



نقش نمایه‌سازی در بازیابی اطلاعات

○ محمد حسن زاده

دانشجوی دکتری دانشگاه فردوسی مشهد

○ بررسی تأثیر شیوه‌های نمایه‌سازی بر قابلیت بازیابی منابع مشخص و موجود در سه نرم‌افزار کتابخانه‌ای فارسی (پایان‌نامه کارشناسی ارشد)

○ دانشگاه: فردوسی مشهد

○ سال: ۱۳۸۰

○ پژوهشگر: غلامرضا سلگی

○ استاد راهنما: دکتر رحمت‌الله فتاحی

○ استاد مشاور: دکتر محمدرضا داورپناه

در راستای دستیابی به نظام ذخیره و بازیابی مطلوب اطلاعات در کتابخانه‌ها و مراکز اطلاع‌رسانی ضروری است. در دهه گذشته چندین تحقیق در این باره انجام شده است که از بین آنها به «بررسی مشخصات نرم‌افزارهای اطلاع‌رسانی و پیشنهاد الگوی بهینه» توسط شهرام الماسی، «بررسی ساختار و وضعیت ذخیره و بازیابی اطلاعات در پایگاه‌های اطلاعاتی (مرکز اطلاعات و مدارک علمی ایران و سازمان پژوهش‌های علمی و صنعتی ایران)» توسط محمود بابایی، «ریزش کاذب در نرم‌افزارهای کتابخانه‌ای نوسا، پارس آذرخش و کاوش» و غیره اشاره کرد. در این مقاله به تحقیق «بررسی تأثیر شیوه‌های نمایه‌سازی بر قابلیت بازیابی منابع مشخص و موجود در سه نرم‌افزار کتابخانه‌ای فارسی» پرداخته می‌شود. تحقیق مذکور برای پایان‌نامه کارشناسی ارشد رشته کتابداری و اطلاع‌رسانی در دانشگاه فردوسی مشهد انجام گرفته است. محقق در مقدمه با اشاره به پیشرفت کاربرد کامپیوتر در امر فهرست‌نویسی، به وجود آمدن فهرست‌های رایانه‌ای و OPAC ها در اختیار داشتن مطلوب‌ترین شیوه و ابزار برای آماده‌سازی، ذخیره و بازیابی و انتقال اطلاعات را از موارد مهم و مورد توجه در دنیای اطلاعات ارزیابی کرده است. در کشور ما نیز چند سالی است که طراحی و تولید پایگاه‌های اطلاعاتی توسط برخی شرکت‌ها، کتابخانه‌ها و مراکز اطلاع‌رسانی مانند فهرست رایانه‌ای کتابخانه (OPAC) و دیسک فشرده رشد چشمگیری داشته است. ولی به دلایل متعددی کار جدی در زمینه سنجش قابلیت بازیابی اطلاعات در رابطه با بکارگیری شیوه‌های نمایه‌سازی فیلدها انجام نگرفته است. بدین ترتیب همچنان خلأ بررسی نرم‌افزارهای کتابخانه‌ای موجود فارسی در ابعاد و زوایای گوناگون آن‌ها، به‌ویژه در زمینه شیوه‌های نمایه‌سازی فیلدها و تأثیر آن بر قابلیت بازیابی اطلاعات به

همگام با افزایش تولید اطلاعات و دانش و پیشرفت صنعت نشر و افزایش حجم مخازن کتابخانه‌ها طراحی نرم‌افزارهای کتابخانه‌ای نیز رشد چشمگیری داشته است. این رشد در پاسخ به نیاز کتابداران جهت پاسخگویی به پرسش‌های کاربران از طریق منابع موجود در مخزن کتابخانه خود و از طرف دیگر تمایل کاربران به دستیابی سریع و انتخاب منبع دلخواه خود از بین منابع انبوه کتابخانه اتفاق افتاده است و اکنون ما شاهد پیدایش نرم‌افزارها و برنامه‌های سازماندهی و ذخیره و بازیابی اطلاعات در سطح جهان و کشور هستیم. آنچه که مهم است این که این نرم‌افزارها باید تا حد توان از نتایج تحقیقات انجام شده در راستای اعتلای سطح کیفیت بازیابی پاسخ‌های درست و به دور از ریزش و یا افت اطلاعات استفاده کنند و از طرف دیگر تلاش برای آزمون قابلیت‌های این نرم‌افزارها از طرف متخصصان علوم کتابداری و اطلاع‌رسانی ضروری است. یکی از مواردی که در ارزیابی نرم‌افزارها مورد بررسی قرار می‌گیرد استفاده از شیوه‌های نمایه‌سازی است که می‌تواند نقش عمده‌ای در کارایی یک نظام ذخیره و بازیابی اطلاعات داشته باشد. با این مقدمه می‌توان نتیجه گرفت ارزیابی نرم‌افزارهای کتابخانه‌ای کشور

خوبی احساس می‌گردد. محقق با تبیین نیاز به بررسی جدی در این زمینه دو مسأله موجود در پیش روی تحقیق خود را در قالب دو سؤال زیر بیان کرده است:

- ۱- آیا با بکارگیری شیوه‌های متفاوت نمایه‌سازی فیلدها در نرم‌افزارهای فارسی، نتایج بازیابی متفاوتی بدست می‌آید؟
- ۲- تأثیر شیوه‌های نمایه‌سازی متفاوت بر قابلیت بازیابی اطلاعات در نرم‌افزارهای فارسی تا چه میزان است؟

بنابراین اهداف تحقیق مذکور عبارت است از:
الف) ارزیابی میزان کارایی شیوه‌های مختلف نمایه‌سازی در سه نرم‌افزار فارسی (پارس آذرخش، کاوش و نوسا) در زمینه قابلیت بازیابی اطلاعات.

ب) فراهم‌آوری بستر مناسب برای ارزیابی نرم‌افزارهای کتابخانه‌ای دیگر باتوجه به یافته‌های تحقیق. بحث در مورد اینکه قابلیت بازیابی تحت تأثیر عوامل مختلف و متعددی قرار می‌گیرد و این عوامل نقش معینی در شکل‌گیری نتایج جست‌وجو و در نهایت بازیابی اطلاعات دارند در قالب فرضیه تحقیق نمودار شده است و محقق فرضیه خود را بدین شکل بیان کرده است: بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی در سه نرم‌افزار موردتحقیق تفاوت معنی‌داری وجود دارد.

شیوه‌های نمایه‌سازی مدنظر تحقیق عبارتند از: کلمه، عبارت و فیلد کامل. محقق از بین فیلدهای مختلف پایگاه‌های داده‌ای سه فیلد پدیدآور، عنوان و موضوع را انتخاب کرده و در قالب فرضیه‌های فرعی به عنوان زیربخش‌های فرضیه اصلی به منظور بررسی آورده است که به شکل زیر است:

الف) بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی فیلد پدیدآور در سه نرم‌افزار مورد تحقیق تفاوت معنی‌داری وجود دارد.

ب) بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی فیلد عنوان در سه نرم‌افزار مورد تحقیق تفاوت معنی‌داری وجود دارد.

ج) بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی فیلد موضوع در سه نرم‌افزار مورد تحقیق تفاوت معنی‌داری وجود دارد.

باتوجه به اینکه محقق خواسته است تأثیر شیوه‌های نمایه‌سازی را در قابلیت بازیابی نرم‌افزارها بررسی کند. بنابراین شیوه‌های نمایه‌سازی فیلدهای سه‌گانه (عنوان، پدیدآور و موضوع) در نرم‌افزارهای (پارس، آذرخش، کاوش و نوسا) را به عنوان متغیر مستقل و قابلیت بازیابی در نرم‌افزارهای مذکور را به عنوان متغیر وابسته بررسی کرده

است.

جامعه آماری پژوهش از نوع جامعه محدود بوده و متشکل از سه پایگاه اطلاعاتی ایجاد شده در نرم‌افزارهای نوسا، کاوش و پارس آذرخش است. منابع مشخص نمونه‌ای برابر ۳۰۰ عنوان کتاب که همه آن‌ها در سه نرم‌افزار موردنظر، پارس آذرخش نسخه ۷/۰، کاوش نسخه ۴/۵ و نوسا نسخه ۸/۰ انتخاب شده است. در این پژوهش از روش علی - مقایسه‌ای استفاده شده است. محقق در توجیه این انتخاب آورده است: علی است به این دلیل که علت یا عللی نسبت بازیابی (درصد بازیابی) اطلاعات در سه نرم‌افزار از بُعد شیوه‌نمایه‌سازی و تأثیر آن بر بازیابی مورد بررسی قرار گرفته است. مقایسه‌ای است: زیرا سه نرم‌افزار مورد بررسی و میزان بازیابی اطلاعات باتوجه به شیوه‌های مرسوم نمایه‌سازی فیلدها با یکدیگر مقایسه می‌گردند.

محقق برای گردآوری اطلاعات مورد نیاز پژوهش از دو مرحله و شیوه استفاده نموده است: در مرحله اول با مطالعه راهنماهای چاپی مربوط به نرم‌افزارها اطلاعات مربوط به شیوه‌های نمایه‌سازی هر یک از نرم‌افزارها جمع‌آوری شده است و در مرحله دوم جهت گردآوری اطلاعات موردنیاز برای آزمودن فرضیه‌ها از فرم‌های محقق ساخته استفاده شده که نتایج بازیابی متأثر از شیوه‌های مختلف نمایه‌سازی فیلدها (عنوان، نویسنده و موضوع) روی آن‌ها ثبت و سپس مورد تجزیه و تحلیل قرار گرفته است. جهت تبیین تاریخچه نمایه‌سازی و بررسی پیشینه‌های موجود در این زمینه در ایران و خارج از کشور فصلی از تحقیق به این هدف اختصاص یافته است. بدین ترتیب محقق تاریخ نمایه را به طور اعم از شروع نوشتن در تاریخ حیات بشر و بطور اخص به ۳۰۰۰ سال قبل نسبت داده و شواهدی را برای اثبات این امر آورده و سیر تحول تاریخی نمایه‌سازی و شیوه‌ها و ابزارهای بکار رفته در طول تاریخ و اشخاص متقدم در این زمینه را بررسی نموده است. آنچه که مهم است اینکه نمایه‌سازی هم مثل سایر علوم و فنون متأثر از پیشرفت‌ها و حوادث تاریخی، سیاسی و اجتماعی زمان خود بوده است. بنا به استناد محقق بعد از جنگ جهانی دوم به دلیل به وجود آمدن بلوک‌های سیاسی حرکت به سوی مکانیزاسیون نمایه‌سازی و بکارگیری تکنولوژی در این امر و نهایتاً پدید آمدن فهرست‌های رایانه‌ای سریع‌تر و جدی‌تر شده است.

همراه با افزایش تولیدات علمی و اطلاعاتی تلاش بشر هم برای نظم بخشیدن به اطلاعات و بررسی قابلیت‌های این نظام‌ها شکل علمی و سازمان یافته‌تری به خود گرفته است و

در این دوران اندیشمندان و پژوهشگران زیادی به ارائه نظریه‌هایی پیرامون بهبود شیوه‌های ارائه محتوای مدارک، به ویژه نمایه‌سازی خودکار (ماشینی)، شیوه‌های جدید نمایه‌سازی فیلدها و تکامل و تنوع بخشیدن به این شیوه‌ها پرداختند. این نظریه‌پردازان یا افرادی دیگر، نظریه‌های ارائه شده را به بوته آزمایش گذارده و به نتایج خوبی نیز دست یافته‌اند که از این رهگذر، زمینه‌ای غنی برای پژوهش‌های بنیادی و کاربردی آتی و نیز بهینه کردن نظام‌های بازیابی کنونی فراهم آمده است.

محقق در قسمت پیشینه‌های تحقیق به تحقیقات مشابه انجام گرفته در داخل و خارج پرداخته و مواردی مثل «پژوهش‌های سیریل کلوردون» تحت عناوین کرینفیلد ۱ و ۲ مبتنی بر مقایسه زبان نمایه‌سازی پس همارا و پیش همارا، نظام بازیابی اسمارت (Smart) به رهبری سالتون، مطالعات لنکستر در مورد بررسی سودمندی نظام مدلارز (LARS-MED) و پژوهش‌های سیورتز (Sieverts) و همکاران در مورد نظام‌های بازیابی و نرم‌افزارهای موجود پرداخته و اطلاعات مختصری ارائه داده است. به پیشینه‌های موجود در ایران هم اشاره‌ای شده است.

در چنین تحقیقاتی که روی نظام‌های ذخیره و بازیابی اطلاعات انجام می‌گیرد به خاطر اینکه صحبت از یک نظام ماشینی به میان می‌آید این نظام‌ها دارای اجزاء و زیربخش‌ها و اصطلاحاتی هستند که ممکن است برای خوانندگان و یا محققان بعدی ناآشنا باشد بنابراین لازم بود که این مفاهیم و ساختارهای موجود در ذخیره و بازیابی اطلاعات بررسی شود که محقق از عهده آن به خوبی درآمده و فصلی را به این امر اختصاص داده و بطور مفصل ساختار و شیوه‌های ذخیره و بازیابی اطلاعات در نرم‌افزارهای کتابخانه‌ای را توضیح داده و در خلال آن مفاهیم بکار برده شده. در تحقیق خود را هم بررسی کرده و با اشاره به متون و اسناد موجود آنها را تبیین کرده است. این بررسی در قالب ۳ بخش انجام گرفته است: در بخش اول ساختار و شیوه‌های نمایه‌سازی فایل‌ها مورد بررسی قرار گرفته، در بخش دوم راهبردهای جست‌وجو و بازیابی، انواع قابلیت‌های جست‌وجو در نظام‌های بازیابی رایانه‌ای و عملگرهای مختلف تعریف و ارائه گردیده است و در نهایت در بخش سوم از این فصل ویژگی‌های نرم‌افزارهای کتابخانه‌ای فارسی پارس آذرخش، نرم‌افزار کاوش (شرکت مهندسی پویا کاوش) و نرم‌افزار نوسا (شرکت نرم‌افزار و سخت‌افزاری ایران) با استفاده از راهنماهای چاپی این شرکت‌ها از لحاظ ساختار و قابلیت‌های بازیابی و نمایه‌سازی توضیح داده شده است که

مقدمه جامعی برای پرداختن به اصل تحقیق فراهم آورده است. محقق تا به این قسمت از تحقیق خود سعی کرده است که بنیان‌های نظری و فنی تحقیق را فراهم آورده و توجیه نظری و ضرورت انجام آن را بیان کند که برای رسیدن به این هدف از یک ساختار درختی بهره گرفته است که شیوه خوبی برای بیان مطالب دارای زیربخش‌ها و قسمت‌های فرعی زیاد می‌باشد و بررسی قابلیت بازیابی نرم‌افزارهای کتابخانه‌ای هم از این امر مستثنی نیست. این تحقیق می‌تواند منبع قوی و منسجمی برای شکل‌دهی ایده‌ها و موضوعات تحقیقی در آینده باشد و گام مهمی در این راستا به حساب می‌آید. باتوجه به اینکه در سه فیلد مشخص (عنوان، موضوع و پدیدآور) انجام شده است می‌تواند راهنمای خوبی برای شکل‌گیری و انجام تحقیقات مفصل و فراگیر در سایر فیلدها، اجزاء، قابلیت‌های نرم‌افزاری و شیوه‌های مختلف جست‌وجو و... باشد. محقق برای تعیین جامعه آماری تحقیق خود مجموع ۱۵۱۰۴۱ رکورد از منابع پایگاه اطلاعاتی کتابشناسی ملی بر روی دیسک فشرده (نسخه ۱۳۷۹) را مینا قرار داده و برای محاسبه نمونه آماری از فرمول کوجسی و مورگان استفاده کرده و با کمی تغییر ۳۰۰ رکورد به صورت مشترک برای هر سه نرم‌افزار مورد نظر انتخاب و مورد آزمون قرار داده و برای تعیین نمونه‌ها از روش نمونه‌گیری منظم و از طریق شماره رکوردها بهره برده است.

جهت گردآوری اطلاعات و ارزیابی پایگاه‌ها از فرم‌های محقق ساخته استفاده شده است. این فرم‌ها در سه شکل به طور مجزا برای عنوان، پدیدآور و موضوع طراحی شده‌اند که هر فرم از قسمت‌های مختلفی مثل عناوین (برای ثبت ۳۰۰ عنوان منبع مشخص)، شماره رکورد در پایگاه منبع مورد جست‌وجو (برای ثبت ۳۰۰ شماره رکورد منابع مشخص)، (شیوه نمایه‌سازی فیلدها که در هر سه نرم‌افزار یکسان بوده است)، جست‌وجو از طریق عملگر (بین سه نرم‌افزار یکسان بوده است) و نتایج بازیابی اطلاعات که ردیف‌های «الف» برای نرم‌افزار پارس آذرخش، «ب» برای نرم‌افزار کاوش و «ج» برای نرم‌افزار نوسا در صورت پیدا شدن عدد (۱) و در صورت پیدا نشدن عدد (۰) درج می‌شده است. بدین ترتیب همه منابع مشخص در هر سه نرم‌افزار مورد جست‌وجو قرار گرفته و نتایج ثبت شده و به وسیله نرم‌افزار SPSS مورد تجزیه و تحلیل قرار گرفته و با کمک آزمون آماری X^2 مورد ارزیابی و سنجش قرار گرفته است که میزان معنی‌داری تفاوت‌ها بین نتایج جست‌وجو در هر فیلد در هر نرم‌افزار یا نرم‌افزارهای دیگر برای آزمون فرضیه‌های چهارگانه (یک فرضیه اصلی و سه فرضیه

فرعی) یکبار گرفته شده است.

براساس داده‌های گردآوری شده، شیوه محاسبه امتیازات بدین صورت بوده است که هر نرم‌افزاری که قابلیت بازیابی در هر یک از سه فیلد عنوان، پدیدآور و نویسنده را با توجه به انواع شیوه‌های نمایه‌سازی مربوطه (کلمه، عبارت و فیلد کامل) دارا باشد عدد (۱) می‌گیرد و اگر فاقد این قابلیت باشد عدد (۰) گرفته و در نهایت مجموع امتیازات (درصد بازیابی) نسبت به کل و به درصد بیان گردیده است. مجموع امتیازات کسب شده در هر بخش تفاوت درصد بازیابی هر نرم‌افزار را در مقایسه با سایر نرم‌افزارها نشان می‌دهد، نتایج تجزیه و تحلیل نیز به صورت جدول و نمودار دو بعدی در زیر مبحث هر فیلد به صورت مجزا و فیلدها در یک جا و در نهایت به صورت کلی برای بیان نتیجه نهایی ارائه شده است.

در قسمت آزمون فرضیه‌ها معلوم می‌شود:

الف. وجود تفاوت معنی‌دار بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی متفاوت در فیلد عنوان در سه نرم‌افزار مورد تحقیق مربوط به فرضیه اول تأیید گردیده است.

ب. وجود تفاوت معنی‌دار بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی متفاوت در فیلد پدید آمد در سه نرم‌افزار مورد تحقیق، مربوط به فرضیه دوم رد شده است.

ج. وجود تفاوت معنی‌دار بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی مختلف در فیلد موضوع در سه نرم‌افزار مورد تحقیق مربوط به فرضیه سوم تأیید گردیده است.

د. در مجموع، وجود تفاوت معنی‌دار بین قابلیت بازیابی متأثر از شیوه‌های نمایه‌سازی متفاوت در فیلدهای عنوان، پدیدآور و موضوع، در سه نرم‌افزار مورد تحقیق مربوط به فرضیه اصلی تأیید گردیده است. بنا به نظر محقق تفاوت موجود بین شیوه‌های نمایه‌سازی و قابلیت بازیابی منابع مشخص و موجود در سه نرم‌افزار مورد تحقیق عمدتاً ناشی از مسائل و مشکلاتی است که در حوزه‌های نمایه‌سازی، برنامه‌نویسی، ورود اطلاعات، جست‌وجو و بازیابی اطلاعات وجود دارد.

محقق بر تحقیق خود نتایجی را بیان کرده است که عیناً نقل می‌شود:

الف. نتیجه عمده تحقیق حاضر شناسایی اصول و شاخص‌های موجود در زمینه شیوه‌های نمایه‌سازی است که پیاده‌سازی آنها توسط طراحان نرم‌افزار در محیط‌های نرم‌افزاری موجود می‌تواند بر کیفیت و کمیت بازیابی و در نتیجه بر میزان مفید بودن جست‌وجوهای رایانه‌ای براساس جست‌وجوی فیلدهای عنوان، پدیدآورندگان و موضوعات که به صورت کلمه، عبارت، فیلد کامل نمایه‌سازی گردیده‌اند، بیفزاید. ب. دومین نتیجه تحقیق آن است که زمینه‌های ارزیابی

و ارتقای نرم‌افزارهای کتابخانه‌ای در زمینه راهبردهای جست‌وجو و بازیابی مورد شناسایی قرار گرفت.

ج. در تدوین شاخص‌های ارزیابی نرم‌افزارهای کتابخانه‌ای تلاش شد تا این شاخص‌ها بر اصول و کارکردهای مورد بحث استوار باشد.

در عین حال با توجه به اینکه تغییر و تحول در همه جنبه‌های نرم‌افزارهای کتابخانه‌ای امری طبیعی است که بسیار رخ می‌دهد نمی‌توان نتایج بدست آمده از فرضیه‌های تحقیق را برای همیشه مطلق فرض نمود، بلکه با توجه به شتاب سریع تحولات نرم‌افزارهای کتابخانه‌ای و همچنین وجود رقابت، احتمال برآورده شدن هریک از شاخص‌های مورد نظر این تحقیق از طرف هریک از این نرم‌افزارها وجود خواهد داشت. بر این اساس و برای سنجش این که آیا شاخص‌های تدوین شده در این تحقیق در مورد ویرایش‌های جدید نرم‌افزارها نیز اعتبار خواهد داشت یا خیر، ضروری است. این معیارها در مورد نسخه‌های جدید نرم‌افزارها نیز مورد بررسی قرار گیرد. در نهایت محقق با تکیه بر نتایج به دست آمده میانگین بازیابی توسط نرم‌افزارهای فارسی را محاسبه کرده و بازیابی با میانگین تقریباً ۸۲ درصد را نقطه ضعفی برای نرم‌افزارهای کتابخانه‌ای در حوزه بازیابی اطلاعات دانسته است برای حل این مشکل پیشنهاداتی را ارائه داده است تا طراحان و برنامه‌نویسان نظام‌های ذخیره و بازیابی اطلاعات کتابخانه‌ای فارسی با عمل به آنها از سندرم نظام‌های کوچک مبتنی بر از دست دادن قدرت بقا در درازمدت در امان مانده و بتوانند انتظارات کاربران خود را در دستیابی به اطلاعات جامع و مانع و به دور از ریزش کاذب و افت اطلاعات برآورده سازند. علاوه بر پیشنهادات در راستای فرضیه‌های تحقیق محقق توصیه کرده است با توجه به اینکه جست‌وجو و بازیابی از اطلاعات در پایگاه‌های اطلاعاتی یکی از فاکتورهای اساسی در نرم‌افزارهای کتابخانه‌ای به حساب می‌آید و از طرفی عملکرد نرم‌افزارهای کتابخانه‌ای موجود فارسی با همه محاسن و توانمندی‌هایی که دارا بوده دارای نواقص و ضعف‌هایی هستند. بنابراین توصیه می‌شود همان‌گونه که در نیم قرن اخیر در کشورهای غربی به‌ویژه انگلستان و آمریکا - پژوهش‌های گسترده‌ای در حوزه بازیابی اطلاعات و ارزشیابی نرم‌افزارهای کتابخانه‌ای در این حوزه انجام گرفته و دستاوردهای چشمگیری را نیز به همراه داشته است. پژوهشگران کتابداری و اطلاع‌رسانی و برنامه‌نویسان رایانه‌ای در ایران نیز به انجام مطالعه و تحقیق عمیق‌تر در این حوزه پرداخته، نواقص و ضعف‌ها را جهت برطرف کردن آنها مشخص نمود و مهم‌تر اینکه به خلق و ارائه نظریات نو و بدیع مبادرت ورزند.