

به نام خدا



وزارت علوم، تحقیقات، و فناوری

پژوهشگاه علوم و فناوری اطلاعات ایران

(ایرانداک)

خلاصه گزارش طرح پژوهشی سازمانی

عنوان طرح پژوهشی

ارزیابی کیفیت فراداده پایان نامه ها و رساله های الکترونیکی در پایگاه گنج

تاریخ شروع طرح پژوهشی	۱۴۰۳/۱۲/۰۱	تاریخ پایان طرح پژوهشی	۱۴۰۴/۱۱/۷
-----------------------	------------	------------------------	-----------

نام و نام خانوادگی مجری طرح پژوهشی	فریبرز درودی
------------------------------------	--------------

نام و نام خانوادگی همکاران طرح پژوهشی	بهروز رسولی، نصرالله پاک نیت، صبا روشن ضمیر
---------------------------------------	---

درباره طرح پژوهشی

(بیشینه ۳۰۰ واژه درباره مقدمه، مسئله، و هدف های پژوهش و آنچه که انجام، بررسی، یا تعیین شده است.)

پایگاه اطلاعاتی گنج، از زمره پایگاه های معتبر علمی در سطح کشور است که با هدف مدیریت منابع اطلاعاتی علمی و به ویژه پایان نامه ها و رساله های تحصیلات تکمیلی (پارسا) طراحی و راه اندازی شده است. بازیابی اطلاعات در این پایگاه بر اساس کاوش در فراداده های سازماندهی شده انجام می شود. از این رو، توجه به ساختار مهم کیفیت فراداده دارای اهمیت بالایی است. مزایای ارزیابی فراداده شامل: بهبود دسترسی به اطلاعات، افزایش دقت و صحت اطلاعات، تسهیل بهره گیری از داده ها، افزایش بهره وری، توسعه وضعیت حفاظت و نگهداری اطلاعات، همکاری بهتر و اشتراک گذاری مناسب اطلاعات، افزایش اعتبار پژوهش ها، و پاسخگویی دقیق به نیازهای کاربران. هدف اصلی پژوهش نیز عبارت است از شناسایی، ارزیابی و تحلیل مشکلات و چالش های مرتبط با کیفیت فراداده ها و ارائه راهکارهای مناسب برای بهبود آن ها تا میزان دسترسی به اطلاعات علمی افزایش یابد. روش پژوهش رویکرد نظام مند برای تحلیل و بررسی کیفیت فراداده ها در پایگاه گنج است. برای این کار نمونه ای هدفمند و متناسب با حجم اطلاعات انتخاب شده و مولفه های کیفیت فراداده بر روی آنها مورد مطالعه قرار می گیرد. همچنین با بهره گیری از روش های آماری، تحلیل نهایی انجام گرفته و به ارائه الگوی مناسب کیفیت فراداده منجر شد. از منظر نوآوری و ارزش افزوده پژوهش باید به این نکته مهم اشاره کرد که تا کنون پژوهشی در این رابطه و با ویژگی ها و عمق مورد نظر و نیز بر اساس مولفه های تعیین شده به انجام نرسیده است. نتایج علمی بیان شده در

پژوهش پیش‌رو، برای برنامه‌ریزی علمی و هماهنگی در ساختار پایگاه اطلاعات علمی ایران (گنج) حائز اهمیت است و می‌تواند در گسترش فعالیت‌های پژوهشی و اشاعه علمی پایان‌نامه‌ها و رساله‌های تحصیلات تکمیلی برای مدیران اطلاعاتی و فناوری ایرانداک سودمند باشد و در نهایت در ارائه اطلاعات و دسترس‌پذیری بهتر برای دانشجویان و پژوهشگران موثر واقع شود.

روش پژوهش

(بیشینه ۳۰۰ واژه درباره گام‌های پژوهش و چگونگی گردآوری و تحلیل داده‌ها (ابزار، جامعه، نمونه، روش نمونه‌گیری، روش‌های تجزیه و تحلیل داده‌ها).

روش‌شناسی این پژوهش شامل یک رویکرد نظام‌مند برای ارزیابی کیفیت فراداده‌ها در پایگاه گنج است. بخش نخست پژوهش با رویکردی کمی و با استفاده از روش تحلیل محتوای ساختاریافته و بررسی مقابله‌ای انجام شد. با توجه به حجم انبوه رکوردها در پایگاه اطلاعاتی گنج، امکان بررسی تمام پیشینه‌ها میسر نبود؛ لذا از روش نمونه‌گیری خوشه‌ای تصادفی^۱ استفاده شد. انتخاب این روش به منظور اطمینان از بازنمایی متناسب تمامی حوزه‌های دانش و مقاطع تحصیلی در نمونه نهایی صورت گرفت.

از آنجا که در این پژوهش یکی دیگر از اهداف واکاوی چالش‌های موجود در سازماندهی اطلاعات پایگاه گنج بود، از روش تحقیق کیفی نیز بهره‌برداری شد. از آنجا که کیفیت فراداده امری است که مستقیماً با تجربه زیسته و مهارت‌های تخصصی نمایه‌سازان در هم تنیده، استفاده از رویکرد پدیدارشناختی برای درک موانع فنی و محتوایی ضرورت داشت. در این راستا، استراتژی مطالعه موردی انتخاب شد تا تمرکز بر زیرساخت‌های خاص ایرانداک حفظ شود. برای جمع‌آوری داده‌های، از تکنیک گروه کانون استفاده گردید که به پژوهشگران اجازه می‌داد تا علاوه بر نظرات فردی، از تعاملات میان متخصصان و تضارب آراء آن‌ها برای شناسایی ریشه‌ای مشکلات بهره‌برداری کنند.

گام‌های اجرایی پژوهش به شرح زیر بود:

۱. انجام مطالعه‌های نظری درباره کیفیت فراداده؛
۲. گزینش نمونه‌ای از رکوردها (تعداد نمونه: ۶۶۵؛ با سطح اطمینان ۹۹ درصد و فاصله اطمینان ۵) برای بررسی فراداده آن‌ها؛
۳. بررسی میزان کامل بودن فیلدهایی که برای توصیف یک رکورد استفاده شده‌اند؛
۴. بررسی میزان دقت تکمیل فراداده؛

¹ Random Cluster Sampling

۵. شناسایی مشکلات و چالش‌های موجود در کیفیت فراداده رکوردها؛

۶. ارائه پیشنهاد‌های اجرایی و کاربردی.

یافته‌های طرح پژوهشی

(بیشینه ۵۰۰ واژه درباره یافته‌ها و دستاوردهای کلیدی پژوهش در پاسخ به پرسش‌ها یا بررسی فرضیه‌های پژوهش.)

پژوهش حاضر با هدف ارزیابی کیفیت و جامعیت فراداده‌های پایگاه گنج و شناسایی چالش‌ها و مشکلات موجود در مرحله نمایه‌سازی و تعریف فراداده انجام شده است. این پژوهش از رویکردی ترکیبی بهره می‌برد که شامل دو مرحله اصلی است. در مرحله نخست، کیفیت و جامعیت فراداده‌های یک نمونه آماری از رکوردهای گنج به صورت نظام‌مند بررسی و با فایل‌های تمام‌متن پی‌دی‌اف مربوطه که از واحد فناوری اطلاعات ایرانداک دریافت شده بود، مقایسه شد. در مرحله دوم، با برگزاری یک گروه کانونی متشکل از هفت نفر از کارکنان واحد نمایه‌سازی ایرانداک و دو نفر از اعضای تیم پژوهش، تجربیات، چالش‌ها و مشکلات عملی آنان در فرایند نمایه‌سازی پایان‌نامه‌ها و رساله‌ها و تعریف فراداده‌ها مورد بحث و تحلیل قرار گرفت.

در مرحله نخست پژوهش، جامعه آماری شامل کلیه پایان‌نامه‌ها و رساله‌های ثبت‌شده در پایگاه گنج در بازه زمانی سال‌های ۱۳۹۱ تا ۱۴۰۰ بود. به منظور دستیابی به نمونه‌ای نماینده از تنوع موضوعی، مقاطع تحصیلی و سال‌های مختلف، از روش نمونه‌گیری خوشه‌ای تصادفی استفاده شد. در این فرایند، ابتدا خوشه‌ها بر اساس گروه‌های تحصیلی مختلف شامل علوم انسانی، علوم پایه، فنی و مهندسی، علوم پزشکی، دامپزشکی، کشاورزی و هنر تعریف شدند و سپس نمونه‌ها به صورت تصادفی از هر خوشه انتخاب شدند. در مجموع، ۶۶۵ رکورد به عنوان نمونه اولیه انتخاب شد که پس از بررسی امکان دسترسی به فایل‌های تمام‌متن و تطبیق آن‌ها با رکوردهای موجود، فراداده‌های ۵۹۴ رکورد مورد ارزیابی نهایی قرار گرفت.

ارزیابی فراداده‌ها بر اساس مجموعه‌ای از فیلدهای اصلی توصیفی و محتوایی انجام شد که شامل اطلاعات کتابشناختی پایه مانند عنوان، سال انتشار، مقطع تحصیلی، نوع مدرک و نام پدیدآور، و نیز عناصر محتوایی و تکمیلی مانند موضوع، رشته و گرایش، چکیده، کلیدواژه‌ها، نام استادان راهنما و مشاور، فهرست منابع فارسی و غیر فارسی و وجود فایل‌های پی‌دی‌اف بود. برای هر رکورد، وجود یا عدم وجود داده در هر فیلد بررسی و موارد تهی ثبت شد. سپس، داده‌های حاصل بر اساس گروه‌های تحصیلی تحلیل شد تا الگوهای تفاوت یا شباهت در میزان

جامعیت فراداده میان حوزه‌های مختلف علمی شناسایی شود.

نتایج مرحله کمی پژوهش نشان داد که در برخی از فیلدهای پایه، مانند عنوان، سال انتشار، مقطع تحصیلی و نوع مدرک، میزان جامعیت بسیار بالا و در اغلب موارد کامل است. این امر نشان‌دهنده اهمیت و اولویت این عناصر در فرایند نمایه‌سازی و کنترل کیفی اولیه رکوردها در پایگاه گنج است. با این حال، در فیلدهای محتوایی و تکمیلی، به‌ویژه اطلاعات مربوط به استادان مشاور، فهرست منابع و کلیدواژه‌ها، میزان قابل توجهی از داده‌های تهی مشاهده شد. این الگو بیانگر آن است که هرچه فیلدها تخصصی‌تر، تفسیری‌تر یا وابسته‌تر به ساختار و کیفیت متن اصلی باشند، احتمال نقص یا عدم تکمیل آن‌ها افزایش می‌یابد.

در کنار تحلیل کمی، یافته‌های گروه کانونی بُعد عمیق‌تری از مسئله کیفیت فراداده را آشکار ساخت. کارکنان واحد نمایه‌سازی ایراداک به چالش‌های متعددی اشاره کردند که ریشه در عوامل ساختاری، فنی و انسانی دارد. از جمله این چالش‌ها می‌توان به ناهمگونی قالب و ساختار پایان‌نامه‌ها و رساله‌های دریافتی از دانشگاه‌های مختلف، نبود استانداردهای یکنواخت در نگارش بخش‌های کلیدی مانند چکیده و فهرست منابع، محدودیت زمان برای نمایه‌سازی حجم بالای منابع، و ابهام یا نقص اطلاعات ارائه‌شده در فایل‌های اولیه اشاره کرد. این عوامل سبب می‌شود که نمایه‌سازان در بسیاری از موارد ناچار به تفسیر شخصی، تصمیم‌گیری موردی یا حتی رها کردن برخی فیلدها به‌صورت تهی شوند.

نتیجه‌گیری و پیشنهادها

(بیشینه ۳۰۰ واژه بحث و نتیجه‌گیری و بیان پیشنهادهای برگرفته از پژوهش).

یافته‌های مربوط به جامعیت فراداده‌ها در پایگاه گنج نشان می‌دهد که اگرچه عناصر پایه و هویتی فراداده‌ای، مانند عنوان، سال انتشار، مقطع تحصیلی و نوع مدرک، در اغلب رکوردها به‌طور کامل ثبت شده‌اند، اما در بخش قابل توجهی از فیلدهای محتوایی و تکمیلی با کاستی‌های جدی مواجه هستیم. میزان بالای فراداده‌های تهی در فیلدهایی همچون فهرست منابع فارسی و غیر فارسی، فهرست نوشته‌ها، کلیدواژه‌ها و در برخی موارد چکیده، بیانگر آن است که ساختار کنونی فراداده در گنج بیشتر بر حداقل‌های توصیف کتابشناختی تمرکز دارد و کمتر پاسخگوی نیازهای تحلیلی، موضوعی و پژوهشی کاربران است.

یکی از نکات کلیدی که از نتایج این پژوهش استنباط می‌شود، محدودیت ذاتی شمای فراداده‌ای فعلی پایگاه

گنج است. این شما از نظر تعداد و تنوع فیلدها، ظرفیت کافی برای بازنمایی دقیق و غنی محتوای پارساها را ندارد و در بسیاری موارد، نمایه‌سازان حتی در صورت تمایل و توان، امکان ثبت اطلاعات تکمیلی مهم را در قالب ساختار موجود ندارند. نبود فیلدهای تفصیلی برای ثبت انواع مختلف نقش‌های پدیدآورندگان، اطلاعات روش‌شناختی پژوهش، داده‌های مربوط به نسخه‌های مختلف اثر، یا حتی تفکیک دقیق اجزای منابع و پیوست‌ها، سبب شده است که بخشی از اطلاعات ارزشمند یا اصلاً ثبت نشود یا به صورت پراکنده و غیرقابل بازیابی باقی بماند.

افزون بر این، نتایج نشان می‌دهد که جامعیت پایین برخی فیلدها الزاماً به کیفیت پایین فراداده در گنج نیست، بلکه در بسیاری موارد اطلاعات مربوطه در فایل تمام‌متن وجود ندارد. برای نمونه، فهرست منابع یا اطلاعات استادان مشاور ممکن است در متن پایان‌نامه‌ها درج نشده باشند. در مجموع، یافته‌های مربوط به جامعیت فراداده‌ها نشان می‌دهد که بهبود کیفیت فراداده در پایگاه گنج صرفاً با افزایش دقت یا آموزش نمایه‌سازان محقق نخواهد شد، بلکه مستلزم بازطراحی اساسی شمای فراداده‌ای است.

در مجموع، این یافته‌ها بر ضرورت بازنگری در رویکردهای ارزیابی کیفیت فراداده تأکید می‌کند. از یک سو، نیاز به تدوین و به‌کارگیری استانداردهای نگارشی دقیق‌تر، ابزارهای یکسان‌سازی خودکار و کنترل واژگانی برای کاهش ناسازگاری‌های صوری کاملاً مشهود است. از سوی دیگر، باید در شاخص‌های ارزیابی دقت، امکان تشخیص موارد اصلاحی و آگاهانه نمایه‌سازان نیز پیش‌بینی شود. چنین رویکردی می‌تواند به درک واقع‌بینانه‌تری از کیفیت فراداده در پایگاه گنج منجر شود و همزمان نقش تخصصی و ارزش‌افزوده نیروی انسانی در فرایند سازماندهی اطلاعات علمی را به‌درستی منعکس کند.