان‌های موضوعی و یا زبان طبیعی ذخیره می‌شوند و در دسترس کاربران قرار می‌گیرد. سازمان استاندارد ملی اطلاعات ۳ یک فهرست را به‌عنوان "یک راهنمای سیستماتیک طراحی شده برای نشان دادن موضوعات یا ویژگی‌های اسناد به‌منظور سهولت در بازیابی اسناد یا بخش‌هایی از اسناد" تعریف می‌کند (اندرسون، ۱۹۹۷) ساخت نمایه‌ساز ماشینی برای زبان فارسی به‌عنوان یک سیستم نرم‌افزاری پیچیده، باید یک مدل مفهومی جامع با روش‌شناسی درخور، طراحی و سیستم بر پایه آن در طول زمان ساخته شود. طراحی چنین سیستمی باید براساس نیازمندی‌های زبان فارسی و تجربه‌های موجود در کشور از سویی، و منابع علمی و تجربه‌های موفق کشورهای دیگر انجام شود. برای اینکه مدل پیشنهادی، عملی شود، باید بستری مناسب برای اجرای مدل نیز فراهم باشد (جلالی‌منش، علیدوستی، خسروجردی، ۱۳۹۲)، نرم‌افزارهایی که کار نمایه‌سازی در داخل ایران را انجام می‌دهند، عموماً نمی‌توانند اطلاعات را از متون قدیمی (اسکن شده) استخراج کنند. همچنین استاندارد یکسانی برای تهیه فایل‌های متنی وجود ندارد که این خود باعث از بین رفتن امکان استخراج اطلاعات و نمایه‌سازی آن‌ها می‌شود. همچنین برای متون تخصصی، گاهی اصطلاحنامه کاملی وجود ندارد.

از دید نویسنده این مقاله، طبق بررسی‌های که انجام شده، آن چیزی که به‌عنوان خلأ چه در قالب پژوهش و چه در عمل احساس می‌شود، فقدان الگویی برای این نرم‌افزارها از لحاظ نمایه‌سازی خودکار است. لذا پژوهش حاضر درصدد است تا ضمن استخراج و تدوین یک چارچوب برای نرم‌افزارهای نمایه‌سازی مدل مطلوب طراحی نرم‌افزارهای نمایه‌سازی را تدوین

جامع بشری در طول زندگی خود، همیشه به دنبال کسب دانش و اطلاعات بوده است. جهت دستیابی به این هدف، انسان‌ها باید روزآمدترین خبرها و اطلاعات را درزمینه‌های مختلف علمی جمع‌آوری می‌کردند. پس از جنگ جهانی دوم، با افزایش سریع حجم اطلاعات، توسعه شبکه‌های متعدد کتابشناختی، پیشرفت علم و تکنولوژی رشد تولیدات علمی و به‌تبع آن رشد اسناد و مدارک چاپی و همچنین تولید، گسترش و توسعه اطلاعات الکترونیکی در رشد سریع و آنی اطلاعات اینترنتی باعث حجم اطلاعات قابل‌دسترس شده است. این عصر را عصر توانایی انتخاب بهترین اطلاعات در زمان مناسب می‌توان دانست. جهت کشف بهترین اطلاعات در بهترین زمان، اطلاعات باید در دسترس، قابل جستجو و بازیابی در کمترین زمان باشند. اطلاعاتی که در دسترس نباشند، یا به‌سختی قابل جستجو و بازیابی باشند، در واقع اطلاعات مفید نیستند. یکی از بهترین روش‌ها که باعث می‌شوند اطلاعات سریع‌تر در دسترس قرار گیرند، نمایه‌سازی آن‌هاست. نمایه‌سازی باعث می‌شود مهم‌ترین اطلاعات متن، به‌راحتی قابل جستجو و بازیابی شوند و بیشتر در دید قرار گیرند. در این مقاله الگویی جهت ایجاد یک نرم‌افزار نمایه‌سازی کامل ارائه شده است.

نمایه‌سازی از نگاه متخصصان در دو معنای عام و خاص به کار می‌رود. در معنای عام می‌توان نمایه‌سازی را فرایند تجزیه و تحلیل داده‌ها و اطلاعات کتابشناختی مدارک دانست. هدف از این کار پردازش داده‌های کتابشناختی برای واردکردن آن به یک سامانه اطلاعاتی است. و در معنای خاص معمولاً، منظور از نمایه‌سازی، تجزیه و تحلیل موضوعی مدارک است (توکلی‌زاده راوری، ۱۳۹۴)، با افزایش سریع حجم اطلاعات علمی، روش‌های دستی و قدیمی نمایه‌سازی و بازیابی کارایی خود را از دست داد. با توجه به سرعت افزایش تولید و مصرف اطلاعات نیاز به استفاده از کامپیوتر در عرصه‌های مختلف و از جمله در عرصه ذخیره و بازیابی اطلاعات ضروری گردید.

در نرم‌افزارهایی که نمایه‌سازی را انجام می‌دهند، مشکلات متعددی مشاهده می‌شود. از جمله این مشکلات می‌توان به «مشخص نبودن میزان کیفیت اطلاعات استخراج شده، عدم استفاده از زبان طبیعی» (گلوب و سورگل، ۲۰۱۵) و «یکدست نبودن روش ذخیره‌سازی و بازیابی اطلاعات» (ابو سبا کاظمینی، ۱۳۹۰) اشاره کرد.

اهداف پژوهش

هدف کلی در این مقاله، شناسایی مؤلفه‌های طراحی نرم‌افزار نمایه‌سازی خودکار منابع فارسی است.

سؤالات پژوهش

الگوی مناسب برای طراحی نرم‌افزارهای نمایه‌سازی چه الگویی است؟

نرم‌افزار نمایه‌سازی چه بخش‌هایی باید داشته باشد؟

هر بخش از این سامانه، شامل چه زیر بخش‌هایی باشد؟

نمایه‌سازی خودکار

نمایه‌سازی ماشینی یا نمایه‌سازی خودکار، روشی از نمایه‌سازی است که توسط ماشین انجام می‌شود. نمایه‌سازی ماشینی، نوعی نمایه‌سازی است که در آن با استفاده از الگوریتم رایانه‌ای، واژه‌های کلیدی یک مدرک از عنوان یا متن استخراج شده، سپس، در قالب مدخل‌های نمایه، مرتب و سازمان‌دهی می‌گردند (فیدل^۴، ۱۹۴۵).

جهت استفاده از روش نمایه‌سازی ماشینی داده‌ها باید به صورت ماشین‌خوان درآیند. اینکه واژگان از کدام محدوده متن انتخاب شوند به نرم‌افزار به کار گرفته شده بستگی دارد. (نیاکان، ۱۳۸۳) در این نوع نمایه‌سازی همه امور از انتخاب کلیدواژه، شماره‌گذاری، ترتیب‌بندی و غیره از طریق کامپیوتر انجام می‌گیرد.

نکته‌ای که در نمایه‌سازی خودکار باید مورد توجه قرار گیرد آن است که هر نوع نمایه‌سازی خودکار باید دو فرایند پایه را مورد نظر قرار دهد: اول، باید به شیوه‌ای واحدهای متنی را مجزا نماید. این عمل مستلزم فهم زبان طبیعی است. دوم، بر اساس واحدهای متنی یک‌زبان، کلمات نمایه را گزینش و استخراج کند. باید متذکر شد که به دلیل تنگناهای زبانی در زبان‌های نمایه‌سازی موجود، بهره‌گیری از ابزارهای نمایه‌سازی مثل اصطلاحنامه یا نظام رده‌بندی چندان مفید نیست (گیلوری، ۱۳۷۹).

می‌توان گفت که نمایه‌سازی خودکار با اختراع رایانه‌های بزرگ هم‌زمان بود. اولین نمایه‌سازی خودکار که توسط اچ پی لون انجام گرفت بر بسامد واژه‌ها در نمایه‌سازی متون تأکید داشت. این تحقیقات ادامه یافت و اکنون به روش‌های نمایه‌سازی خودکار پیچیده‌ای انجام شده است که بر استفاده از

کند. به عبارت دیگر مسئله اصلی این پژوهش این است که چارچوب، شاخص و اصول مطلوب نرم‌افزارهای نمایه‌سازی به چه صورتی است؟

هدف هر نمایه اطلاعاتی است که در فرایند نمایه‌سازی ذخیره شده است. بهترین نمایه آن است که در سریع‌ترین زمان ممکن، بهترین و مرتبط‌ترین اطلاعات را بازیابی کند یعنی از جامعیت^۱ و مانعیت^۲ لازم برخوردار باشد و برای انجام سریع‌تر و بهتر این کار نیاز به نرم‌افزاری مناسب و کارآمد احساس می‌شود (گلوب^۳، ۲۰۱۶).

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

نیاز به نرم‌افزارهای مناسب در حوزه ذخیره و بازیابی اطلاعات باعث شده است تا سازمان‌های ملی و بین‌المللی در اکثر نقاط دنیا فعالیت‌هایی را از سال‌ها پیش در این راستا انجام دهند. به دنبال همین تحقیقات در ایران نیز تلاش‌های گسترده‌ای در زمینه برآورده ساختن نیازهای کشور در رابطه با طراحی و توسعه نرم‌افزارهای کتابخانه‌ای آغاز شد اما نرم‌افزارهای ایرانی به‌ویژه در زمینه نمایش اطلاعات دارای مشکلاتی هستند (نوروزی و نعمتی، ۱۳۸۹).

با توجه به اهمیتی که نمایه‌سازی در بازیابی اطلاعات دارد تاکنون نمایه‌سازی در کتابخانه‌ها و سازمان‌ها و پایگاه‌ها به شکل دستی صورت می‌گرفت اما با ظهور تکنولوژی رایانه‌ای و هوش مصنوعی این کار به شکل خودکار و با استفاده از ماشین صورت می‌گیرد. برای زبان فارسی نیز نمایه‌سازی خودکار مهم و با اهمیت است. فقدان نرم‌افزاری که بتواند به شکل کاملاً خودکار منابع فارسی را نمایه‌سازی کند ما را بر آن داشت تا به ارائه الگویی برای این نوع نرم‌افزارها بپردازیم.

1. Comprehensiveness
2. Specificity
3. Golub

در تحلیل آماری متن، عناصری چون بسامد واژه‌های مهم محاسبه می‌شود و برای نمایه‌سازی اهمیت دارد. معمولاً واژه‌های غیر اطلاعاتی شامل نام اشیاء، حروف اضافه، حروف ربط، و واژه‌های رایج از قبیل «چون»، «مانند» و مانند آن در نظر گرفته نمی‌شود. ممکن است واژه‌ها براساس بسامد و تکرارشان رتبه‌بندی شوند. با توجه به این مهم که نمایه‌سازی توسط ماشین انجام می‌شود لازم است که متن مدرک قابل خواندن با ماشین باشد و ماشین بتواند هریک از کلمات را به‌طور انفرادی و جداگانه شناسایی کند و بتوانیم قوانینی برای انتخاب کلمات و عبارات موردنظر توسط رایانه تدارک ببینیم (یوسفی، ۱۳۷۷).

نمایه‌سازی خودکار به سه روش انجام می‌شود. روش زبان‌شناختی، که در این روش با کمک تحلیل‌های شکل‌شناسی و ساختار نحوی و معنایی متن مدرک، توصیفگرها را استخراج نماید. دوم روش‌های آماری، که مشخص می‌کند معنی هر مفهوم منفرد در مدرک با حضور آن در جایگاه‌های مختلف مدرک ارتباط تنگاتنگ دارد. همه روش‌های آماری بر این عقیده است که در نمایه‌سازی بار معنایی هر اصطلاح واحد در مدرک به تعداد آن در مدرک بستگی دارد؛ بنابراین لغات درون متن شمارش می‌شود و ارتباط آن‌ها ارزش‌گذاری می‌شود. روش سوم نیز روش مبتنی بر احتمالات است. در این روش‌ها تئوری احتمالات برای مدل‌سازی ریاضی مراحل بازیابی به کار گرفته می‌شود. در حالی که در توزیع آماری، اصطلاحات یک مدرک مورد استفاده قرار می‌گیرد. می‌توان گفت در این روش با انجام عملیات ریاضی بر روی موارد بازیابی شده به این نتیجه رسید که چه تعداد از موارد که یافت شده با موضوع موردنظر مرتبط بوده است، و با به‌کارگیری احتمالات یک تخمین احتمالی برای توزیع توصیفگر برای اصطلاح نمایه‌ای در نظر گرفت.

همچنین در تقسیم‌بندی دیگر، نمایه‌سازی را به دو روش حفظ اصطلاح و حذف اصطلاح تقسیم‌بندی می‌کنند که هرکدام از این روش‌ها، تقسیم‌بندی خاص خود را دارند. نمایه‌سازی‌های کوییک و کووک از مهم‌ترین روش‌های نمایه‌سازی حفظ اصطلاح هستند.

اصطلاحات واحد و استفاده از بخش بیشتری از اطلاعات مانند مفهوم عبارات تأکید دارند (سالتون^۱، ۱۹۷۱).

در آلمان از ۱۹۷۰ مطالعات ریخت‌شناسی آغاز شد و طرح‌های مختلفی به اجرا درآمد که از جمله آن‌ها طرح پاسات^۲ بود که در زمینه مطالعه زبان‌های کامپیوتری بود.

در حدود سال ۱۹۸۰ علاقه به سیستم‌های هوشمندی که بسیاری از دانشمندان در جهت تحلیل محتوا به آن اعتماد داشتند، ایجاد شد. از نمونه‌های آن سیستم دانشگاه ییل و نیز برلین در آلمان را می‌توان نام برد.

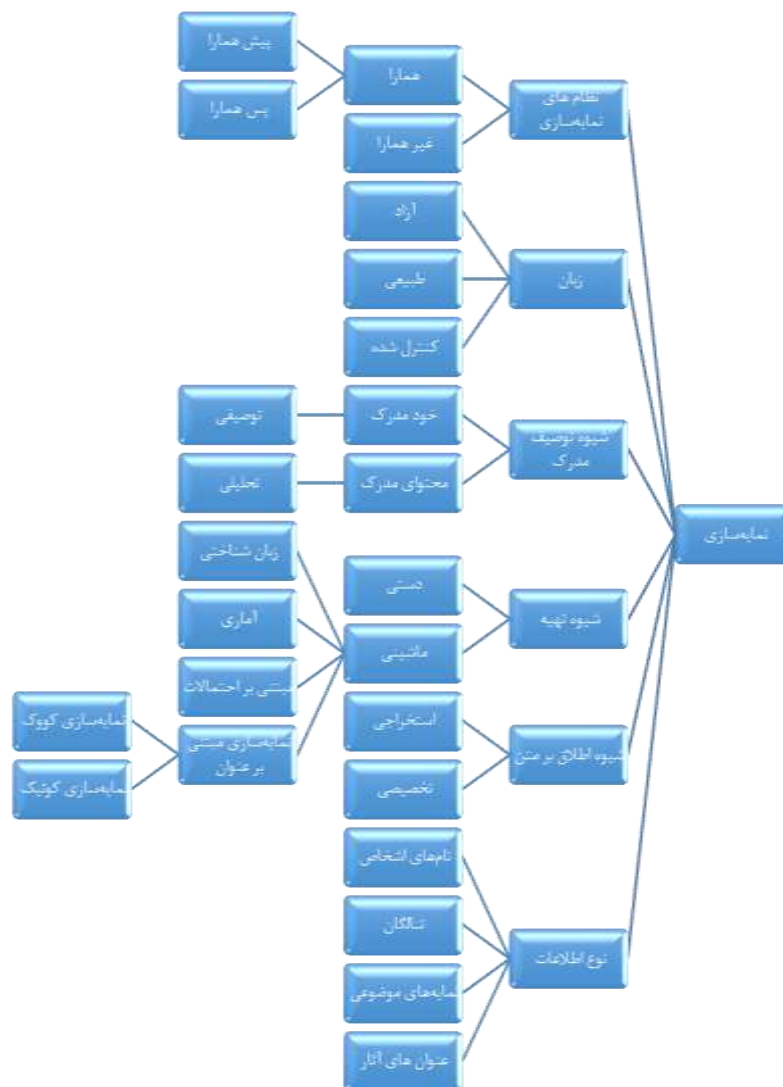
از سایر طرح‌های نمایه‌سازی ماشینی می‌توان طرح کتابخانه ملی پزشکی در آمریکا^۳، طرح دانشگاه فنی دارمشتات با عنوان «تحلیل عنوان و چکیده مدارک با موضوع فیزیکی برای پایگاه اطلاعاتی فیزیکی^۴، طرح‌های شرکت زیمنس^۵، طرح کتابخانه ملی پزشکی آمریکا برای نمایه‌سازی نشریات پزشکی^۶، تحلیل عنوان و چکیده مدارک علمی برای نمایه‌سازی مدارک مرکز دکومانناسیون آمریکا^۷، سیستم بازیابی هوشمند اطلاعات در مدارک بیولوژی و پزشکی بر پایه روش‌های آماری و زبان‌شناسی در دانشکده پزشکی هاروارد^۸ و سیستمی برای تحلیل اطلاعات بیولوژیکی و پزشکی بر مبنای تکنیک‌های استفاده از زبان طبیعی از طریق دانشگاه ملی پزشکی آمریکا^۹ همچنین طرح گروه کارنگی در آمریکا برای نمایه‌سازی تحلیلی و فهرست‌نویسی بر مبنای ارتقای سیستم کانسروی تیس^{۱۰} را نام برد (نیاکان، ۱۳۸۳).

به هر روی با پیشرفت‌هایی که در این زمینه‌ها حاصل شده، تلاش انسان برای رسیدن به سطوح بالاتر روزافزون شده و تحقیقات و مطالعات امروزی جز موارد ذکر شده شامل موارد پیشرفته‌تری مانند تحلیل هوشمندانه متون نیز می‌شود.

اصول نمایه‌سازی ماشینی

نمایه‌سازی ماشینی از دو دستورالعمل عمده و کلی پیروی می‌کند: اول تحلیل آماری متن و دوم تحلیل نحوی متن.

1. Salton
۲. PASSAT
3. AIMS
4. AIR/X
5. ISA, COPS
6. Indexing Aid system
7. MAI
8. SAPHIR
9. SPECIALIST
10. CONSTRVE/TIS



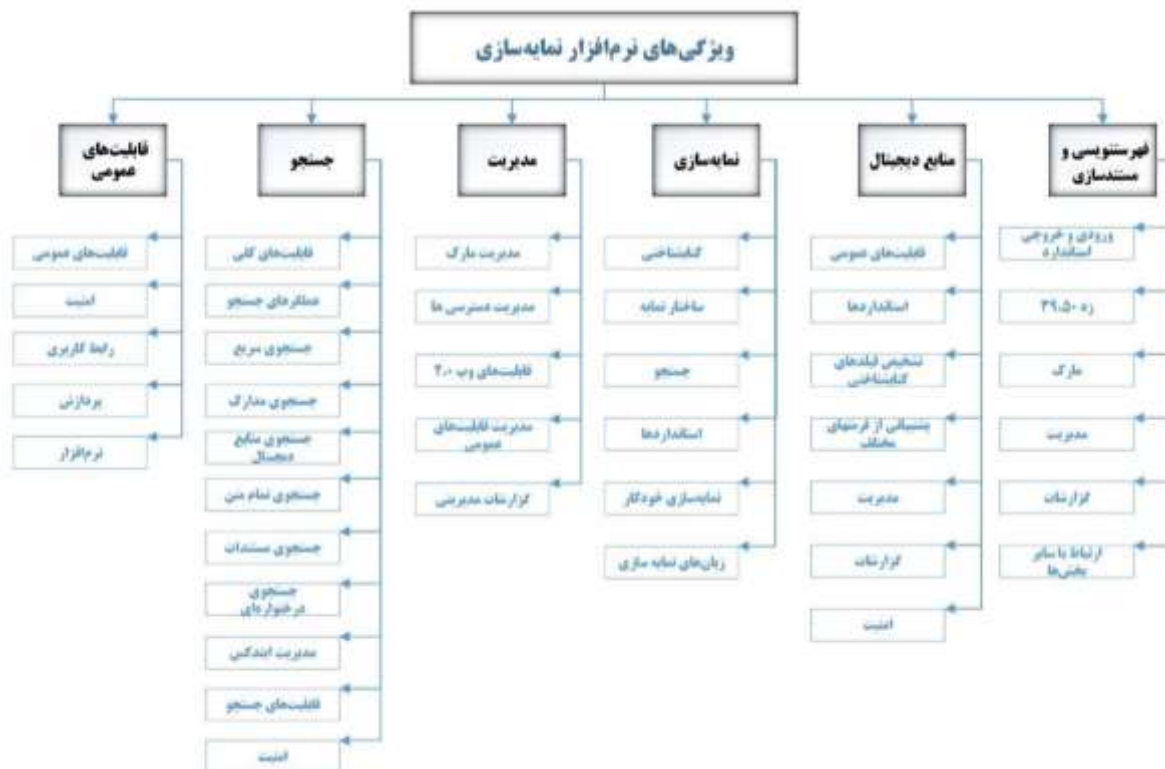
نمودار ۱. انواع روش‌های نمایه‌سازی

ارائه الگوی نرم‌افزار نمایه‌سازی خودکار

می‌توان گفت یک نرم‌افزار نمایه‌سازی کاملاً خودکار به بخش‌های مختلفی از جمله فهرست‌نویسی جهت ورود اطلاعات کتابشناختی و سند‌شناختی، مستندسازی جهت ایجاد انواع مستندات و اصطلاحنامه‌ها، منابع دیجیتال و نمایه‌سازی جهت یکپارچگی منابع دیجیتال و نمایه‌های استخراج شده از آن‌ها، مدیریت کلی نرم‌افزار و جستجو جهت استفاده کاربران نیاز دارد. استفاده از استانداردهای مختلف در بخش‌های فهرست‌نویسی و منابع دیجیتال و همچنین استفاده از استانداردهای برنامه‌نویسی نیز جزو الزامات یک نرم‌افزار نمایه‌سازی است.

برای این کار ما نرم‌افزار را به بخش‌های مختلفی تقسیم کرده‌ایم که نرم‌افزار باید این بخش‌ها را دارا باشد و برای هر بخش زیر بخش‌هایی مشخص شده است. (تصویر ۱) در این پژوهش، کلیه ویژگی‌هایی که اعلام شده است، مواردی هستند که باید در یک نرم‌افزار نمایه‌ساز وجود داشته باشند. در صورتی که پیاده‌کننده این ابزار، می‌تواند از سایر قابلیت‌ها و موارد نیز استفاده کند تا نرم‌افزاری جامع‌تر و کامل‌تر در اختیار جامعه استفاده‌کننده قرار دهد. همچنین در این الگو از استانداردهای پیشنهادی برای ورود اطلاعات، افزودن منابع دیجیتال، نحوه نمایه‌سازی و جستجو، استفاده شده است. قطعاً با به‌روز شدن استانداردها، می‌توان از استانداردهای جدید استفاده کرد. به‌عنوان نمونه در بخش فهرست‌نویسی و

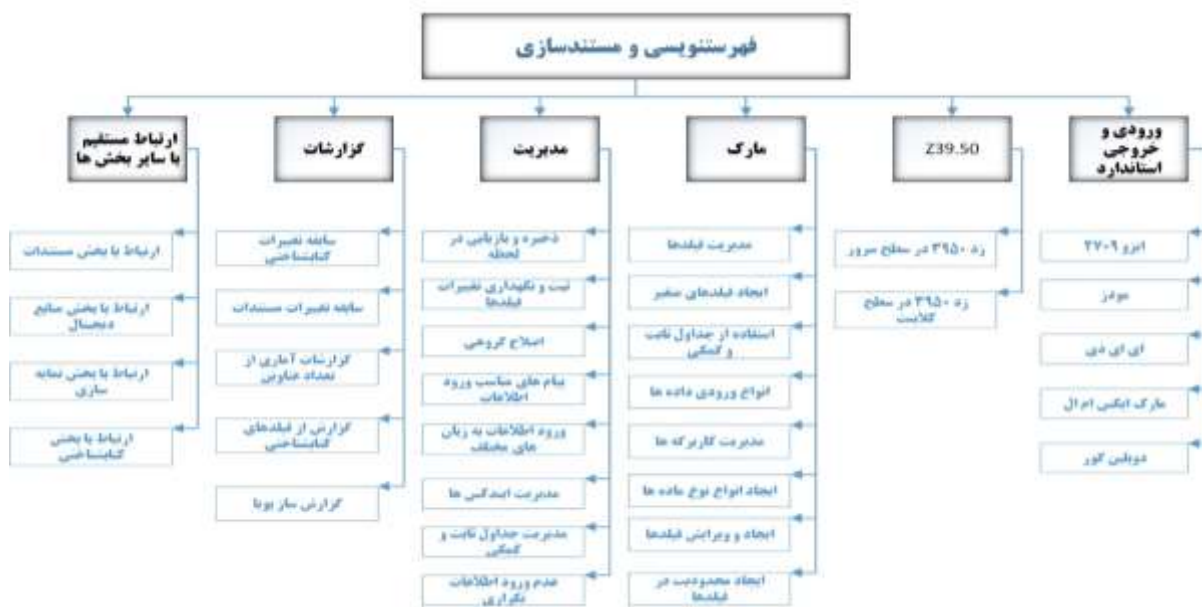
مستندسازی، مارک به‌عنوان یک استاندارد بومی‌سازی شده، یاد شده است.



شکل ۱. الگوی کلی نرم‌افزار نمایه‌سازی

کردن به نرم‌افزار را داشته باشند. هرچقدر این ورودی‌ها بیشتر باشند یا از استانداردهایی استفاده شود که استفاده بیشتری دارند، نیازهای بیشتری پاسخگو خواهد بود.

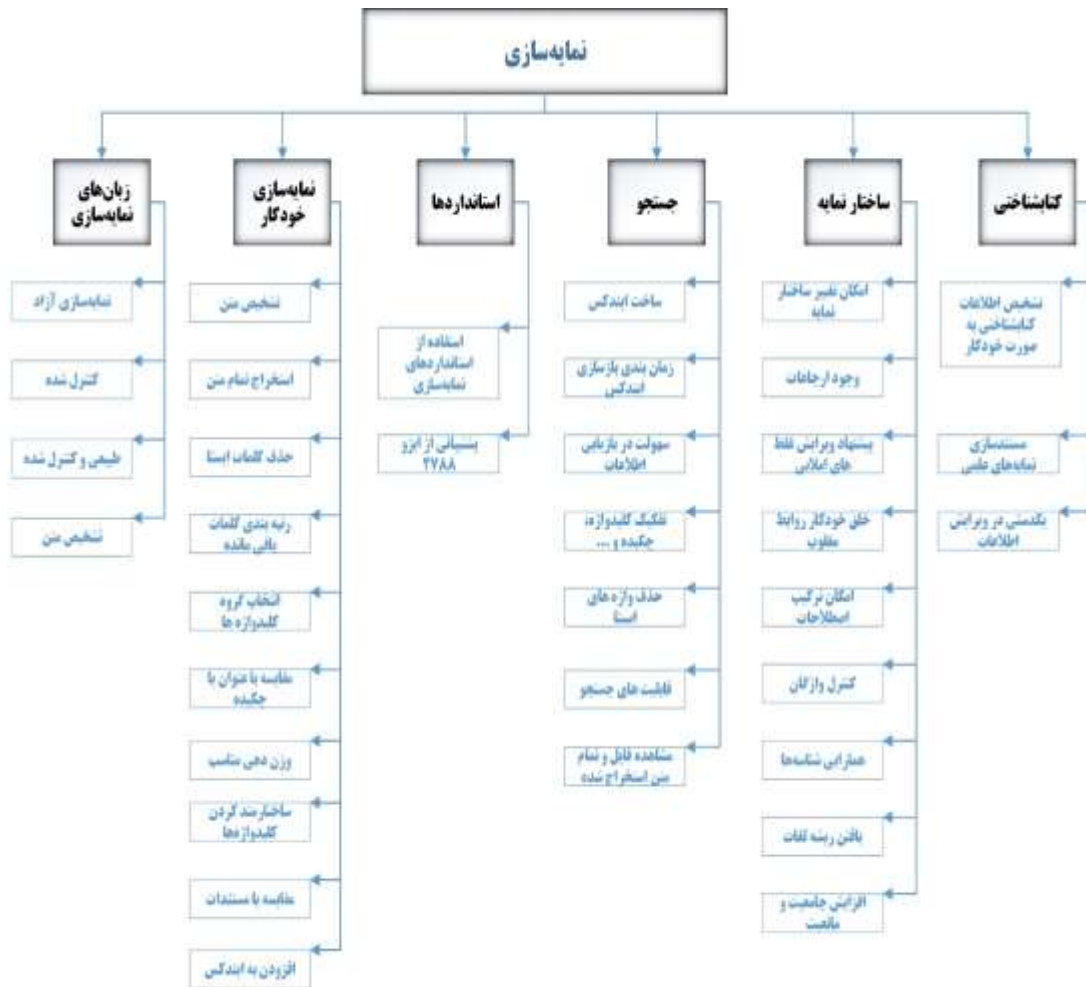
هر نرم‌افزار نیاز به بخشی دارد که منابع را وارد نرم‌افزار کرده و اطلاعات آن‌ها را ثبت کند این بخش را می‌توان بخش فهرست‌نویسی و مستندسازی نامید (شکل ۲) در این الگو، ورودی‌های مختلفی برای نرم‌افزار در نظر گرفته شده که این امکان را می‌دهد که اطلاعات در فرمت‌های مختلف قابل وارد



شکل ۲. الگوی بخش فهرست‌نویسی و مستندسازی

همچنین استفاده از وب‌سرویس‌های مختلف جهت انتقال مستقیم اطلاعات از پایگاه داده هم باید در این نرم‌افزار دیده شود (ریکاردو^۶، ۲۰۰۴).

جهت انتقال اطلاعات از یک سامانه به سامانه دیگر، وجود استانداردهای ایکس ام ال^۱، مارک^۲، طرح آرشیوی باز^۳، زد ۴۳۹/۵۰^۴، طرح فراداده‌ای توصیف شیء^۴، استاندارد توصیف آرشیوی^۵ رمزگذاری شده، استاندارد کدگذاری و انتقال متادیتا و



شکل ۳. الگوی بخش نمایه‌سازی

در این بخش، علاوه بر رعایت استانداردهای منابع دیجیتال، مانند متس^۷ باید از استانداردهای نمایه‌سازی مانند ایزو ۵۹۶۳ نیز استفاده نماید. استانداردهایی مانند متس، اطلاعات منابع دیجیتال را بین پایگاه‌های مختلف قابل انتقال می‌کند. این استاندارد از ۷ بخش اصلی تشکیل شده است.

در قدم اول، یک نمایه‌ساز خودکار باید قابلیت افزودن منابع دیجیتال را داشته باشد. با توجه به این مهم که حجم منابع دیجیتال می‌تواند بسیار زیاد باشد، بنابراین پایگاه داده و یا فایل سیستمی باید قابلیت ذخیره این حجم داده را داشته باشد.

از قابلیت‌های دیگر نرم‌افزار نمایه‌سازی خودکار این است که باید بتواند بخش‌های مختلف فایل دیجیتال را به شکل مستقل شناسایی کند. به عبارت دیگر، باید بتواند عنوان، چکیده و متن را از منبع دیجیتال شناسایی و تفکیک نماید و آن اطلاعات را در جایگاه مخصوص خود در فیلدهای فهرست‌نویسی قرار دهد.

1. XML = eXtensible Markup Language
 2. MARC = Machine Readable Cataloging
 3. OAI = Open Archival Initiative
 4. MODS = Metadata Object Description Schema
 5. EAD = Encoded Archival Description
 6. Ricardo
 7. METS = Metadata Encoding and Transmission Standard

همان‌طور که مشاهده می‌شود، در مرحله اول، جداسازی و حذف کلمات غیر مرتبط رخ می‌دهد. با استفاده از تکنولوژی‌های استخراج محتوا^۲ کلیه اطلاعات یک فایل متنی، استخراج می‌شود و موارد غیر متنی مانند جداول و تصاویر از متن اصلی حذف می‌شوند و نتیجه این مرحله تنها یک فایل متنی ساده^۳ خواهد بود و سپس از طریق تکنولوژی حذف کلمات زائد^۴، علائم نگارشی از متن حذف می‌شود. علت حذف علائم نگارشی در جستجوها توسط کاربر است که تقریباً در همه حالات، کاربر علائم نگارشی را تایپ نمی‌کند. بعد از حذف علائم نگارشی و تصاویر و جداول نوبت به حذف اطلاعات حشو در اطلاعات استخراج شده می‌رسیم. برای این کار باید از یک لیست استاندارد از کلمات زائد^۵ که در هر زبانی وجود دارد استفاده کرد. با این کار همه کلمات زائد از فایل اصلی هم خارج می‌شوند. در مرحله دوم نیاز است که حتماً کلمه به ریشه اصلی خود برگردانده شود. اگر لغات متن به ریشه تبدیل نشوند، دو لغت کاملاً مشابه و هم‌معنی، دو کلمه مجزا در نظر گرفته می‌شوند. روش‌های مختلفی برای پیدا کردن ریشه لغات وجود دارد که یکی از آن روش‌ها به نام پورتر^۶ می‌باشد.

در مرحله بعد زمان پردازش داده‌ها و استخراج کلمات کلیدی فرا می‌رسد. در حال حاضر با یک لیست بزرگ از کلمات و عبارات روبرو هستیم که نیاز به پردازش دارد. براساس الگوریتم‌های مشخص و با در نظر گرفتن آستانه خاصی، می‌توان عباراتی که کمتر در این لیست استفاده شده است را حذف کرد (دهقانی محمدی، ۱۳۹۲).

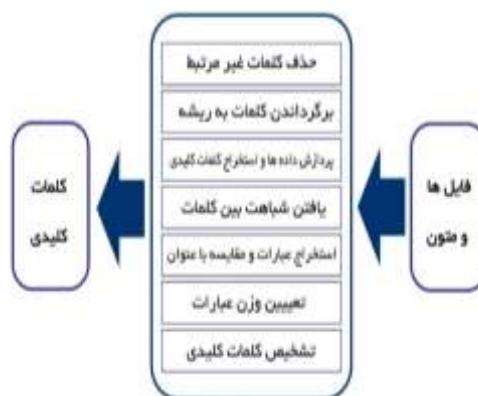
یکی از مهم‌ترین بخش‌های نمایه‌سازی در این ساختار، بخش پیدا کردن مشابهت‌ها است. پیدا کردن مشابهت‌ها به معنی یافتن لغات مشابه و مقایسه آن‌ها با مستندات است که در بخش مستندسازی وجود دارد. یکی از روش‌های پیدا کردن مشابهت‌ها می‌تواند مقایسه عبارت‌های استخراج شده با عنوان مقاله باشد. از آنجایی که هر ریشه لغت می‌تواند معانی مختلفی داشته باشد، باید این ریشه‌ها با عنوان مقایسه شود و نزدیک‌ترین حوزه برای دو عبارت که با هم شباهت دارند انتخاب شود. در واقع نرم‌افزاری مانند وردنت^۷ می‌تواند کلمه سرشاخه دو عبارت مختلف را از اصطلاحنامه استخراج نماید.

در یکی از این بخش‌ها، اطلاعات کتابشناختی مدرک، که در قالب استاندارد مارک یا دابلین کور ارائه شده، ذخیره می‌شود. بنابراین حتماً باید ورودی و خروجی‌های استاندارد مارک یا دابلین کور در این نرم‌افزار لحاظ شده باشد.

یک نرم‌افزار نمایه‌سازی، جهت ارائه دقیق اطلاعات به کاربران، باید بیشترین دقت در استخراج اطلاعات مفید از یک منبع دیجیتال را دارا باشد. به همین جهت باید روشی برای استخراج نمایه‌ها از منبع دیجیتالی وجود داشته باشد. ورودی فایل دیجیتال باید براساس استانداردهای خاصی باشد تا دقیق‌ترین نمایه‌ها از آن‌ها استخراج شود.

امروزه برای فایل‌های متنی نیز استانداردهای خاصی ارائه شده است که از جمله آن‌ها می‌توان به استاندارد پی دی اف / ای^۱ اشاره کرد. این فایل‌های متنی، نوعی از فایل‌های پی دی اف هستند که می‌توانند آن‌ها را به مدت طولانی نگهداری کنند. در واقع حرف «ای» در آخر این نوع فایل‌ها به معنای آرشیو است. نخستین شماره این استاندارد با عنوان پی دی اف / ای ۱ در سال ۲۰۰۵ میلادی توسط شرکت ادوبی منتشر شد و با عنوان ایزو ۱۹۰۰۵ به صورت بین‌المللی مورد تأیید قرار گرفت. هر کدام از نسخه‌های این استاندارد، ضمن پیشرفت، قابلیت نگهداری فایل‌های مختلفی در فایل پی دی اف را دارا شدند و در سال ۲۰۱۸ آخرین نسخه از این استاندارد با شماره پی دی اف / ای ۴ منتشر شد. علاوه بر اینکه استانداردهای فایل‌های پی دی اف باعث نگهداری فایل متنی به مدت طولانی می‌شوند، امکان استخراج محتوا را نیز برای آن‌ها فراهم می‌کند. فایل‌های متنی باید امکان استخراج محتوا را داشته باشند.

نرم‌افزار نمایه‌ساز طبق کدهای خاصی امکان استخراج محتوا از فایل متنی را دارد. استخراج اطلاعات از فایل‌های متنی طبق تصویر زیر، دارای مراحل مشخصی باید باشد.



شکل ۴. معماری پردازش و استخراج اطلاعات

2. Content Extraction
3. Plain Text File
4. Symbol Reduction
5. Stop List
6. Porter
7 wordNet

1. PDF/A

تعیین واژه‌های ایستا، بازسازی خودکار اندیس، بازیابی اشتباهات، پیشنهاد اشتباهات و غلط‌گیری از نمایه‌سازی ایجاد شده و را در خود داشته باشد.

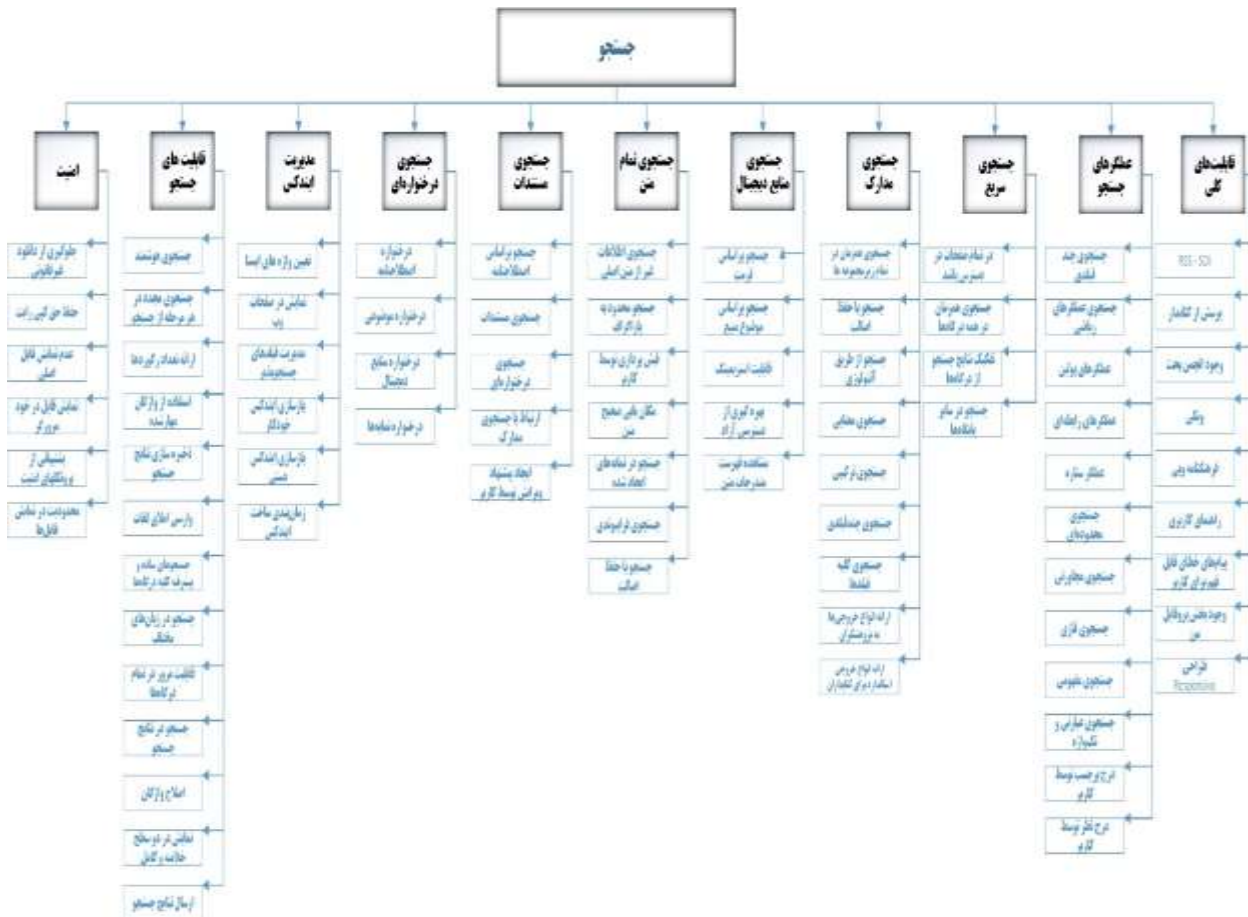
کلیه موارد اعلام شده مربوط به مراحل آماده‌سازی منابع دیجیتال برای استفاده کاربران بود. از نظر دیگر، کاربران نیز باید امکان استفاده از قابلیت‌های این نرم‌افزار را نیز داشته باشند تا حداکثر استفاده از اطلاعات استخراج شده توسط کاربر انجام شود.

کاربران نیاز به انجام انواع جستجوها در نرم‌افزار نمایه‌سازی را خواهند داشت. جستجوهای نظیر اطلاعات کتابشناختی، جستجوی منابع دیجیتال، جستجوی نمایه‌های ایجاد شده، جستجوی تمام متن، جستجو در اصطلاحنامه و جستجو به صورت هم‌زمان در کلیه این اطلاعات. بنابراین صفحات کاربری باید قابلیت جستجو در همه این درگاه‌ها را با هم داشته باشد. علاوه بر این درگاه‌ها باید جستجوی هم‌زمان در همه این درگاه‌ها را داشته باشد و نتیجه را در قالب بخش‌های مختلف به کاربر ارائه دهد.

مرحله بعد، تعیین بهترین وزن برای عبارات می‌باشد. براساس تعداد تکرار لغات و مشابهت‌هایی که در یک متن به کار رفته است، امکان وزن‌دهی برای آن‌ها فراهم می‌شود. به این صورت می‌توان موضوعات اصلی را از متن یک مقاله یا پایان‌نامه یا به صورت کلی یک مدرک استخراج کرد. این بخش نیز مانند بخش قبل نیازمند الگوریتم‌های خاص کامپیوتری می‌باشد.

بخش نمایه‌سازی باید بتواند روش‌های نمایه‌سازی آزاد، کنترل شده و ترکیبی از این دو را داشته باشد. با توجه به موارد گفته شده، هنگامی که کلمات کلیدی استخراج می‌شوند، در صورتی که ارتباط مستقیم با بخش مستندسازی داشته باشند، می‌توانند کلمات کلیدی را به لغات مستند شده تبدیل نمایند (نمایه‌سازی کنترل شده). ضمناً استخراج لغاتی که در متن وجود دارند اما در مستندات وجود ندارند نیز امکان نمایه‌سازی را ایجاد می‌کنند (نمایه‌سازی آزاد یا طبیعی). ترکیبی از این دو روش می‌تواند بهترین کیفیت در نمایه‌سازی را ارائه دهد.

بعد از استخراج محتوا، یکی از مهم‌ترین بخش‌های نمایه‌ساز خودکار، آماده‌سازی اطلاعات برای جستجوی کاربران است. بخش نمایه‌سازی، باید امکان تغییر ساختار نمایه، امکان

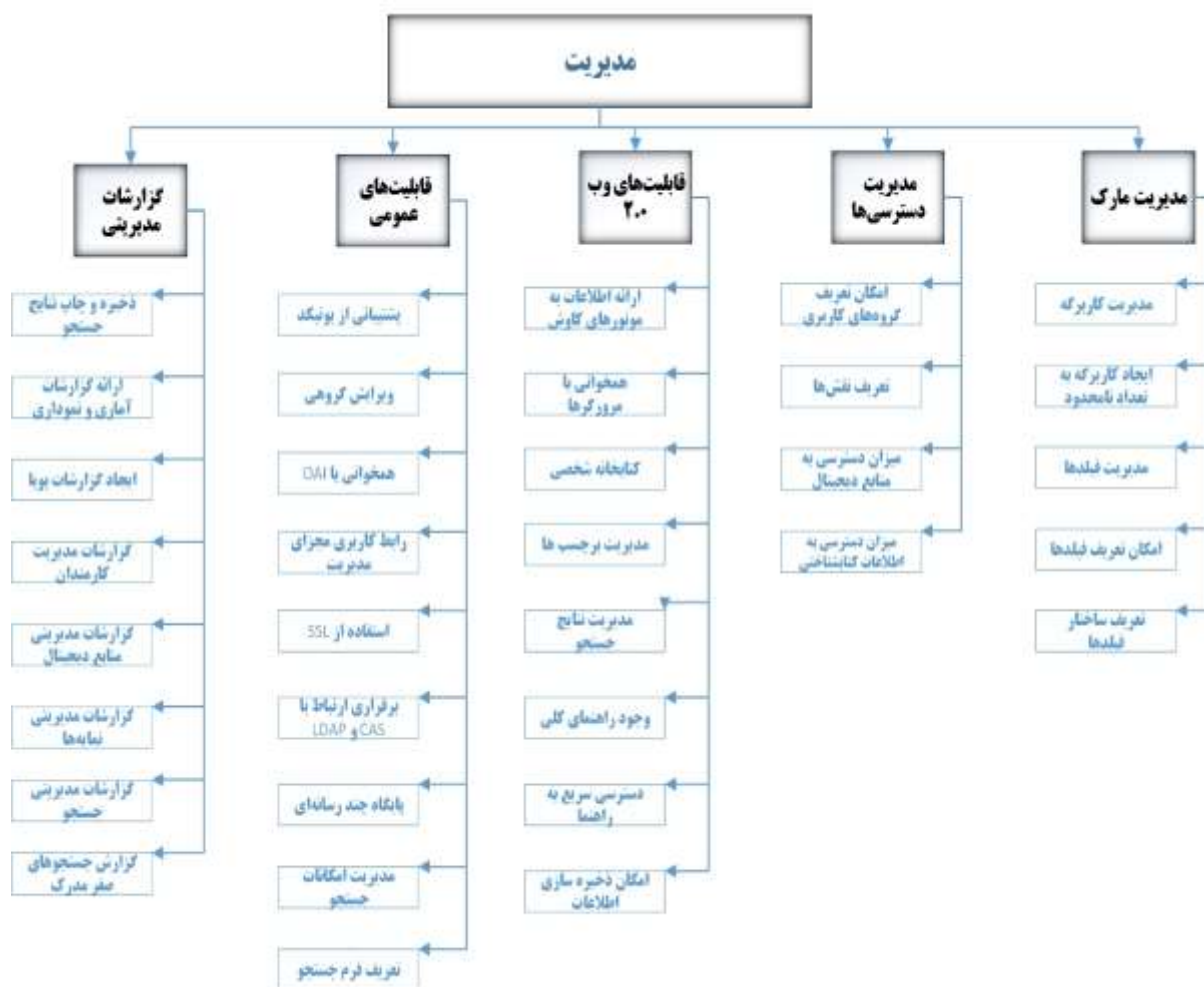


شکل ۵. الگوی بخش جستجو

باشد. امکان مشاهده رکورد، منبع دیجیتال، دسترسی به چند صفحه یا همه صفحات و همچنین امکان بارگذاری منبع و جستجو در متن فایل‌های متنی باید با دسترسی‌های مشخصی قابل کنترل باشند. جهت جلوگیری از داندلود غیرقانونی منابع باید بیشترین امنیت بر روی این منابع اعمال شود. برای حل این مشکل استانداردهای امنیتی زیادی تعریف شده‌اند که حتماً پیاده‌سازی یک یا مجموعه‌ای از این استانداردها با سطح امنیتی بالا پیشنهاد می‌شود.

هر نرم‌افزار نیاز به بخش مدیریتی دارد که بتواند فعالیت‌های نرم‌افزار را مدیریت کند. جهت هماهنگی بخش‌های مختلف و انجام تنظیمات موردنیاز هر مرکز و اینکه این بخش باید به صورت کامل از بخش کاربران جدا باشد و کاربران به این بخش دسترسی نداشته باشند. در مدیریت سیستم تنظیمات زیر باید وجود داشته باشد.

کاربرانی که کارهای پژوهشی انجام می‌دهند، گاهی نیاز به فیش‌برداری خواهند داشت. بنابراین نرم‌افزار باید بتواند فیش‌برداری، که از متن استخراج شده انجام می‌شود را، در جایی ذخیره کند و در کل این قابلیت را در اختیار کاربران قرار دهد. همچنین اطلاعات کاربر مانند منابع دیجیتالی که مشاهده کرده یا سابقه جستجوهای که انجام داده و ذخیره کرده، یا رکوردهای کتابشناختی که نشانه‌گذاری کرده برای مطالعه در زمان مناسب ذخیره کند. این موارد به این معناست که کاربر باید بخشی مانند «پروفایل من» داشته باشد که اطلاعات جامع و کاملی از تاریخچه و کارهایی که انجام داده است مشاهده نماید. در بخش جستجو، مشاهده برخی از منابع، فقط برای برخی از کاربران آزاد است و برای بقیه کاربران ممکن است ممنوع باشد. بنابراین مشاهده منابع دیجیتال باید براساس دسترسی‌های خاصی محدود شود. این دسترسی‌ها باید، تا سطح یک کاربر هم ریز شود و کنترل آن کاملاً در اختیار مرکز



شکل ۶. الگوی بخش مدیریت

کلاینت ۱ گفته می‌شود. معمولاً نرم‌افزارهای تحت وب به هیچ ابزار کمکی غیر از ابزارهای رایج در ویندوز نیاز ندارند. به عبارت دیگر، نصب هیچ‌گونه نرم‌افزار جانبی بر روی کلاینت‌ها نباید صورت بگیرد.

از دیگر قابلیت‌های نرم‌افزار تحت وب می‌توان به همیشه در دسترس بودن، سادگی استفاده، سرعت و امنیت بالا، نیاز به کلاینت با حداقل سخت‌افزار، قابل اجرا در همه سخت‌افزارها مانند تبلت، موبایل و لب تاب و ... را اشاره کرد.

یک نرم‌افزار نمایه‌ساز، با توجه به اطلاعاتی که در آن وارد می‌شود، باید از زبان‌های مختلف پشتیبانی کند. یکی از بهترین استانداردهایی که می‌تواند همه زبان‌های موجود در دنیا را پشتیبانی نماید، استاندارد یونیکد ۲ است. حداقل زبان‌هایی که یک نرم‌افزار کتابخانه و نمایه‌ساز باید داشته باشد، فارسی، عربی و انگلیسی است. بقیه زبان‌ها نیز باید قابلیت پیاده‌سازی در مدت‌زمان کوتاهی را داشته باشند.

بخش بعدی که در مدیریت سیستم خیلی اهمیت دارد، دارا بودن بخش راهنما در نرم‌افزار است. کاربران و کارمندان، علاوه بر دریافت پشتیبانی از مرکز ایجادکننده نرم‌افزار باید بتوانند با مطالعه راهنمای محصول، تا حد زیادی نیاز خود را برطرف نمایند.

بحث و نتیجه‌گیری

با توجه به موارد ذکر شده، مشاهده می‌شود که برخلاف تصور بسیاری که گمان می‌برند طراحی چنین نرم‌افزاری امکان‌پذیر نمی‌باشد، پیاده‌سازی و راه‌اندازی یک نرم‌افزار نمایه‌سازی می‌تواند به صورتی که گفته شد، پیاده‌سازی شود. در واقع روش‌های مختلفی برای ایجاد این نرم‌افزارها وجود دارد (موهان^۳، ۲۰۱۶). ضمناً اگر به کلیه موارد گفته شده، قابلیت‌های هوش مصنوعی نیز اضافه شود، پردازش اطلاعات موردنیاز کاربران بسیار تسهیل‌تر و در دسترس‌تر خواهد بود.

در برخی از نرم‌افزارهای کتابخانه دیجیتال داخلی، قابلیت‌های جستجوی تمام متن، نمایه‌سازی منابع دیجیتال به صورت دستی، جستجوهای پیشرفته منابع دیجیتال و استانداردهای کتابخانه‌ای قابل مشاهده است. اما بخش‌های اصلی نمایه‌سازی، از جمله حذف عبارات اضافی، انتخاب کلیدواژه مناسب، رتبه‌بندی کلیدواژه‌ها و در نهایت خروجی مناسب نمایه‌سازی از یک فایل متنی در نرم‌افزارها وجود ندارد.

تنظیمات کاربرگه‌های ورود اطلاعات که توسط مدیر یا مدیران سیستم تعریف می‌شوند. قبلاً اعلام شد که قابلیت ایجاد به تعداد نامحدود از کاربرگه‌ها باید وجود داشته باشد که در مدیریت سیستم، این کاربرگه‌ها و اطلاعات درون آن‌ها قابل تعریف و پیاده‌سازی است.

نرم‌افزار باید امکان ورود اطلاعات مدارک مختلف کتابخانه‌ای شامل نشریه، کتاب، مقاله و ... یا مدارک غیر کتابخانه‌ای شامل انواع اسناد فنی و کتابخانه‌ای، لوح‌های فشرده، نسخ خطی و ... را داشته باشد. در واقع هر نوع ماده‌ای که در سراسر دنیا وجود دارد، باید بتواند در قالب یک کاربرگه وارد سیستم شود.

همچنین اطلاعاتی که قرار است به کاربر در صفحه نمایش کامل رکورد نشان داده شود نیز باید قابل تعریف باشد. از آنجایی که امکان تعریف کاربرگه‌های مختلف با فیلدهای مختلفی وجود دارد، باید این امکان نیز وجود داشته باشد که اطلاعات این فیلدها نمایش داده شود. علاوه بر نمایش اطلاعات، در هنگام جستجو باید کلیه فیلدهایی که قابلیت ایجاد دارند نیز قابل جستجو شوند و در ایندکس قرار گیرند.

قابلیت‌های دیگری که در بخش مدیریت نیاز به کنترل دارند و باید در نرم‌افزار وجود داشته باشند می‌توانند شامل امکان اصلاح گروهی، امکان تعریف سطوح دسترسی کاربران و مدیریت کاربران باشند. کلیه دسترسی‌هایی که هر کاربر می‌تواند از آن برخوردار باشد، در بخش مدیریت باید تعریف شوند. همچنین مدیریت کاربران شامل مشاهده میزان دانلودها و استفاده‌های هر کاربر به صورت مجزا و به صورت کلی نیز از قابلیت‌هایی است که در بخش مدیریت کاربران باید وجود داشته باشد.

ساختار کلی یک نرم‌افزار، پایه و اساس آن است و مشخص می‌کند این نرم‌افزار بر چه اساسی کار می‌کند و چگونه قابل دسترس خواهد بود.

مشخص‌ترین قابلیت‌هایی که باید در این نرم‌افزار دیده شود، تحت وب بودن است. تحت وب بودن با تحت اینترنت بودن تفاوت‌های زیادی دارد. نرم‌افزار تحت وب به نرم‌افزاری گفته می‌شود که قابلیت اجرا تاز طریق مرورگرهای رایج را چه در محیط اینترنت و چه در محیط شبکه داخلی داشته باشد. نرم‌افزارهای تحت وب معمولاً دارای یک سرور مرکزی هستند که همه تغییرات در آن رخ می‌دهد. به سیستم‌هایی که کاربران و کارمندان از طریق آن‌ها به سرور متصل می‌شوند، اصطلاحاً

پیشنهادها

پیشنهاد می‌شود الگوی ارائه شده برای طراحی یک نرم‌افزار نمایه‌ساز در این پژوهش، به صورت عملیاتی و در محیط اصلی پیاده‌سازی شود.

ساخت یک نرم‌افزار نیازمند زیرساخت‌هایی است از جمله منابع کاملاً دیجیتال و همچنین منابع مالی جهت گسترش آن در تمام سازمان‌ها و نمایه‌سازی خودکار منابع است.

امروزه با ظهور هوش مصنوعی و با استفاده از آن در نرم‌افزار نمایه‌سازی می‌توان میزان دقت و کارآمدی را در این‌گونه نرم‌افزارها بالا برد.

سپاسگزاری

از معاونت محترم پژوهشی دانشگاه شهید بهشتی به خاطر حمایت معنوی در پژوهش حاضر سپاسگزاری می‌شود همچنین از داوران محترم که در ارائه نظرهای علمی بنده را یاری کردند قدردانی می‌نمایم.

معیارهای مهمی که باید در یک نرم‌افزار نمایه‌ساز وجود داشته باشد، اما در حال حاضر در وضعیت نامطلوب به سر می‌برند، مانند تشخیص اطلاعات کتابشناختی، بررسی قالب‌های فایل‌های متنی، ارجاع به مستندات ایجاد شده در سیستم، پیشنهاد تغییر نمایه، استفاده از استانداردهای نمایه‌سازی، خلق خودکار انواع روابط مقلوب، تشخیص ریشه لغات، ایجاد ایندکس نمایه، زمان‌بندی ایجاد ایندکس و ... باید در یک نرم‌افزار نمایه‌ساز وجود داشته باشد و با وجود آن‌ها نرم‌افزار کارایی بهتری خواهد داشت.

فناوری‌های جدید که پیامدهای بالقوه‌ای برای یادگیری دارند تقریباً روزانه در حال توسعه هستند. اما قطعاً مدیران باید در مورد استفاده کردن یا توسعه دادن این فناوری‌ها تصمیم‌گیری کنند (اوستون، ۲۰۰۸) به همین دلیل باید این فناوری‌های جدید مورد بررسی قرار گیرند. در این مقاله معرفی سطحی از قابلیت‌هایی که باید در یک نرم‌افزار نمایه‌ساز وجود داشته باشد، عنوان شد. قطعاً ارائه اطلاعات کامل در مورد جزئیات هر پارامتر نیاز به نوشتن کتاب‌های مختلفی دارد. همچنین بررسی این قابلیت‌ها هم بعد از پیاده‌سازی آن‌ها الزامی خواهد بود.

References

- Abusba Kazemini, A. (2010). *Comparison of comprehensiveness and hindrance of retrieved information based on pre-Hamara and post-Hamara storage systems in Persian library software*. Master's thesis. Faculty of Psychology and Educational Sciences. University of Esfahan. (In Persian)
- Andersen, J. (2004). *Analyzing the role of knowledge organization in scholarly communication: an inquiry into the intellectual foundation of knowledge organization*. Ph.D. dissertation. Copenhagen: Royal School of Library and Information Science.
- Chowdhury, G. (2004). Indexing and Abstracting in Theory and Practice (3rd ed.). *The Electronic Library*, 22(4), 364-365.
- Dehghani Mohammadi, A. (2012). *Providing a method for automatically extracting keywords from scientific texts and implementing a template example*. Master's thesis. Faculty of electronic education. Shiraz University. (In Persian)
- Ertz, D., Guzew-Krzemińska, B., Thor, G., Lúbek, A., & Kukwa, M. (2018). Photobiont switching causes changes in the reproduction strategy and phenotypic dimorphism in the Arthoniomycetes. *Scientific Reports*, 8(1), 1-14.
- Fidel, R. (1945). American Society for Information Science (1994). *Challenges in indexing electronic text and images*. Published for the American Society for Information Science by Learned Information, Medford, NJ
- Golub, K. (2015). Evaluating automatic subject indexing: A framework. In *7th ISKO Italy Meeting, Bologna, 20 April, 2015*.
- Golub, K. (2016). Semi-automated subject indexing of Swedish resources. *International Conference on Dublin Core and metadata applications*. Linnaeus University.
- Golub, K., & Soregel, D. (2016). A framework for Evaluating Automatic Indexing or classification in the context of retrieval. *Journal of the association for information science and technology*, 67(1), 3-16.
- Golub, K., Lykke, M., & Tudhope, D. (2014). Enhancing social tagging with automated keywords from the Dewey decimal classification. *Journal of Documentation*, 70(5), 801-828.
- Golub, K., Lykke, M., Tudhope, D. (2014). Enhancing social tagging with automated keywords from the Dewey decimal classification. *Journal of Documentation*, 70(5), 801-828

- Golub, K., Soergel, D., Buchanan, G., Tudhope, D., Lykke, M., & Hiom, D. (2016). A framework for evaluating automatic indexing or classification in the context of retrieval. *Journal of the Association for Information Science and Technology*, 67(1), 3-16.
- Guillory, A. (1379). Automatic indexing (past, present, future). *Information research and public libraries (Payam former library)*, 10(4), 17-25. (In Persian)
- Jacquemin, C., Daille, B., Royauté, J., & Polanco, X. (2002). In vitro evaluation of a program for machine-aided indexing. *Information processing & management*, 38(6), 765-792.
- Jalali Manesh, A., Alidosti, S., & Khosrowjerdi, M. (2012). Machine Indexing of Persian Sources: An Integrated Model for the Research Institute of Information Science and Technology of Iran. *Information processing and management research paper*. 29 (2): 425-451. (In Persian)
- Mirjalili, H. (1379). Investigating how to policy and use index language in the internal databases of information centers in Tehran. 10 (49). 42-51. (In Persian)
- Mohan, M. J., Sunitha, C., Ganesh, A., & Jaya, A. (2016). A study on ontology based abstractive summarization. *Procedia Computer Science*, 87, 32-37. Mohan, M. J., Sunitha, C., Ganesh, A., & Jaya, A. (2016). A study on ontology based abstractive summarization. *Procedia Computer Science*, 87, 32-37.
- Montejo Ráez, A. (2002). *Toward conceptual indexing using automatic assignment of descriptors* (No. CERN-ETT-2002-005).
- Nowrozi, A., & Velayati, K. (1389). *Thematic indexing: conceptual indexing*. Tehran: Chapar Publications. (In Persian)
- Nowrozi, Y., & Nemati, S. (1389). Evaluation of comprehensive online library software of Pars Azarakhsh, Noosa and Profile in information retrieval. *Information research and public libraries*. 16 (60). 23-43. (In Persian)
- Nyakan, SH. (2013). *Machine indexing*. Tehran: Center for Scientific Information and Documents of Iran. (In Persian)
- Owston, R. (2008). *Models and Methods for Evaluation. Handbook of Research on educational communications and technology (3rd Ed.)* New York: Routledge.
- Ricardo da Silva (2004). An OAI compliant Content based image search component. *In digital libraries*.
- Salton, G. (1971). Automatic indexing using bibliographic citations. *Journal of Documentation*.
- Tavaklizadeh Rawri, M. (2014). Two-stage gap model for automatic indexing of Persian texts. *Information Research and Public Libraries*, 21(1), 40-12. (In Persian)
- Torres, R. D. S., Medeiros, C. B., Gonçalves, M. A., & Fox, E. A. (2004, June). An OAI compliant content-based image search component. In *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004*. (p. 418). IEEE.
- Yousefi, A. (1377). Principles and methods of computer indexing. *Book Quarterly*, 9(2), 48-60. (In Persian)