



وزارت علوم، تحقیقات و فناوری
مرکز اطلاعات و مدارک علمی ایران

بررسی میزان همخوانی کلید واژه های عنوان و
توصیفگرهای نمایه سازان در پایگاه
چکیده پایان نامه های ایران

مجری:
محمّد ایرانشاهی

داور:
سرکار خانم دکتر شهین نعمت زاده

اردیبهشت ۱۳۸۳

چکیده:

پژوهش حاضر با هدف ارزیابی میزان همخوانی کلید واژه های عنوان پایان نامه های مندرج در "پایگاه چکیده پایان نامه های ایران" با توصیفگرهای اختصاص یافته به این پایان نامه ها صورت پذیرفت. جامعه مورد بررسی، کلیه رکوردهای موجود در این پایگاه از ابتدا تا پایان سال ۱۳۸۱ بود که ۶۵۵۳۶ رکورد را شامل می شد. پس از حذف رکوردهای تکراری و رکوردهایی که فاقد مشخصه های لازم برای بررسی بودند جامعه تحقیق بر اساس دسته بندی صورت پذیرفته در "پایگاه چکیده پایان نامه های ایران" و "فصلنامه چکیده پایان نامه های ایران" به ۶ حوزه علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، علوم کشاورزی و هنر تقسیم شد. سپس با استفاده از فرمول $pq / (d^2(N-1))$ حجم نمونه در هر حوزه محاسبه شد. چون تعداد نمونه محاسبه شده در هر یک از حوزه ها نزدیک به ۱۰۰ بود، به منظور افزایش دقت در مقایسه حوزه ها به طور مساوی از هر حوزه ۱۰۰ رکورد به عنوان نمونه تحقیق به طور تصادفی گزینش شد. سپس به منظور پاسخگویی به سوالات تحقیق در ارتباط با تفاوت متوسط تعداد کلید واژه های عنوان در حوزه های مختلف، میزان همخوانی کامل، همخوانی نسبی و عدم همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان در این حوزه ها، سنجش رابطه تعداد توصیفگرها با تعداد کلید واژه های عنوان و نیز رابطه تعداد کلید واژه ها با افزایش موارد همخوانی کامل، شیوه نگارش توصیفگرها و سنجش میزان بازیافت از طریق کلید واژه های عنوان و توصیفگرهای نمایه سازان؛ در قالب فرم محقق ساخته ای اطلاعات مورد نظر گردآوری شده و با استفاده از جداول، نمودارها و اشکال توصیفی با استفاده از فرمول های آماری ضریب همبستگی پیرسون و آزمون مجذور کای به سوالات مورد بحث پاسخ داده شد. نتایج به دست آمده بیانگر این بود که بین تعداد متوسط کلید واژه های عنوان در حوزه های مورد بررسی تفاوت معنی

داری وجود دارد و ۴۵ درصد توصیفگرهای نمایه سازان را کلید واژه های عنوان تشکیل می دهد که اگر موارد همخوانی نسبی را نیز در نظر بگیریم این رقم به ۶۷ درصد افزایش می یابد. از طرف دیگر یافته های تحقیق نشان داد که همبستگی بالایی بین تعداد کلید واژه های عنوان، تعداد توصیفگرها و تعداد همخوانی کامل وجود دارد. در عین حال با گذشت زمان و شاید به دلایلی چون تدوین اصطلاحنامه ها و آموزش نمایه سازان میزان وابستگی به کلید واژه های عنوان کاهش یافته است. نسبت تقریباً مساوی بین توصیفگرهای تک واژه (۴۸ درصد) و ترکیب واژه (۵۲ درصد) و فزونی توصیفگرهای مفرد (۸۷ درصد) در مقایسه با توصیفگرهای جمع (۱۳ درصد) یکی دیگر از یافته های این تحقیق بود. در پایان نیز بازیافت بیشتر از طریق توصیفگرها در مقایسه با کلید واژه های عنوان، اثر بخشی طرح هایی را نشان داد که به منظور افزایش ارتباط بین توصیفگرهای هم موضوع در این پایگاه صورت پذیرفته است. این تحقیق با پیشنهاداتی به منظور رفع نقایص مربوط به " پایگاه چکیده پایان نامه های ایران"، اعم گردآوری، ثبت و ضبط، استفاده بهینه از توصیفگرها و کلید واژه ها و انجام تحقیقات مشابهی بر روی دیگر پایگاه های مرکز به پایان رسیده است.

پیشگفتار

افزایش سریع حجم مدارک و توسعه رشته های علمی موجب پیدایش روشهای گوناگونی برای ضبط محتوای اطلاعاتی آنها شده است. برای محققانی که در زمینه های گوناگون علمی دست در کار پژوهش هستند مطالعه کلیه مدارکی که در زمینه تخصص آنها منتشر شده امکان پذیر نیست. حتی در بسیاری از موارد فرصت مطالعه چکیده مقاله ها و گزارشهای تحقیقی نیز میسر نمی باشد. علاوه بر این چه بسیار تحقیقاتی که در نقاط مختلف جهان بعمل آمده و نشانه ای از آنها در دست نیست. با توجه به موارد فوق و به منظور جلوگیری از دوباره کاری و ایجاد سرعت و دقت در فعالیتهای پژوهشی، صاحب نظران دانش اطلاع رسانی بر آن شدند تا با استفاده از وسائل متعدد از هر مدرک علمی نشانه ای بدست دهند. یکی از این راه ها "نمایه سازی" است که جزء وظیفه اساسی بسیاری از مراکز اطلاعاتی و انتشاراتی جهان به شمار می رود. در تحقیق حاضر دو شیوه رایج در نمایه سازی منابع غیر کتابی، یعنی نمایه سازی به زبان طبیعی و نمایه سازی به زبان آزاد در مرکز اطلاعات و مدارک علمی ایران مورد تحقیق و بررسی قرار گرفته است تا بر اساس یافته های آن بتوان با دقت و سرعت بیشتر و هزینه های کمتر نسبت به تهیه نمایه ها اقدام نمود.

فهرست مندرجات

فصل اول : معرفی پژوهش.....	۶
۱-۱- مقدمه :.....	۷
۲-۱- بیان مسأله :.....	۱۰
۳-۱- هدف های پژوهش :.....	۱۳
۴-۱- سوالات پژوهش :.....	۱۴
۵-۱- تعریف عملیاتی اجزا متغیرها :.....	۱۴
فصل دوم :بررسی مطالعات گذشته.....	۱۹
۱-۲- مروری بر مبانی نظری :.....	۲۰
۲-۱-۱- نمایه سازی به زبان طبیعی :.....	۲۱
۱.نمایه کوئیک یا نمایه درون بافتی.....	۲۳
۲.نمایه کووک یا نمایه برون بافتی.....	۲۵
۳.نمایه جایگشتی.....	۲۷
توانایی ها و ضعف های نمایه گردان.....	۲۸
اصلاح نمایه گردان.....	۳۱
۲-۱-۲- نمایه سازی به زبان آزاد :.....	۳۳
۳-۱-۲- نمایه سازی به زبان کنترل شده :.....	۳۶
ویژگی زبان های نمایه سازی:.....	۳۷
۲-۲- پیشینه تحقیق:.....	۳۹
۱-۲-۲- مطالعات انجام شده در خارج از ایران :.....	۴۰
۲-۲-۲- مطالعات انجام شده در داخل کشور:.....	۴۵
فصل سوم: روش پژوهش و گردآوری اطلاعات.....	۴۸
۱-۳- روش پژوهش:.....	۴۹
۲-۳- جامعه آماری :.....	۴۹
۳-۳- روش نمونه گیری :.....	۴۹
۴-۳- ابزار گردآوری داده ها:.....	۵۳
۵-۳- روش تجزیه و تحلیل اطلاعات:.....	۵۳
فصل چهارم: یافته ها و تحلیل اطلاعات.....	۵۵
۱-۴- مقدمه:.....	۵۶
۲-۴- چه تفاوتی بین تعداد متوسط کلید واژه عنوان پایان نامه در حوزه های علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی ، کشاورزی و هنر وجود دارد ؟.....	۵۶

- ۳-۴- بین تعداد کلید واژه های عنوان پایان نامه ها با تعداد توصیفگرهای اختصاص یافته چه رابطه ای وجود دارد؟..... ۶۴
- همان گونه که داده های جدول فوق نیز نشان می دهد در کلیه حوزه ها همبستگی مثبتی بین تعداد کلید واژه های عنوان و تعداد توصیفگرها وجود دارد..... ۶۸
- ۴-۴- چند درصد توصیفگرها، از کلید واژه های عنوان گزینش شده اند؟ به عبارت دیگر همخوانی کامل کلید واژه های عنوان با توصیفگرها چقدر است؟..... ۶۹
- ۵-۴- در حوزه های علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر چه تفاوتی بین میزان همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان وجود دارد؟..... ۶۹
- ۶-۴- میزان همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان در سال های مختلف چه تفاوتی دارد؟..... ۸۱
- ۷-۴- آیا با افزایش تعداد کلید واژه های عنوان، میزان همخوانی آنها با توصیفگرها نیز افزایش می یابد؟..... ۸۹
- ۸-۴- از نظر شیوه نگارش توصیفگرها چگونه هستند؟..... ۹۱
- ۹-۴- در "پایگاه چکیده پایان نامه های ایران"، بین میزان بازیافت با استفاده از کلید واژه های عنوان و توصیفگرهای نمایه سازان چه تفاوتی وجود دارد؟..... ۹۴
- ۱۰-۴- توصیفگرهای نمایه سازان تا چه حد با توصیفگرهای اصطلاحنامه های فارسی موجود همخوانی دارد؟..... ۱۰۲
- ۴-۱۱- توصیفگرهای نمایه سازان تا چه حد با رسم الخط فرهنگستان زبان و ادب فارسی همخوانی دارد؟..... ۱۰۴
- فصل پنجم: نتیجه گیری و پیشنهادها..... ۱۰۵**
- منابع..... ۱۱۲**
- فارسی :..... ۱۱۳
- انگلیسی :..... ۱۱۶

فهرست جداول

جدول شماره ۱- تعداد رکورد های حذف شده قبل از تفکیک حوزه.....	۵۰
جدول شماره ۲- عناوین حذف شده پس از تفکیک رشته های اصلی.....	۵۲
جدول شماره ۳- تعداد متوسط کلید واژه در عناوین پایان نامه های هر یک از حوزه های شش گانه.....	۵۶
جدول شماره ۴- آزمون ANOVA برای محاسبه اختلاف تعداد متوسط کلید واژه های عنوان.....	۵۹
جدول شماره ۵- توزیع فراوانی پایان نامه ها بر حسب تعداد کلید واژه های عنوان ، به تفکیک حوزه.....	۶۰
جدول شماره ۶- تعداد متوسط واژه ها ، در عناوین پایان نامه های هر یک از حوزه های شش گانه.....	۶۱
جدول شماره ۷- تعداد متوسط واژه های غیر موضوعی در عناوین پایان نامه های هر یک از حوزه های شش گانه.....	۶۲
جدول شماره ۸- مقایسه تعداد کلید واژه و واژه های غیر موضوعی در عناوین پایان نامه های حوزه های شش گانه.....	۶۳
جدول شماره ۹- محاسبه مجذور کای در خصوص تفاوت تعداد کلید واژه ها و واژه های غیر موضوعی عنوان.....	۶۳
جدول شماره ۱۰- محاسبه مجذور کای در خصوص تفاوت تعداد واژه های غیر موضوعی.....	۶۴
جدول شماره ۱۱- توزیع فراوانی تعداد کلید واژه های عنوان و تعداد توصیفگرها به تفکیک حوزه.....	۶۵
جدول شماره ۱۲- ضریب همبستگی تعداد کلید واژه های عنوان و تعداد توصیفگرها به تفکیک حوزه.....	۶۸
جدول شماره ۱۳- وضعیت همخوانی کلید واژه های عنوان با توصیفگرها ، به تفکیک حوزه.....	۷۰
جدول شماره ۱۴- مقایسه همخوانی و عدم همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان به تفکیک حوزه.....	۷۱
جدول شماره ۱۵- مقایسه بررسی همخوانی کامل کلید واژه های عنوان با توصیفگرها ، به شیوه دستی و ماشینی.....	۷۹
جدول شماره ۱۶- آزمون مجذور کای برای سنجش تفاوت محاسبه تعداد توصیفگرها به شیوه دستی و ماشینی.....	۸۰
جدول شماره ۱۷- آزمون مجذور کای برای سنجش تفاوت محاسبه تعداد همخوانی کامل، به شیوه دستی و ماشینی.....	۸۰
جدول شماره ۱۸- محاسبه ماشینی تعداد توصیفگرها و همخوانی کامل آنها با کلید واژه های عنوان در کل جامعه تحقیق.....	۸۱
جدول شماره ۱۹- توزیع فراوانی پایان نامه ها بر حسب سال نگارش ، به تفکیک حوزه.....	۸۲
جدول شماره ۲۰- توزیع فراوانی همخوانی کلید واژه های عنوان با توصیفگرها به تفکیک سال نگارش.....	۸۶
جدول شماره ۲۱- نتیجه آزمون مجذور کای در خصوص تفاوت موارد همخوانی در سال های مختلف.....	۸۸
جدول شماره ۲۲- توزیع فراوانی تعداد کلید واژه و همخوانی ها (همخوانی کامل و نسبی) به تفکیک حوزه.....	۸۹
جدول شماره ۲۳- ضریب همبستگی تعداد کلید واژه های عنوان و تعداد همخوانی کامل در حوزه های شش گانه.....	۹۱
جدول شماره ۲۴- نسبت تعداد توصیفگرهای تک واژه به تعداد واژه های ترکیبی، به تفکیک حوزه.....	۹۲
جدول شماره ۲۵- نسبت توصیفگرهای مفرد به توصیفگرهای جمع ، به تفکیک حوزه.....	۹۳
جدول شماره ۲۶- نسبت کلید واژه های مفرد به کلید واژه های جمع ، به تفکیک حوزه.....	۹۴
جدول شماره ۲۷- تعداد رکوردهای دارای موارد همخوانی نسبی.....	۹۷
جدول شماره ۲۸- تعداد بازیافت از طریق کلید واژه ها و توصیفگرهای دارای همخوانی نسبی، به تفکیک حوزه.....	۹۷
جدول شماره ۲۹- آزمون مجذور کای در خصوص میزان بازیافت از طریق کلید واژه های عنوان و توصیفگرهای نمایه سازان.....	۹۸
جدول شماره ۳۰- تعداد رکوردها دارای حداقل یک مورد همخوانی کامل.....	۱۰۰
جدول شماره ۳۱- تعداد بازیافت از طریق کلید واژه ها و توصیفگرهایی که دارای همخوانی کامل، به تفکیک حوزه.....	۱۰۰
جدول شماره ۳۲- تعداد رکوردهای دارای کلید واژه و توصیفگر مترادف به تفکیک حوزه.....	۱۰۱
جدول شماره ۳۳- تعداد بازیافت از طریق کلید واژه ها و توصیفگرهایی مترادف، به تفکیک حوزه.....	۱۰۲

- جدول شماره ۳۴ - وضعیت همخوانی توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه های فارسی ۱۰۳
- جدول شماره ۳۵ - میزان همخوانی توصیفگرها با رسم الخط فارسی ۱۰۴

فهرست نمودارها

- نمودار شماره ۱- سیر خطی تعداد کلید واژه های عنوان بر حسب تاریخ ۵۹
- نمودار شماره ۲- توزیع فراوانی تعداد کلید واژه ها و توصیفگرها ۶۷
- نمودار شماره ۳- توزیع فراوانی پایان نامه ها بر حسب سال نگارش ، به تفکیک حوزه ۸۳
- نمودار شماره ۴- فارغ التحصیلان دانشگاه ها و مؤسسات آموزش عالی از سال ۱۳۵۷ تا سال ۱۳۷۶ ۸۵
- نمودار شماره ۵- مقایسه موارد همخوانی و عدم همخوانی توصیفگرها و کلید واژه های عنوان به تفکیک سال ۸۷
- نمودار شماره ۶- توزیع فراوانی کلید واژه های عنوان و همخوانی ها. ۹۰

فهرست شکل ها

- شکل شماره ۱- استفاده از دستور Pivot table جهت مشخص شدن تعداد عناوین تکراری ۵۲
- شکل شماره ۲- استفاده از دستور Find به منظور شناسایی توصیفگرها ۷۳
- شکل شماره ۳- استفاده از دستور Count برای شمارش تعداد توصیفگرها. ۷۴
- شکل شماره ۴- دستور Mid برای ثبت توصیفگرها ۷۵
- شکل شماره ۵- دستور Find برای یافتن موارد همخوانی کامل ۷۶
- شکل شماره ۶- دستور Count برای شمارش تعداد همخوانی کامل ۷۷
- شکل شماره ۷- درصد همخوانی کامل در هر عنوان ۷۸

فصل اول

معرفی پژوهش

۱-۱- مقدمه :

"عنوان" نخستین بخش از هر اثر است که نظر خواننده را به خود جلب می کند. از دیدگاه خواننده یا پژوهشگر "عنوان" دارای ارزش زیادی است. "عنوان" هر اثر جلوه ای از هویت آن اثر است و به تعبیری نخستین جلوه ای از متن است که خواننده با آن روبه رو می شود (داورپناه، ۱۳۷۵، ۱).^۱ عنوان کتاب ها، نشریات، پایان نامه ها و مقالات معمولاً در خلاصه ترین شکل ممکن و با حروفی درشت تر از متن چاپ شده و ناظر بر محتوی اثر است. هر کتاب، مقاله، داستان، شعر و یا فیلم جدا از نام پدید آور، با نام عنوانش نیز شناخته می شود. گاهی اوقات پیش می آید که نام نویسنده ای را فراموش می کنیم ولی عنوان اثرش را به خاطر داریم؛ یا این که نویسندگانی را با آثارشان می شناسیم. مثلاً مولوی را با "مثنوی"، سعدی را با "بوستان" و "گلستان"؛ و جک لندن را با "سپید دندان" (حدادیان مقدم، ۱۳۷۷، ۵۶). در صفحات "وب" نیز عناوین برای جلب نظر مخاطبان و نمایه شدن در موتورهای جستجو از اهمیت زیادی برخوردار هستند. از این رو، "عنوان" می بایست جامع و مانع باشد. "عنوان" ظرفی است که مضمون آن اندیشه اصلی متن است. در نظر بگیریم عنوانی را با این عبارت "شناخت گیاهان ایران" (که محتوای آن تنها به بررسی گیاهان مناطق غرب ایران پرداخته باشد)، یا عنوانی را با عبارت "تاریخچه مدارس نظامی ایران" (که محتوای آن به بررسی انواع مختلف مدارس ایران- از جمله مدارس نظامی- پرداخته باشد) انتخاب هر دو نوع عنوان "مغشوش" کردن کالایی است که عرضه می شود، منتهی نوع غش هر یک با دیگری متفاوت است (حرّی، ۱۳۷۲، ۱۷۸).

نارسایی عنوان همیشه به این دلیل نیست که ظرف و مضمون با هم ناسازگارند. گاه عنوان نقش اصلی خود را ایفا نمی کند و رساننده مباحث متن نیست، حتی ممکن است به صورت استعاری بیان شده باشد. این امر در ارتباط با آثار هنری یا برخی مباحث علوم انسانی بیشتر مشاهده می شود. عناوینی چون "هنر عشق ورزیدن"، "رنج و سرمستی"، یا "خسی در میقات" اطلاعات خاصی درباره متن به دست نمی دهد؛ یا وب سایتی

^۱ کتابنامه این تحقیق بر اساس شیوه نامه دانشگاه شیکاگو (The Chicago Manual of Style) تنظیم شده است.

با عنوان *What the Thunder said* گویای این نیست که حاوی مطالبی درباره زندگی و آثار T.S. Eliot است (تیلور، ۱۳۸۱، ۱۶۷). نیومارک (Newmark, 1988, 57) عناوین را به دو گروه عناوین توضیحی (Descriptive) و تلمیحی (Allusive) تقسیم بندی کرده است. به نظر وی در متون غیر ادبی، عنوان توضیحی به اختصار به موضوع اثر اشاره می کند و هدف متن را به نحوی در خور بیان می دارد و عنوان تلمیحی نیز نوعی ارتباط ارجاعی یا استعاره ای با موضوع دارد. درباره آثار علمی، پژوهشی و اطلاعاتی استفاده از عنوان تلمیحی مطلوب نیست. بسیاری از ناشران متون علمی می کوشند میان این دو نقش عنوان - یعنی زیبایی و اطلاع رسانی - تلفیق ایجاد کنند، و این کار را با استفاده از عنوان فرعی انجام می دهند. در این گونه تلفیق ها، عنوان اصلی عمدتاً نقش زیبایی شناسانه و عنوان فرعی نقش اطلاعاتی دارد. مجلات معتبر دنیا نیز یکی از شرایط قبول مقاله را برابری عنوان با محتوی متن تعیین می کنند، و این کار را به دو دلیل انجام می دهند :

۱- به دلیل آن که صورت مسأله تحقیق باید به گویا ترین وجه در عنوان متجلی باشد؛

۲- به این دلیل که عنوان، پایگاه موضوعی بسیاری از سازمان هایی است که به کار نمایه سازی ماشینی مقاله ها و مدارک دیگر می پردازند.

فرضی که ورای چنین گزینشی نهفته است این است که مجموعه کلید واژه های عنوان، جلوه ای از موضوع کلی مدرک است. روش استفاده از کلید واژه های عنوان برای تعیین و منعکس کردن موضوع مدارک هنوز هم در انواع نمایه سازی ها معمول است. حتی اگر عنوان رساننده موضوع متن نباشد، برای این منظور از عنوان ساختگی - که به چکیده عنوانی معروف است - استفاده می کنند. یعنی عنوانی را در کنار عنوان ساخته مولف به کار می برند که تجلی مباحث متن باشد (حرّی، ۱۳۷۲، ۱۷۹). همزمان با افزایش فوق العاده اطلاعات در طول جنگ جهانی دوم و ضرورت دستیابی سریع به منابع اطلاعاتی، روش های مختلف نمایه سازی ابداع شد که دو روش عمده آن بر اساس استفاده از واژه های مدرک (نمایه سازی طبیعی) یا استفاده از واژگان موضوعی (نمایه سازی کنترل شده) استوار بود .

از آنجا که در نمایه سازی موضوعی، تجزیه و تحلیل محتوا نیاز به کوشش فکری نمایه ساز و استفاده از یک زبان نمایه را داشت، طبعا^۱ سرمایه گذاری مادی و معنوی قابل توجهی را ایجاب می کرد. علاوه بر این، وقت گیر بودن مراحل کار از یکسو و حجم فزاینده اطلاعات و مدارک از سوی دیگر موجب می شد تا نمایه ها نتوانند به موقع در دسترس خواستاران قرارگیرند (حریری و توفیق، ۱۳۶۲، ۱). بنابر این تاخیر در نمایه سازی موضوعی مدارک و هزینه های گزاف مربوط به آن موجب شد تا نمایه سازان از کلید واژه های عنوان، به عنوان نقاط بازیابی (Access Point) استفاده کنند. اصولاً^۲ فکر استفاده از واژه های مدرک برای نمایه سازی قبل از پیدایش نمایه های موضوعی وجود داشته است. از قرن سیزدهم میلادی از این شیوه برای تهیه نمایه نوشته های مذهبی استفاده می شد. منابعی چون کشف آیات قرآن مجید را می توان در زمره اینگونه نمایه ها محسوب داشت. این شیوه به تدریج کاربردهای دیگری یافت و در شکلی محدودتر برای تهیه سایر راهنماها توصیه و استفاده شد. از جمله پانیزی (Panizzi) با اعتقاد بر این که هیچ گاه دو فهرستنویس در انتخاب موضوع برای کتابی واحد، رای واحدی ندارند و فقط واژه های منتخب از عنوان کتاب می تواند مانع عدم تجانس واژه های موضوعی شود، این روش را برای فهرستنویسی کتاب ها توصیه کرد و قواعد آن را که مشتمل بر ۹۱ قانون بود در سال ۱۸۵۰ برای فهرست کتابخانه ملی انگلستان وضع کرد. این شیوه با اندک تغییراتی در مناطق آلمانی زبان، هلند، اسپانیا و سوئد به صورت گسترده ای به کار گرفته شد و کرستادرو (Crestadro) از آن با تعدیلاتی برای فهرست کتابخانه عمومی منچستر استفاده کرد (آزادیان، ۱۳۵۲، ۱۰). علاوه بر این کلید واژه های عنوان در نمایه های موضوعی نیز حضور یافتند که از آن جمله به نمایه سازی درون بافتی (کوئیک^۳)، نمایه سازی برون بافتی (کووک^۴) و نمایه سازی جایگشتی^۳ می توان اشاره کرد.

^۱ Keyword In Context (KWIC)

^۲ Keyword Out of Context (KWOK)

^۳ Permuterm Index

در پی تحقیقات مختلفی که متخصصان علوم کتابداری و اطلاع رسانی انجام دادند نقش و اهمیت عنوان مدارک در فعالیت های ساماندهی و اشاعه اطلاعات بیش از پیش مشخص شد که به برخی از آنها در فصل دوم تحقیق اشاره شده است.

۱-۲- بیان مسأله :

اصولاً مواد غیر کتابی از نظر محتوی اطلاعاتی اهمیت زیادی دارند. معمولاً آثار پژوهشگران پیش از آن که به صورت کتاب در اختیار علاقه مندان قرار گیرد، در نسخه های محدود به شکل مقاله در نشریات یا سمینارها، پایان نامه ها، انتشارات دولتی و امثال اینها منتشر می شود. یکی از این منابع اطلاعاتی با ارزش پایان نامه ها هستند که همه ساله مقدار زیادی از امکانات مالی دولت و بخش قابل توجهی از نیرو و وقت اعضای هیأت علمی دانشگاه ها صرف آن می شود تا نتایج حاصل از آن پاسخی باشد برای یک یا بخشی از یک مسأله علمی یا اجتماعی که در عمل این دانش فردی به دلیل عدم انتشار به دانش اجتماعی تبدیل نمی شود (دیانی، ۱۳۷۸، ۱۴). بنابراین ضرورت وجود مراکزی احساس شد که به نگهداری و ساماندهی منابع غیر کتابی پرداخته و یافته های آن را به عنوان اطلاعات ارزشمند، به نوعی دسترس پذیر سازد.

در ایران پیشنهاد ایجاد یک مرکز ملی مدارک ابتدا توسط دبیرسازمان پیمان مرکزی "سنتو"^۱ دکتر مروین اسمیت (Mervin Smith) مطرح شد. در پائیز سال ۱۳۴۶ (دسامبر ۱۹۶۷) دکتر جان هاروی (John Harvey)، بورسیه فولبرایت (FullBright Scholar) در دانشگاه تهران براساس درخواست مروین اسمیت طرحی جهت ایجاد "مرکز اسناد ایران" و "مرکز آماده سازی کتاب تهران" به دکتر مجید رهنما، وزیر وقت علوم و آموزش عالی ایران ارایه داد. با موافقت دکتر رهنما و تصویب سازمان برنامه و بودجه از مهرماه سال ۱۳۴۷ دو مرکز فوق رسماً افتتاح و آغاز به کار کردند. در اردیبهشت سال ۱۳۴۸ پس از تأسیس "موسسه تحقیقات و برنامه ریزی علمی و آموزشی"، "مرکز اسناد ایران" به عنوان یکی از مراکز چهارگانه موسسه فوق فعالیت خود را

¹ CENTO: Central Treaty Organization

آغاز کرد. این مرکز در سال ۱۳۵۰ به "مرکز مدارک علمی ایران" تغییر نام یافت. در سال ۱۳۵۸ "موسسه تحقیقات و برنامه ریزی علمی و آموزشی" با کلیه واحدهای تابعه آن در وزارت فرهنگ و آموزش عالی ادغام شد. در سال ۱۳۶۱ موسسه فوق منحل گردید و "مرکز مدارک علمی ایران" تحت پوشش حوزه معاونت پژوهشی وزارت فرهنگ و آموزش عالی قرار گرفت. در نیمه دوم سال ۱۳۷۰ اساسنامه جدید مرکز به تصویب شورای گسترش آموزش عالی و به امضاء وزیر فرهنگ و آموزش عالی رسید. مطابق اساسنامه جدید "مرکز مدارک علمی ایران" به عنوان پژوهشکده شناخته شده و به "مرکز اطلاعات و مدارک علمی ایران" تغییر نام یافت (معترف، ۱۳۷۷، ۱۸). در شرح وظائف "مرکز اطلاعات و مدارک علمی ایران" (بند ۲ اساسنامه جدید) آمده است:

الف: تحقیق و مطالعه در نحوه تهیه اطلاعات و مدارک علمی تولید شده در داخل و خارج از کشور،

ب: تحقیق و بررسی روشهای سازماندهی و طبقه بندی اطلاعات و مدارک علمی به منظور تشکیل بانکهای اطلاعاتی و مراکز مدارک علمی در کشور،

ج: تحقیق و مطالعه پیرامون چگونگی اشاعه اطلاعات و مدارک علمی از طریق خدمات کتابخانه ای و به کارگیری تکنولوژی مناسب،

د: تحقیق و بررسی در مورد فعالیت کتابخانه های تخصصی و مراکز مدارک اطلاعات علمی کشور برای دست یابی به راههای مناسب بهبود فعالیت های آنها،

ه: تحقیق و بررسی در مورد فراهم نمودن امکانات مناسب مبادله اطلاعات و مدارک علمی با دیگر کشورها و استفاده از شبکه اطلاعات منطقه ای و بین المللی،

و: گردآوری، تجزیه و تحلیل و سازماندهی اطلاعات علمی به منظور تشکیل بانکهای اطلاعاتی،

ز: برگزاری دوره های تخصصی عالی و کوتاه مدت در چهارچوب ضوابط ومقررات آموزش عالی،

ح: گسترش اطلاعات علمی از طریق ایجاد شبکه اطلاع رسانی علمی کشور،

ط: انتشار نشریات در زمینه اطلاع رسانی،

ی: جمع آوری مدارک علمی کشور شامل پایان نامه ها، انتشارات دولتی و مستندات طرحهای تحقیقاتی،

ک: هماهنگی در گردآوری، تجزیه و تحلیل، سازماندهی و تقویت مراکز اطلاعاتی و کتابخانه‌های تخصصی کشور.

این مرکز در راستای وظایف قانونی خود به ایجاد پایگاه‌های اطلاعاتی مختلف پرداخت که یکی از آنها پایگاه "چکیده‌های پایان‌نامه‌های ایران (دکتر و کارشناسی ارشد)" است. تهیه مدارک یا تهیه پایان‌نامه‌های مربوط به این پایگاه بر عهده واحد مدیریت اسناد و اطلاعات است. کارکنان این واحد گاهی اوقات از طریق مکاتبه با دفاتر تحصیلات تکمیلی یا معاونت‌های پژوهشی دانشکده‌های مختلف و گاهی از طریق مراجعه به این مراکز، یک نسخه پایان‌نامه و یا اطلاعات مربوط به آن را در قالب کاربرگه‌های نمایه‌سازی دریافت می‌کنند. برخی از دانشکده‌ها نیز در مرحله تسویه حساب دانشجویان مقطع کارشناسی ارشد یا دکتری یک نسخه از پایان‌نامه این دانشجویان را به منظور ارسال به مرکز اطلاعات و مدارک علمی ایران دریافت کرده و در دوره‌های زمانی مختلف به مرکز ارسال می‌دارند (مانند دانشگاه امام صادق و دانشگاه تربیت مدرس). گاهی اوقات نیز دانشجویان ملزم هستند شخصاً یک نسخه از پایان‌نامه خود را به مرکز ارائه داشته و نامه‌اعلام وصول دریافت کنند (مانند دانشکده فنی و مهندسی و دانشکده علوم دانشگاه تهران). از سال ۱۳۸۰ نیز در سایت اینترنتی مرکز^۱ فرم‌هایی طراحی شده که علاقه‌مندان می‌توانند اطلاعات مربوط به پایان‌نامه‌های خود را در آن درج کنند. اطلاعات این قسمت پس از ویرایش به پایگاه اصلی افزوده می‌شود. این پایگاه شامل اطلاعات مربوط به پایان‌نامه‌های مختلف از سال ۱۳۱۴ به بعد است که در شش حوزه علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر طبقه‌بندی شده‌اند. از بهار ۱۳۷۲ اطلاعات موجود در این پایگاه در قالب "فصلنامه چکیده پایان‌نامه‌های ایران (دکتر و کارشناسی ارشد)" منتشر شد. این نشریه اطلاعات مربوط به پایان‌نامه‌های نگارش یافته از سال ۱۳۶۸ به بعد را پوشش داده و هر شماره آن شامل ۱۰۰۰ رکورد است. تا پایان سال ۱۳۸۱؛ چهل شماره از آن منتشر شده است. علاوه بر این از سال ۱۳۷۸، امکان جستجو و دسترسی به اطلاعات موجود در این پایگاه از طریق آدرس اینترنتی <http://database.irandoc.ac.ir> نیز فراهم شده است. نمایه‌سازان مرکز

^۱ <http://database.irandoc.ac.ir/scripts/wxis.exe>

پس از دریافت پایان نامه در قالب فرم های خاصی (پیوست شماره ۱) اطلاعات مربوط به این فیلدها را کامل می کنند: شماره مدرک، نام و نام خانوادگی، عنوان پایان نامه، استاد راهنما، درجه (دکتری، کارشناسی ارشد)، رشته، محل تحصیل، تاریخ نگارش، نام دانشگاه، نام دانشکده، بر ساخت (شامل: صفحه، تصویر، جدول، نمودار، نقشه، واژه نامه و کتابنامه)، زبان نگارش، توصیفگر فارسی و توصیفگر انگلیسی. اختصاص توصیفگرها با استفاده از عنوان، چکیده، مقدمه، پیشگفتار، فهرست مندرجات، نتایج، منابع و مآخذ، فرهنگ ها و واژه نامه های تخصصی صورت می پذیرد و پایان نامه ها به دو شیوه طبیعی و کنترل شده نمایه سازی می شوند (طلوعی، ۱۳۸۱، ۵۱).

حال مسائلی که در این خصوص مطرح می شود این است که در نمایه سازی پایان نامه ها "عنوان" تا چه حد می تواند در ارائه محتوی این مدارک نقش داشته باشند؟، نمایه سازان مرکز اطلاعات و مدارک علمی ایران تا چه حد از عنوان پایان نامه ها در نمایه سازی بهره می گیرند؟، این بهره گیری در حوزه های مختلف، چه تفاوتی دارد؟، آیا با افزایش کلید واژه های عنوان میزان همخوانی بین کلید واژه ها و توصیفگرها افزایش می یابد؟ و... با توجه به این موارد اگر میزان همخوانی کلید واژه های عنوان و توصیفگرها زیاد باشد می توان از کلید واژه های عنوان، که زبان طبیعی یا به عبارتی زبان استفاده کنندگان است در نمایه سازی استفاده کرد و با بهره گیری از شیوه های نمایه سازی ماشینی در وقت و هزینه نمایه سازی دستی، ویراستاری، حروفچینی، تصحیح و به طور کلی ذخیره و بازیابی اطلاعات صرفه جوئی نمود.

۱-۳- هدف های پژوهش :

هدف این پژوهش بررسی ساختار و اجزای تشکیل دهنده عنوان پایان نامه ها در حوزه های مختلف، نشان دادن تشابهات و تمایز های عنوان پایان نامه های هر حوزه و نیز بررسی ارتباط این اجزاء (به ویژه کلید واژه های آن) با توصیفگرهایی است که به منظور نمایه سازی در پایگاه "چکیده پایان نامه های ایران" از آنها استفاده شده است. نتایج حاصل از این تحقیق نشان خواهد داد که عنوان پایان نامه ها تا چه حد بیانگر محتوی موضوعی است، و به منظور نمایه سازی ماشینی پایان نامه ها - که عمدتاً با استفاده از کلید واژه های عنوان

صورت می گیرد و در مقایسه با نمایه سازی کنترل شده از هزینه کمتر و سرعت بیشتری برخوردار است - تا چه حد می توان به عنوان متکی بود. علاوه بر این یافته های این پژوهش متخصصان نمایه سازی را در اتخاذ تصمیم های مناسب برای افزایش نقاط دستیابی به منابع اطلاعاتی و نیز بهبود شیوه های ذخیره و بازیابی اطلاعات پایان نامه ها کشور یاری می دهد.

۱-۴- سوالات پژوهش :

- ۱- چه تفاوتی بین تعداد متوسط کلید واژه های عنوان پایان نامه ها در حوزه های علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر وجود دارد؟
 - ۲- چند درصد توصیفگرها، از کلید واژه های عنوان گزینش شده اند؟
 - ۳- در حوزه های علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر چه تفاوتی بین میزان همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان وجود دارد؟
 - ۴- آیا با افزایش تعداد کلید واژه های عنوان، میزان همخوانی آنها با توصیفگرها نیز افزایش می یابد؟
 - ۵- بین تعداد کلید واژه های عنوان پایان نامه ها با تعداد توصیفگرهای اختصاص یافته چه رابطه ای وجود دارد؟
 - ۶- از نظر شیوه نگارش، توصیفگرها چگونه هستند؟
 - ۷- در پایگاه "چکیده پایان نامه های ایران"، بین میزان بازیافت با استفاده از کلید واژه های عنوان و توصیفگرهای نمایه سازان چه تفاوتی وجود دارد؟
 - ۸- چه ميزانی از توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه های فارسی موجود همخوانی دارد؟
 - ۹- توصیفگرهای نمایه سازان تا چه حد با رسم الخط فارسی همخوان است؟
- ۱-۵- تعریف عملیاتی اجزا متغیرها :

- واژه (Word): یک نماد زبانی که دارای کارکرد خاصی در فرایند برقراری ارتباط بوده و معنی و مفهوم آن وابسته به متن است (هلموت فلبر، ۱۳۸۱، ۱۴۵). در این تحقیق کلیه کلمات یا عبارات های دارای بار معنایی و بدون بار معنایی (مانند حروف ربط، حروف اضافه و ...) به عنوان "واژه" در نظر گرفته شده است.

- کلید واژه (Key word): کلید واژه اصطلاح مهمی است که غالباً از عنوان یا چکیده مدرک به دست می آید و مبین محتوی موضوعی آن است (استلاکینن، ۱۳۷۸، ۶۰). معمولاً هر کلید واژه، یک موضوع (subject) از یک متن یا مدرک است که مفهوم معینی را در بر دارد، و این مفهوم در نقطه معینی از متن (یا نقاط معینی از آن) تشریح شده است. محمدی فر (۱۳۸۰، ۶۷ و ۶۴)، کلید واژه را بر دوتنوع می داند:

۱- کلید واژه استخراج شده از عنوان یا کلید واژه عنوانی (Title-based keyword): کلید واژه ای است که در عنوان و یا تیتیر یک مطلب یا مدرک موجود است، و برای استخراج آن نیازی به مطالعه متن مطلب یا مدرک نیست.

۲- کلید واژه استخراج شده از متن یا کلید واژه متنی (Text-based keyword): کلید واژه ای است که از داخل یک مدرک یا مطلب استخراج شده است. در داخل متن، معمولاً تعداد زیادی کلیدواژه وجود دارد که هر یک فقط جزء کوچکی از محتویات متن را نشان می دهد. برای انتخاب کلید واژه های یک متن، معیارهای دقیقی وجود ندارد و معمولاً آنها را بر اساس تجربه انتخاب می کنند.

در تحقیق حاضر به منظور اجتناب از نظرات و تجربیات شخصی در انتخاب کلید واژه های عنوان؛ و نیز با توجه به ضرورت دانش تخصصی در تشخیص کلید واژه رشته های مختلف، تمام کلمات و عبارت های دارای بار معنایی در عناوین پایان نامه ها، به عنوان کلید واژه مورد بررسی قرار گرفته و تفاوت آن با واژه در این است که کلمات یا لغات بی معنی را شامل نمی شود.

- **توصیفگر (Descriptor):** کلمه یا عبارتی را گویند که دارای بار اطلاعاتی بوده و منعکس کننده محتوی مدرک باشد (حریری، ۱۳۶۱، ۶۷). توصیفگر عملاً همان موضوع یا کلید واژه است که از بین واژگان استاندارد شده انتخاب می شود و بهتر است که در نمایه به کار برده شود. مثلاً یک سازمان نشر ممکن است برای مبحث کامپیوتر حدود ۵۰۰ توصیفگر استاندارد انتخاب کرده باشد و از نمایه سازان درخواست کند که موضوعات و کلید واژه های منتخب خود را با آنها تطبیق دهند. برای نمونه، در توصیفگرهای این مبحث، ممکن است به جای واژه کامپیوتر، از واژه رایانه استفاده شده باشد، و به جای چهار واژه "جی یو آی"، "سی ال آی"، "سیستم ناظر" و "سیستم عامل" فقط از یک واژه "سیستم عامل" استفاده شده باشد. گاهی از اصطلاحنامه به عنوان فهرست توصیفگرها استفاده می شود، زیرا در برابر هر مدخل اصطلاحنامه، مشخصات واژه نیز آمده است و نمایه ساز با استفاده از آن می تواند تصمیم بگیرد که یک واژه را در نمایه به کار برد یا خیر (محمدی فر، ۱۳۸۱، ۸۷). در تحقیق حاضر منظور از توصیفگر واژه های انتخاب شده توسط نمایه سازان مرکز اطلاعات و مدارک علمی ایران است که از آنها به منظور وصف محتوی، یا موضوع پایان نامه ها در نمایه سازی استفاده شده است. نمایه سازان این مرکز توصیفگرها را از عنوان، چکیده، متن پایان نامه ها یا منابعی چون اصطلاحنامه ها، واژه نامه ها یا فرهنگ های تخصصی گزینش می کنند.

- **حوزه (Class, Field):** حوزه یا رده بزرگترین جزء از رده بندی دانش بشری است که مجموعه ای از رشته های تخصصی نزدیک به هم را می پوشاند: مثلاً حوزه علوم پایه، حوزه پزشکی و حوزه هنر که هر یک از آنها، ده ها یا صدها رشته تخصصی نزدیک به هم را در بر می گیرد. تعداد حوزه های کل دانش بشری محدود و دارای فهرست معلومی است، که بستگی به تعریف و نحوه رده بندی دارد. به عنوان مثال، در رده بندی کنگره حدود ۲۰ رده مختلف وجود دارد که می توان هر یک را، یک حوزه به حساب آورد (محمدی فر، ۱۳۸۰، ۵۸). در تحقیق حاضر عنوان پایان نامه ها در رشته های مختلف

در شش حوزه علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر بررسی شده است. گفتنی است این حوزه ها بر اساس دسته بندی اعمال شده در پایگاه و نشریه "چکیده پایان نامه های ایران" صورت گرفته است. در این پایگاه کلیه پایان نامه ها در شش طبقه علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر دسته بندی شده اند.

- رشته (Discipline): رشته یک دانش تخصصی است که جزئی از یک حوزه محسوب می شود. مثلاً رشته علوم کتابداری و اطلاع رسانی، رشته مهندسی نساجی و رشته دندان پزشکی (محمدی فر، ۱۳۸۰، ۶۰). در این تحقیق منظور از رشته عنوان دوره های تحصیلی در مقطع کارشناسی ارشد است، که در پایگاه "چکیده پایان نامه های ایران" به عنوان تقسیمات فرعی شش حوزه علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، علوم کشاورزی و هنر در نظر گرفته شده و اطلاعات مربوط به پایان نامه ها ذیل هر یک دسته بندی شده است. مانند رشته علوم کتابداری و اطلاع رسانی، رشته روان شناسی، رشته متالورژی و رشته دامپزشکی.

- انطباق یا همخوانی (Matching): منظور هر یک از انواع انطباق یا همخوانی است که به شرح زیر می تواند واقع شود:

۱- همخوانی کامل (Overlap): به این معنی که کلید واژه عنوان پایان نامه، دقیقاً به همان صورت به عنوان توصیفگر استفاده شده باشد. مثلاً کلید واژه "حق مؤلف" که در عنوان به کار رفته، دقیقاً به همین شکل نیز (حق مؤلف) به عنوان توصیفگر گزینش شده باشد.

۲- همخوانی نسبی (Relative Matching): بدین معنی که توصیفگر با کلید واژه عنوان دارای اختلاف جزئی است. مانند اختلاف مفرد و جمع ("پایگاههای اطلاعاتی" به عنوان کلید واژه و "پایگاه اطلاعاتی" به عنوان توصیفگر)، اختلاف صورت های مختلف املائی ("طهران" به عنوان کلید واژه و "تهران" به عنوان توصیفگر)، اختلاف عبارات اسمی یا مصدری ("نفی" به عنوان کلید واژه و "نفث" به عنوان توصیفگر) و

اختلاف شکل مقلوب اسامی ("محمد تقی بهار" به عنوان کلید واژه و "بهار، محمد تقی" به عنوان توصیفگر).

- **عدم همخوانی:** بدین معنی که کلید واژه های عنوان هیچ گونه همخوانی با توصیفگر های اختصاص داده شده توسط نمایه سازان نداشته باشند. به عنوان مثال در پایان نامه ای با عنوان "نقش نقاشی در تعلیم و تربیت کودک" از توصیفگرهای "تصویرها / رنگ ها / مسائل روانی" استفاده شده که هیچ کدام با کلید واژه های عنوان همخوانی ندارند.

- **بازیافت (Recall):** عبارت است از درصد مدارک مربوطی که عملاً از میان کل مجموعه مدارک مربوط در فایل بازیابی شده است (پائو، ۱۳۷۸، ۱۱۵). بازیافت توانمندی نظام را در بازیابی مدارک مربوط مورد سنجش قرار می دهد (پائو، ۱۳۷۹، ۴۰۳). در این تحقیق منظور از میزان بازیافت تعداد پایان نامه هایی است که به واسطه جستجو از طریق کلید واژه عنوان یا توصیفگرها و در پاسخ به جستجوی یک موضوع خاص در پایگاه "چکیده پایان نامه های ایران" بازیابی شده اند.

- **شیوه نگارش:** در این تحقیق شیوه نگارش توصیفگرها با توجه به مواردی چون تک واژه، ترکیب واژه، مفرد و جمع بودن بررسی و ارزیابی شده است.

- **تک واژه (Uniterm):** تک واژه ساده ترین واژه واحد، مفرد و بسیطی است که یک مفهوم را القاء می کند (پوری سلطانی، ۱۳۷۹، ۱۰۲). در این تحقیق منظور از تک واژه، وجود کلماتی در عنوان پایان نامه ها است که از یک ریشه تشکیل شده (مانند قدرت) و یا کلماتی که از ترکیب یک ریشه با پسوند یا پیشوند ساخته شده باشند (مانند قدرتمند).

- **واژه های ترکیبی (Compound word):** واژه های ترکیبی یا مرکب، کلمه ای است که حداقل از دو کلمه بسیط تشکیل می شود (محمدی فر، ۱۳۸۱، ۲۵۲). در تحقیق حاضر همین تعریف مد نظر بوده و کلماتی چون: "کتابخانه" یا "بانک اطلاعاتی" واژه های ترکیبی محسوب شده اند.

فصل دوم

بررسی مطالعات گذشته

۲-۱- مروری بر مبانی نظری :

در دوران پس از جنگ جهانی دوم حجم انتشارات، خصوصاً انتشارات علمی و تخصصی به طور فوق العاده‌ای افزایش یافته است. در هر موضع صدها کتاب، مجله و گزارش انتشار می‌یابد و پدیده‌ای را به وجود می‌آورد که گاه "انفجار انتشارات" و گاه "انفجار اطلاعات" نامیده می‌شود. دیگر تنها "نوشته" برای ثبت اطلاعات و انتقال آن کافی نیست و از وسائلی چون فیلم، میکروفیلم، صفحه، نوار و ... برای ثبت و انتقال اطلاعات به استفاده می‌شود. یکی از نتایج این امر آن است که روشهای قراردادی طبقه‌بندی به طور عمده به کار تنظیم کتاب در قفسه کمک می‌کند و سرعنوانهای موضوعی نیز به طور کلی برای کتاب مناسب است. از مشکلات دیگر به کار بردن روشهای قراردادی طبقه‌بندی و سرعنوانهای موضوعی رشد سریع حجم فهرست است که در نتیجه آن یک سلسله مسائل چه از نظر نگهداری فهرست و چه از نظر جستجو در آن، به وجود می‌آید (سینایی، ۱۳۵۱، ۱). به طور کلی به جهت مسائلی چون نبودن روش های قراردادی رده بندی و سرعنوان های موضوعی برای تجزیه و تحلیل مواد غیر کتابی، کمی وقت، حجم زیاد اطلاعات، تحول در تخصصی شدن علوم، به وجود آمدن نیازهای اطلاعاتی جدید، دگرگونی در اهمیت منابع اطلاعاتی و ... در نظام های ذخیره و بازیابی اطلاعات تحولاتی به وجود آمد که فعالیت نمایه سازی از جمله مهم ترین آنها ست. در نظام رده بندی و فهرستنویسی برای باز نمودن محتوی موضوعی کتاب ها و انتقال مفاهیم مندرج در آنها به استفاده کننده، از جدول های رده بندی و سرعنوان موضوعی استفاده می شود که شامل واژه های بسیط یا مرکب از دو یا چند واژه به شکل عبارت و گاه شبه جمله است که با نظم الفبایی تدوین و ارائه شده است. اختصاص دو یا چند سر عنوان موضوعی به کتاب یا مدرک، تنها محتوی کلی آن را باز می نماید، در حالی که در مورد مدارک غیر کتابی، به ویژه گزارش های علمی و فنی که گستردگی موضوعی آنها بسیار بالا است، سر عنوان های مذکور نمی توانند مفاهیم خاص و متعدد را منتقل کنند (آقا بخشی، ۱۳۷۶، ۳). از این رو، برای باز نمودن محتوی موضوعی مدارک غیر کتابی از روش هایی چون نمایه سازی و ابزارهایی مانند اصطلاحنامه ها استفاده می شود.

به طور کلی تا کنون نمایه سازی به سه شیوه یا زبان صورت پذیرفته که عبارت اند از:

الف - نمایه سازی به زبان طبیعی^۱؛

ب - نمایه سازی به زبان آزاد^۲؛

پ - نمایه سازی کنترل شده^۳.

۲-۱-۱- نمایه سازی به زبان طبیعی :

نمایه سازی به زبان طبیعی، مبتنی بر مدارکی است که نظام با آنها سر و کار دارد. در این شیوه واژه های طبیعی موجود در مدارک (واژه های عنوان، چکیده یا بر گرفته از تمام متن) برای نمایه سازی به کار می رود. یکی از ویژگی های نمایه سازی به زبان طبیعی عدم به کارگیری واژگان کنترل شده در آن است. این امر اجازه می دهد تا انواع متفاوت کلمات موجود در زبان طبیعی در نمایه منعکس شود. بیشتر نظام هایی که به نمایه سازی به زبان طبیعی روی می آورند فهرستی از واژه های غیر مجاز فراهم کرده و در نظام ذخیره می کنند که مانع نمایه سازی واژه های ناخواسته می شود. در نمایه سازی به زبان طبیعی که فقط از فهرست غیر مجاز استفاده می کند، زبان نمایه سازی باز است و نمایه سازی به تدریج و با تحول زبان طبیعی مدارکی که به نظام راه می یابد تغییر می کند. روش دیگر این است که همراه با واژه های غیر مجاز، واژه های مجاز نیز به رایانه داده شود. واژگان مجاز شامل تمام کلمات یا عباراتی می شود که مدخل های سودمندی در حوزه موضوعی هستند که نمایه سازی در آن موضوع انجام می گیرد. بنابر این، واژگان مجاز باید قبلاً در رایانه ذخیره شده باشد. فهرست واژگان مجاز، مانند فهرست واژگان غیر مجاز می تواند هر چند وقت یکبار به نسخه ای چاپی تبدیل شده و مورد بررسی و تجدید نظر قرار گیرد. فهرست واژگان مجاز گاه اصطلاحنامه نامیده می شود و در واقع نوعی اصطلاحنامه زبان طبیعی است. در فرایند نمایه سازی گاه به گاه اصطلاحاتی جدید در مدارک وجود دارد که در فهرست واژگان مجاز و فهرست واژگان غیر مجاز وجود ندارند. در این موارد نمایه ساز درباره افزودن واژه به هر

¹ Natural language Indexing

² Free language Indexing

³ Controlled Vocabulary Indexing

یک از این دو فهرست یا نادیده گرفتن واژه تصمیم می گیرد (تیلور، ۱۳۸۱، ۱۷۹). اصولاً فکر استفاده از واژه های مدرک برای نمایه سازی به قبل از پیدایش نمایه های موضوعی باز می گردد. از قرن سیزدهم میلادی از این شیوه برای تهیه نمایه نوشته های مذهبی استفاده می شد و این نمایه های موضوعی - الفبایی وسیله مهمی برای بازیابی اطلاعات نوشته های مذکور به شمار می آمدند. منابعی چون کشف آیات قرآن مجید را می توان در زمره اینگونه نمایه ها محسوب داشت. این شیوه به تدریج کاربردهای دیگری یافت و در شکلی محدودتر برای تهیه سایر راهنماها توصیه و استفاده شد. از جمله پانیتسی (Panizzi) با اعتقاد بر اینکه هیچگاه دو فهرست نویس در انتخاب موضوع برای کتابی واحد، رای واحدی ندارند و فقط واژه های منتخب از عنوان کتاب می تواند مانع عدم تجانس واژه های موضوعی شود، این روش را برای فهرست نویسی کتابها توصیه کرد و قواعد آن را که مشتمل بر ۹۱ قانون بود در سال ۱۸۵۰ میلادی برای فهرست کتابخانه ملی انگلستان وضع کرد. این شیوه با اندک تغییراتی در مناطق آلمانی زبان، هلند، اسپانیا و سوئد به صورت گسترده ای به کار گرفته شد و کرستادرو (Crestadro) از آن با تعدیلاتی برای فهرست کتابخانه عمومی منچستر استفاده کرد (حریری و توفیق، ۱۳۶۲، ۳). به طور کلی طرح نمایه سازی تک واژه ای بر مبنای پذیره های زیر قرار دارد:

۱. اصطلاح های موجود در خود مدرک، گویاترین واژه های محتوای آن است، زیرا آنها واژه های اصلی از خود پدید آورنده می باشد.
۲. نمایه ساز به عنوان واسطه عمل کرده و بودن تفسیر وی از واژه های به کار رفته توسط خود نویسنده، نمایه موجود را یکپارچه تر و گزینش اصطلاح ها را عینی تر می کند.
۳. هر اصطلاح محتوایی مستقیماً به منزله برچسب رده برای طبقه ای از مدارک به کار می رود.
۴. اگر جستجو گر بتواند موضوع مورد جستجو را از روی مفاهیم تشکیل دهنده آن تحلیل کند، بازیابی ثمر بخش تر است.
۵. از آن جا که روی مفاهیم مبتنی بر منطق بولی است، اگر از رایانه برای پردازش استفاده شود، سرعت بازیابی تا حد بسیاری بالا می رود (پائو، ۱۳۸۱، ۲۰۰).

اختراع کامپیوتر و نرم افزاری که به وسیله کامپیوتر می توانست حروف الفبا را تشخیص دهد، سرآغاز تحولی عظیم در داده پردازی و تهیه منابع اطلاعاتی گردید. استفاده از کامپیوتر برای تولید نمایه هایی که می توانست به سرعت تهیه و در بین متخصصان توزیع شود، اولین اقدام در جهت کاربرد عملی و گسترده کامپیوتر در زمینه اطلاع رسانی بود. اساس تهیه این نمایه ها همان واژه گزینی از مدارک بود که بر حسب روشهای متفاوتی که در پیش گرفته می شد، واژه ها به طور خودکار از عنوان، چکیده یا متن کامل مدرک و یا ترکیبی از آنها استخراج می شدند و به صورت اصطلاحات نمایه، وسیله بازیابی اطلاعات و مدارک قرار می گرفتند. متداول ترین نوع این نمایه ها، نمایه هایی بود که بر اساس عنوان تهیه می شد و در این کار فرض بر این بود که عنوان هر اثر مبین محتوای موضوعی آن اثر است، لذا نمایه هایی که بر پایه عنوان مدارک ساخته شوند می توانند جوینده را در یافتن مدارک متناسب با نیازش یاری کنند.

روال کار در تهیه نمایه های ماشینی عنوان، معمولاً گردش دادن عنوان مدرک است و به همین جهت این نمایه ها به "نمایه گردان" معروف شده اند. مراد از "نمایه گردان"، نمایه ای است که در آن واژه های مهم عنوان های مجموعه ای از مدارک در نظم ویژه ای به گردش در می آیند و هر یک از این واژه ها به نوبت مدخلی برای ردیابی و بازیابی اطلاعات قرار می گیرند. بنابراین، تعداد مدخلهایی که برای بازیابی یک مدرک وجود دارد، حداقل برابر با تعداد کلیدواژه های عنوان آن است.

انواع نمایه گردان

۱. نمایه کوئیک^۱ یا نمایه درون بافتی

یکی از ابتدایی ترین و در عین حال موفق ترین نمایه های ماشینی عنوان "نمایه کوئیک" یا درون بافتی است. گر چه نام کوئیک معمولاً با نام هانز پیترلوهن (Hans Peter Luhn) قرین است، اما در واقع لوهن و هربرت اولمان (Herbert Ohlman) را می توان مشترکاً از بدعت گذاران این روش دانست. این دو همزمان ولی مستقل درصدد برآمدند تا با چرخاندن عنوان مجموعه ای از مدارک، نمایه هایی را تولید کنند و در

¹ Keyword In Context (KWIC)

کنفرانس بین المللی اطلاعات علمی که در اواخر ۱۹۸۵ در واشنگتن تشکیل می شد، بین شرکت کنندگان توزیع نمایند. بنابر اظهار "اولمان" در همان زمان دست کم پنج نفر دیگر در آمریکا روی این روش مشغول آزمایش بودند (معرف زاده، ۱۳۶۶، ۴۸). این روش توسط سازمانهای مختلف اطلاع رسانی، به ویژه برای نمایه سازی منابع فنی، به سرعت پذیرفته شد و جوامع حرفه ای از لحاظ نقشی که این نمایه می توانست در اشاعه اطلاعات جاری داشته باشد سخت بدان علاقه مند شدند.

در نمایه درون بافتی تمام کلمات موجود در عنوان مدرک مورد نمایه سازی توسط برنامه ای که به کامپیوتر داده می شود با سیاهه ای از واژه های بازدارنده^۱ مطابقت داده می شود. کلماتی که در این سیاهه وجود ندارند، به طور خودکار کلیدواژه محسوب می شوند. مدخلی که با هر یک از این کلیدواژه ها درست می شود به صورتی است که این واژه را در درون بافت خود، یعنی همراه با بقیه کلمات آن عنوان، در شکل و نظم طبیعی نشان می دهد. این بافت شامل همه واژه های بازدارنده نیز می شود و به این ترتیب، گرچه واژه های غیرمجاز در نمایه نامه مدخل قرار نمی گیرند، اما همراه با سایر کلمات عنوان آورده می شوند تا محیط کلیدواژه و معنا و مفهوم آنرا دقیق تر آشکار سازند و ارتباط میان کلیدواژه های عنوان را بهتر نمایش دهند. ابداع کنندگان این روش چنین پنداشته بودند که چشم در دنبال کردن واژه های الفبایی شده مدخل، کلمات محیطی هر دو جانب مدخل را نیز می گیرد و این کلمات محیطی به شخص کمک می کنند تا به فوریت دریابد که عنوان مربوط با نیاز وی مناسبت دارد یا خیر.

از نظر شکل ظاهری، در نمایه درون بافتی برای هر مدخل فقط یک خط جا در نظر گرفته می شود که این خط معمولاً شامل سه بخش است: واژه نمایه یا مدخل، بافت و نشانه بازیابی یا ارجاع به اطلاعات کامل کتابشناسی مدرک. واژه های نمایه به صورت ستونی در مرکز یا سمت چپ صفحه و یا سمت راست آن (در فارسی) به ترتیب الفبایی ظاهر می شوند و با همین کلمات است که جوینده، جستجوی خود را در نمایه آغاز می کند. برای مشخص کردن پایان عنوان، در هر نمایه از علامت قراردادی خاصی نظیر علامت جمع یا مساوی و

¹ Stop List

یا خط اریب (/) و غیره استفاده می شود تا استفاده کننده بتواند به سهولت ابتدای عنوان را پیدا کرده و عنوان را در شکل طبیعی خود به سرعت مرور نماید. به عنوان مثال در این شیوه، مدرکی با عنوان "اقتصاد مسکن در ایران" بدین شکل نمایه می شود:

اقتصاد مسکن در ایران/

ایران / اقتصاد مسکن در

مسکن در ایران/ اقتصاد

در عنوان فوق واژه "در" بار اطلاعاتی چندانی نداشته و به عنوان کلمه غیرمجاز یا بازدارنده محسوب شده است.

از قدیمی ترین و معتبرترین نمایه های درون بافتی می توان عناوین شیمی (Chemical Titles) را نام برد که از ۱۹۶۱ شروع به انتشار کرده است. در نمایه درون بافتی با توجه به محدودیتی که از لحاظ گنجایش هر خط یا هر شناسه وجود دارد، چنانچه طول عنوان از حد تعیین شده بیشتر باشد، قسمتهایی از آن حذف می شود. این عمل گرچه تناسب موردنظر را در صورت ظاهری نمایه ایجاد می کند؛ اما باعث مبهم شدن عنوان گردیده و موجب می شود که مراجعه کننده به حدس و گمان متوسل شود یا با فرض مناسب بودن مدرک با نیازش، به جستجوی سایر اطلاعات کتابشناسی آن بپردازد، در حالیکه چیزی جز اتلاف وقت نصیبش نمی شود.

۲. نمایه کووک^۱ یا نمایه برون بافتی

برای جبران مسائل ناشی از محدود بودن جا، کوتاه شدن عنوان و نیز مشکلات خواندن نمایه درون بافتی در اثر استفاده از حروف بزرگ در چاپ، نوع دیگری از نمایه عنوان پا به عرصه وجود گذاشت که آنرا اختصاراً "کووک" نامیدند. در این نمایه، هر کلیدواژه به ترتیب از عنوان خود خارج شده و مقدم بر سایر اجزای آن قرار می گیرد. سپس عنوان به ترتیب طبیعی خود - در حالی که کلیدواژه هم جزو آن است - در

¹ Keyword Out of Context (KWOK)

زیر این واژه یا به دنبال آن می آید. به این ترتیب برای هر واژه مهم یک مدخل ساخته می شود. برای مثال

عنوانی که قبلاً ذکر شد به صورت زیر نمایه می شود:

اقتصاد	اقتصاد مسکن در ایران
ایران	اقتصاد مسکن در ایران
مسکن	اقتصاد مسکن در ایران

گرچه در این نمایه هم بافت مشخص است اما جستجو به دنبال عبارات یا کلمات ترکیبی معمولاً کندتر صورت می گیرد، زیرا همیشه از تک واژه به صورت مجرد و خارج از بافت خود استفاده می شود. در بعضی از نمایه نامه های برون بافتی برای صرفه جویی در جا یا سرعت بخشیدن به درک موقعیت مفهومی کلیدواژه، از تدابیر مختلفی استفاده می شود. مثلاً برای آنکه هر مدخل، دوباره در محل اصلی خود در داخل عنوان تکرار نشود، به جای آن علامتی نظیر ستاره گذاشته می شود یا عنوان هایی که کلیدواژه مشترک دارند همه زیر یک مدخل جمع می شوند. علاوه بر این، غالب آنها اطلاعات کامل کتابشناسی مدرک را به دنبال عنوان می آورند و به این ترتیب استفاده از نمایه را - به خصوص در مواردی که عناوین از منابع گوناگونی اخذ شده اند - آسانتر می کنند. نمایه برون بافتی را به این ترتیب بیشتر می توان شبیه نمایه نامه های موضوعی دانست با این تفاوت که در نمایه نامه های موضوعی، مدخل ها توسط نمایه ساز برای مفاهیم تعیین می شوند. به اعتقاد بسیاری از صاحب نظران نمایه درون بافتی، هم از نظر استفاده از جا بیشتر مقرون به صرفه است (چون هر مدخل فقط به اندازه یک خط جا اشغال می کند) و هم از لحاظ بازیابی مطالب آسانتر است (زیرا کلیدواژه را در درون بافت عنوان عرضه می کند).

۳. نمایه جایگشتی^۱

اساس این نمایه هم، گردش کلیدواژه های عنوان است. منتهی این گردش به طور محدود با دو واژه نشان داده می شود. کلیدواژه های موجود در هر عنوان به صورت جفت انتخاب می شوند، بدین معنی که اگر در عنوانی پنج کلیدواژه وجود داشته باشد برنامه کامپیوتر از آنها ده جفت اصطلاح می سازد که هر جفت با جا به جا شدن، دو بار در نمایه نامه وارد می شود و در مجموع بیست مدخل را تشکیل می دهند. به عبارت دیگر، برای هر عنوان $(n-1)$ مدخل درست می شود که n تعداد کلیدواژه های موجود در آن است. در این نمایه اصطلاح دوم یا اصطلاح همراه در محدود کردن مفهوم واژه مدخل و اخص شدن معنای آن نقش موثری دارد و در واقع همان طور که صفت اسم را توصیف می کند، این واژه ها نیز واژه مدخل یا موضوع را توصیف می کند. این خصوصیت به استفاده کننده کمک می کند تا نسبت به ارتباط مدرک یا مدارک با نیازش سریع تر و قاطعانه تر تصمیم بگیرد. البته واژه های همراه پس از استخراج از عناوین مربوط، در زیر واژه مدخل به ترتیب الفبایی تنظیم می شوند و در مقابل هر یک نشانه یا نشانه های بازیابی آن قید می گردد و به این ترتیب امکان مراجعه به اطلاعات کتابشناسی مدرک فراهم می آید. در این نمایه ها غالباً غیر از سیاهه غیرمجاز از سیاهه دیگری به نام سیاهه نیمه مجاز نیز استفاده می شود. واژه های این سیاهه گرچه از نظر بازیابی اطلاعات ممکن است مفید باشند اما به سبب آنکه ارزش کلیدواژه ای ندارند مدخل قرار نمی گیرند. به عنوان نمونه در مثال قبلی نمایه جایگشتی به صورت زیر خواهد بود:

اقتصاد	ایران
اقتصاد	مسکن
ایران	اقتصاد
ایران	مسکن
مسکن	اقتصاد

¹ Permuterm Index

مسکن ایران (حریری و توفیق، ۱۳۶۲، ۱-۱۶).

از نمونه های معروف و معتبر این نوع نمایه گردان، نمایه هایی است که توسط موسسه اطلاعات علمی^۱ برای دو نشریه مشهورش - و به عنوان بخش سوم آنها - تهیه می شود.

توانایی ها و ضعف های نمایه گردان

نمایه سازی به شیوه گردان، مانند هر شیوه دیگر، توانایی ها و ضعف هایی دارد که باید قبل از استفاده به خوبی شناخته شوند. یکی از نقاط قوت آن عدم مداخله یا مداخله اندک انسان در آن است و همین امر را می توان منشأ اکثر محاسن و معایبی دانست که با نمایه های گردان همراهند. محاسن و معایب عمده این نمایه ها را می توان به شرح زیر خلاصه کرد:

۱- در مقایسه با نمایه های موضوعی بسیار ارزان، ساده و سریع تهیه می شوند. در تولید آنها نیاز به استفاده از خدمات نمایه سازان، ویراستاران، حروفچین ها، مصحح ها و غیره نیست و این موجب تقلیل فاحش هزینه ها خواهد شد. یک کارشناس نمایه سازی می تواند با استفاده از یک کامپیوتر و یک برنامه نرم افزاری انبوهی از عنوان ها را در مدت کوتاهی با سرعتی فوق العاده زیاد مرور کرده و واژه های مهم آنها را بی آنکه اشتباهی رخ دهد انتخاب کند. صرفه جویی در هزینه و زمان تولید از جمله عوامل عمده ای بوده که توجه ناشران را به تهیه این گونه نمایه ها جلب کرد. لازم به ذکر است که نمایه های گردان به روش دستی نیز تهیه می شود و در این صورت، کار نمایه سازی توسط افرادی انجام می گیرد که الزاماً متخصص موضوعی یا نمایه ساز حرفه ای نیستند بدین جهت کاهش هزینه ها در اینجا هم به چشم می خورد، اما به طور کلی مسائلی که در ارتباط با نمایه های گردان مطرح می شوند عمدتاً متوجه نظام های خودکار است.

^۱ موسسه اطلاعات علمی (Institute for Scientific Information) دو نشریه استنادی تحت نامهای Science Citation Index و

Social Sciences Citation Index به صورت ادواری منتشر می کند که هر کدام از آنها دارای سه بخش است:

Source Index، Citation Index و Permuterm Subject Index "بخش اخیر نمایه جایگشتی عناوین مدارکی است که اطلاعات کامل کتابشناسی آنها در بخش دوم آمده است.

۲- نمایه های گردان منعکس کننده واژگان رایج در حوزه مربوطه هستند. واژه هایی که به صورت مدخل استفاده می شوند کلماتی هستند که نویسندگان در عنوان اثر خود به کار برده است و از آنجا که نویسندگان علی الاصول در زمینه کار خود تبحر دارد از کلماتی برای نمایاندن محتوای مدرک استفاده می کند که دقیق و روزآمد هستند. اما همین خصوصیت باعث ناهماهنگی هایی می شود. نمایه نامه هایی که به صورت ادواری منتشر می شوند، در طول زمان واژگان ثابت و یکنواختی را عرضه نمی کنند، زیرا بعضی از واژه ها بر اثر پیشرفت حوزه مربوطه اشتقاق یافته و هر یک با دو یا چند واژه نو جایگزین می شوند. بعضی دیگر نیز به مرور از رواج افتاده و مفاهیم مورد نظر با واژه های جدید بیان می شوند. بنابراین، استفاده کننده ای که در پی مطلبی در این نمایه ها به جستجوی گذشته نگر می پردازد به لحاظ تغییرات واژگان با مشکلات متعددی رو به رو می شود.

۳- نمایه های گردان با فراهم آوردن بافت نسبی، تصویر روشنتری از موضوع می دهند. در بسیاری از رشته های فنی نظیر متالوژی و فیزیک هسته ای، عناوین حاوی اصطلاحات کاملاً تخصصی و متعددی هستند که همراه با سایر واژه های به کار رفته در آنها، موضوع مورد نظر نویسندگان را بخوبی بیان می کنند و به این ترتیب شاید بتوان گفت که این اصطلاحات در بافت خود از لحاظ توصیف موضوع مفیدتر از مدخلهای موضوعی صرف هستند. اما ایرادی که از اولین روز پیدایش نمایه گردان به آن گرفته شده و تا به حال نیز ادامه دارد این است که عنوان، معمولاً شاخص ایده آلی برای محتوای مدرک نیست، زیرا اگر چه نویسندگان - به ویژه در رشته های علمی - می کوشند طوری عنوان اثر خود را انتخاب کنند که دیگران به راحتی به محتوای آن پی ببرند، ولی معمولاً عبارت بندی آن را بدون توجه به استفاده اش در نمایه سازی انجام می دهند. بعضی از عنوانها حاوی واژه های کافی نیستند. بعضی دیگر حامل اصطلاحات پربار و اخص نمی باشند و بالاخره برخی هم به منظور جلب توجه یا رعایت ایجاز عمداً مبهم و گمراه کننده و یا به زبان استعاره، تمثیل، قیاس و غیره بیان شده اند. مسئله اخیر در رشته های علوم انسانی و اجتماعی بارزتر به چشم می خورد. مثلاً در عنوان هایی نظیر "آنچه خود داشت" (که در

زمینه علوم اجتماعی است) و "پوست سیاه صورتکهای سفید" (که در زمینه سیاسی است) نشانه های مشهودی برای پی بردن به محتوای واقعی مدرک وجود ندارد.

۴- در تهیه این نمایه ها، محتوای موضوعی مدرک توسط نمایه سازان مختلف تفسیر نمی شود و بدین جهت جایی برای اشتباه و اختلاف نظر از لحاظ تعیین و تبیین موضوع باقی نمی ماند. اگر واژه ای در عنوان مدرک وجود داشته باشد که جزو واژه های غیرمجاز نباشد مسلماً مدخلی برای آن منظور خواهد شد و این نوعی یکنواختی در کار پیش خواهد آورد. اما همین امر منجر به پراکندگی موضوعها خواهد شد. این پراکندگی در نمایه های گردان فراوان به چشم می خورد. یک مفهوم مشخص ممکن است توسط نویسندگان مختلف با لفظ های متفاوتی بیان شود و به این ترتیب مدخلهایی که قاعدتاً بایستی در یکجا گردآمده و با واژه واحدی بیان شوند، در محلهای مختلف و متعدد پراکنده می گردند. واژه های زیر نمونه هایی از آن هستند:

کودکان - خردسالان - اطفال

تجارت فرش - بازرگانی قالی

مرض قند - دیابت

با توجه به تعدد مترادفهای بعضی از واژه ها و نبودن هیچ گونه ارجاعی بین آنها؛ جستجوها یا ناقص و ناموفق رها می شوند و یا بار سنگینی از جهت کنترل همه مترادفها بر دوش جوینده گذاشته می شود که در این صورت لازم است جوینده نسبت به این پراکندگی ها آگاهی کامل داشته باشد. مسئله دیگری که در همین ارتباط بر اثر عدم کنترل واژگانی پیش می آید، مسئله املای متفاوت واژه ها و صورت مفرد و جمع آنهاست که موجب می شود به همان شکلی که در عنوان ظاهر شده اند در نمایه هم بیایند و مدخلهای متفاوتی را در محلهای مختلف تشکیل دهند. مسئله مهم دیگر مربوط به هم نویسه هاست و از آنجا که کامپیوتر قادر نیست بین معانی مختلف یک واژه تمایز قائل شود - و این یکی از مسائل مهم و جدی در ترجمه ماشینی است - همه آنها را به صورت واژه واحد به حساب می

آورد. کنترل اطلاعات تمام عنوان هایی که به این صورت زیر مدخلهای مشابه می آیند برای جوینده وقت گیر است. البته در نمایه های درون بافتی وجود واژه های اطراف مدخل موجب سهولت این کنترل می شود. نبودن یک نظم یا ساختار سلسله مرتبه ای نیز عامل دیگری است که در این نمایه ها باعث پراکندگی واژه ها و طولانی شدن جستجو ها می شود.

۵- گاهی طول مدخلهایی که با یک کلیدواژه درست می شود به یک یا چند صفحه بالغ می گردد. جستجو در میان این مدخلها استفاده کننده را کسل و خسته می کند خصوصاً اگر عنوان ها نتوانند با اطلاعات خود در رد فوری مدرک به او کمک کنند.

اصلاح نمایه گردان

ایرادهایی که به خاطر ضعف های فوق از جانب صاحب نظران به نمایه گردان گرفته می شد موجب گردید تا متخصصان و ناشران راههایی را برای رفع و اصلاح آنها چاره اندیشی کنند. تدابیری که بدین منظور شده مختصراً عبارتند از:

۱- اصطلاحاتی که تصور می رود برای راهنمایی جوینده به موضوعات مطرح شده به وسیله نویسندگان مفید باشند به عنوان مدرک اضافه می شوند. این اصطلاحات ترجیحاً از متن مدرک یا چکیده آن یا به عبارت دیگر از زبان نویسندگان گرفته می شوند و یا به وسیله نمایه سازان و ویراستاران به آن افزوده می شوند. این اصطلاحات اضافی یا مکمل می توانند نمایه سازی را جامع تر و عمیق تر کنند و اشکالات ناشی از ابهام یا کم نمایی عنوان ها را از بین ببرند. این واژه ها یا عبارات که به انتهای عنوان های نارسا اضافه می شوند مانند سایر کلیدواژه ها مدخل قرار می گیرند و معمولاً با علامتی نظیر پرانتز یا دو ابرو از سایر واژه ها مشخص می شوند. این واژه ها یا عبارات می توانند در تکمیل عنوان یا بسط آن نقش موثر و مهمی ایفا کنند. مثلاً در این عنوان "دیوار دو طرف دارد (چین نو)" افزودن عبارت (چین نو) ابهام عنوان را از بین می برد و یا در عنوان "رشد انتشارات در سالهای بعد از انقلاب (کتاب)" افزودن کلمه (کتاب) موجب اخص شدن اصطلاح انتشارات می شود. این اصطلاحات یا تعابیر در واقع همان

”یادداشت دامنه” هستند که در اصطلاحنامه هم برای روشن شدن مفهوم یک واژه یا محدود کردن معنای آن استفاده می شوند.

۲- با استفاده از سیاهه مجاز تمام واژه هایی که به عنوان اصطلاحات نمایه قابل قبول هستند کنترل می شوند. سیاهه مجاز عکس سیاهه بازدارنده عمل می کند یعنی موجب حذف همه واژه هایی می شود که در سیاهه نیستند. این سیاهه ها، برخلاف سیاهه های بازدارنده که با داشتن تعداد محدودی واژه می توانند جلوی نمایه شدن اکثر واژه های مفید (از لحاظ جستجو) را بگیرند، باید مشتمل بر چند هزار واژه و اصطلاح باشند تا بتوانند همه واژه های مفید زمینه موضوعی را بپوشانند. استفاده از سیاهه مجاز شیوه ای عملی برای غلبه بر مسئله مترادفهاست.

۳- کلیدواژه های عنوان، قبل از آن که اطلاعات به کامپیوتر داده شود، به صورت دستی تعیین می شوند. برای این کار لازم است علامت مشخصی که برای ماشین هم شناخته شده باشد جلوی هر کلید واژه گذاشته شود. این کار موجب می شود که اصطلاحات بنا به قابلیت کاربریشان در مدرک، به عنوان اصطلاحات نمایه یا مدخل انتخاب شوند و نه صرفاً به خاطر آنکه در عنوان مدرک آمده اند. علاوه بر این، با توجه به نیاز بعضی از زمینه ها می توان از اصطلاحات نمایه ای مرکب برای بیان مفاهیم استفاده کرد و آن ها را به صورت یک اصطلاح علامت گذاری نمود.

۴- عنوانی که مولف برای اثر خود داده است با عنوان مناسبتری جانشین می شود. در این کار که به وسیله ویراستار صورت می گیرد کوشش می شود تا آنچه که نویسنده در مطلب خود ارائه کرده است در عبارتی دقیق بیان شود. عنوان جدید طبیعتاً هم از عنوان اصلی طولانی تر است و هم از لحاظ عبارت بندی استانداردتر است.

۵- درونداد یا برونداد کامپیوتر به منظور کنترل پراکندگی موضوعات ویرایش می شود. کنترل هم نویسه ها و شکل مفرد و جمع واژه ها، کنترل اصطلاحات مرکب، حذف مدخلهایی که از نظر جستجو بی اهمیت

یا کم اهمیت هستند و وارد کردن ارجاعات، به ویژه برای مهار کردن مترادفها، از جمله کارهایی است که می توان به وسیله این ویرایش انجام داد.

اصلاحات و ویرایشهایی از قبیل آنچه در فوق ذکر شد مسلماً موجب بهتر شدن نمایه گردان و سهولت استفاده از آن خواهد شد، ولی آنچه که مورد بحث و انتقاد است این است که آیا این بهبود می تواند توجیهی قوی از نظر زیانهای ناشی از افزایش مخارج و تاخیر زمان انتشار به همراه داشته باشد و آیا این کوششها می تواند کیفیت نمایه های گردان را به حد نمایه های موضوعی نزدیک کند. به نظر بسیاری از متخصصان سرمایه گذاری نیروی انسانی در ویرایش این نمایه ها از لحاظ اقتصادی قابل توجیه نیست و اگر قرار باشد هزینه هایی این چنین که عمدتاً به عهده استفاده کننده گذاشته خواهد شد مدنظر قرار گیرد بهتر است این نیرو در تولید نمایه های موضوعی و با استفاده از فنون دقیق و کامل نمایه سازی بکار برده شود. به علاوه بسیاری از این اصطلاحات عملاً سادگی و مفید بودن رهیافت اولیه را مغلوب می کنند و نتیجه آن چیزی جز یک نمایه سازی ناکافی، غیر خودکار و غیر ارزان نخواهد بود. "لوهن" معتقد بود که نمایه گردان فقط وسیله ای برای آگاهی رسانی جاری است که دارای ارزش موقت می باشد، وسیله ای که می تواند به سازمانهای جوان امکان دهد تا زمان تهیه نمایه های جامع موضوعی از آن برای راهنمایی و روزآمد نگهداشتن اطلاعات مراجعان خود استفاده کنند. او اعتقاد داشت که این نمایه ها بایستی با ابراز فکری قدرتمندی جایگزین گردند (حریری و توفیق، ۱۳۶۲، ۱۷-۳۰).

۲-۱-۲- نمایه سازی به زبان آزاد :

زبان نمایه سازی آزاد نوع دیگری از زبان های نمایه سازی است که تفاوت آن با شیوه قبلی عدم محدودیت به مدرک در دست نمایه سازی است. نمایه سازی به زبان آزاد می تواند توسط انسان یا رایانه انجام گیرد. نمایه سازی با زبان آزاد در رایانه همانند نمایه سازی بر اساس زبان طبیعی است، زیرا رایانه باید چیزی برای گزینش در اختیار داشته باشد و این چیز همان واژه های مندرج در عنوان، چکیده یا متن مدرک است.

هم زبان نمایه‌سازی طبیعی و هم زبان نمایه‌سازی آزاد به مقدار زیاد در تولید نمایه‌ها و دستیابی به پایگاه‌های اطلاعاتی به کار برده می‌شوند. بسیاری از پایگاه‌های اطلاعاتی واژه‌هایی از متن مدرک و واژه‌هایی از زبان نمایه‌سازی کنترل شده را برای نمایه‌سازی به کار می‌برند و نمایه‌ای را که می‌تواند به عنوان گزینش بهترین‌ها از دو حوزه تلقی شود ارائه می‌دهند. در عین حال استفاده از دو شیوه یاد شده در نمایه‌سازی دارای مسائلی است که توجه به زبان نمایه‌سازی کنترل شده را آشکار می‌سازد. برخی از این مسائل عبارتند از:

۱. مترادف‌ها: مترادف‌ها کلمات و اصطلاحاتی هستند که دیکته متفاوت و معانی یکسان و مشابه دارند.

وضعیت‌های مشابه و مشترکی که در مورد مترادف‌ها به وجود می‌آید به شرح زیر است:

الف. بعضی موضوعات هم ریشه‌اند، مثلاً واژه‌های گند زدا، گند زدایی، گند زدا شده یا واژه‌های محاسبه، حساب شده، حسابگر، محاسبه‌گر. گاه مناسب است که این واژه‌ها برابر تلقی شوند و در موقعیت دیگر مهم است که متفاوت دانسته شوند.

ب. بعضی موضوعات هم دارای معنای عام و هم دارای معنای خاص (فنی) هستند. این تفاوت‌ها باید

شناخته شده و بر اساس نوع مخاطب، در نمایه انعکاس پیدا کند. مانند: نمک طعام و کلرید سدیم.

ج. تغییراتی که در طول زمان در استفاده از اصطلاحات رخ می‌دهد نیز مشکلاتی ایجاد می‌کنند.

برای مثال کتابخانه کنگره برای رایانه ابتدا اصطلاح "ماشین حسابگر الکترونیک" را به کار می‌برد اما بعدها مجبور شد آن را به اصطلاح رایج آن یعنی "کامپیوتر" تغییر دهد.

د. بعضی مفاهیم در ویرایش‌های مختلف یک زبان به شکل‌های متفاوت تعریف می‌شوند. مثلاً

اصطلاحات زیر همگی ممکن است معنای واحدی داشته باشند Eyeglasses; Spectacles; Glasses همه به معنی عینک هستند.

۲. هم‌نویسه‌ها (کلمات هم‌املاء): کلماتی که املائی یکسان ولی معنای متفاوت دارند باید شناسایی شوند.

در یک نمایه عمومی برای تشخیص معانی متفاوت یک کلمه متشابه تنها می‌توان از دامنه نما

(یادداشت دامنه ، توضیحگر) و یا از کلمات توصیفی کمک گرفت. مانند: روابط (ریاضی)؛ روابط (اجتماعی).

۳. جمع ها و مفرد ها: تمام اسم ها یک شکل جمع و یک شکل مفرد دارند . معمولاً شکل جمع و مفرد یک کلمه معادل یکدیگر تلقی می شود ، اما گاه نمونه هایی وجود دارد که ضروری است بین شکل جمع و مفرد یک واژه تفاوت قائل شویم. مثلاً *exercice* به معنای تمرین و ورزش است و در حالت جمع به معنی جشن می باشد.

۴. واژه های مرکب: بعضی موضوعات را نمی توان با یک کلمه بیان کرد و برای شرح مشخص آنها به دو یا چند کلمه نیاز است. مانند : بازیابی اطلاعات، ماورای عالم خاکی. دسترس پذیری باید از طریق همه کلمات مهم و معنی دار در اصطلاحات مرکب فراهم شود. بدین منظور گاهی از شکل مقلوب واژه های مرکب استفاده می شود.

۵. موضوعات پیچیده: واژه هایی که بیش از یک مفهوم را شامل می شوند. در اصطلاحنامه ها و سر عنوان های موضوعی از علائم و نشانه های خاصی برای نمایش مفاهیم این واژه ها استفاده می شود. در مورد این موضوعات ترتیب اتصال^۱ بسیار مهم است. برای مثال، بدیهی است که "کتابشناسی تاریخی" مترادف "تاریخ کتابشناسی" نیست. هر دو واژه یعنی "کتاب شناسی" و "تاریخ" در این دو موضوع آورده شده اند و فقط ترتیب قرار گرفتن آنها در زبان طبیعی و نیز حروف ربط بین آنها تعیین کننده موقعیت و مفهوم دو موضوع است.

بنابراین در نظام زبان طبیعی واژه های مورد استفاده در نمایه از عنوان ها، چکیده ها، استناد ها یا تمام متن گرفته شده و کلمات به همان شکلی که هستند آورده می شوند و هیچ تغییری در آنها داده نمی شود، خطاهای نمایه سازی در حداقل میزان خود است و واژه های جدید را به هر شکلی که مورد نیاز است می توان به واژه نامه اضافه کرد. با این حال مسائل یاد شده عمدتاً مربوط به زبان می شوند و لازم است بر

¹ Citation Order

نقایص آنها (زبان طبیعی) که به عنوان اساس کلید واژه های جستجو در یک نظام بازیابی اطلاعات به کار می روند غلبه کنیم. در یک نظام ایده ال و مناسب، جوینده قادر خواهد بود در قبال یک پرسش به زبان طبیعی با استفاده از رایانه متون را جستجو کرده و مجموعه ای از مدارک کافی و مرتبط با موضوع مورد نظر خود را بازیابی کند. تحقیقات و بعضی از نظام های عملیاتی مثل LIBERTAS و STAU.SI.IQ به این سمت حرکت می کنند، اما اغلب نظام ها هنوز دومیورد زیر را پیشنهاد می کنند.

۱. استفاده از نمایه سازی به زبان طبیعی و اجازه دادن به جوینده تا به آن زبان فکر کند.

۲. استفاده از زبان کنترل شده و تقاضا از جوینده تا خودش را با یک واژه نامه مصنوعی تحمیل شده

تطبیق دهد (تیلور، ۱۳۸۱، ۱۸۶-۱۸۹).

۲-۱-۳- نمایه سازی به زبان کنترل شده :

زبان نمایه سازی کنترل شده سومین شیوه نمایه سازی است. در این شیوه نمایه ساز هم بر اصطلاحاتی که برای ارائه موضوعات به کار می رود و هم بر فرایندی که اصطلاحات به مدرکی خاص نسبت داده می شود نظارت می کند. در این گونه نمایه سازی معمولاً فهرستی مستند مشخص کننده اصطلاحاتی است که ممکن است به موضوعات نسبت داده شود. به بیانی دیگر، در نمایه سازی کنترل شده شخصی اصطلاحات یا واژه هایی از یک فهرست واژگان را بر اساس تفسیر ذهنی که از مفاهیم مندرج در مدرک دارد به مدرک اختصاص می دهد. به دلیل تنوع زیاد در واژگان زبان طبیعی، اصطلاحات مورد استفاده در یک نمایه با دقت کنترل می شود. چنین کنترلی ممکن است مانع استفاده از برخی کلمات به عنوان موضوعات یا مدخل های دسترسی شوند. اصطلاحاتی که مورد استفاده قرار خواهند گرفت، احتمالاً خاص می باشند، مترادف ها شناخته شده و احتمالاً حذف می شوند و شکل مرجع کلمات مورد توجه خواهد بود. ساده ترین راه برای به کار بردن چنین کنترلی بر واژه های یک نمایه این است که واژگان قابل قبولی به عنوان یک مجموعه لغات ویژه دسته بندی و ذخیره شوند. چنین سیاهه ای هم در بردارنده تصمیم گیری خاص مربوط به کلمات مرجع است و هم تصمیم گیری درباره شکل کلمات،

مثلاً مفرد یا جمع، اسم یا صفت را مشخص می‌سازد. سه ابزار برای زبان نمایه‌سازی کنترل شده طراحی شده است:

۱- **سرعنوان‌های موضوعی:** سرعنوان‌های موضوعی سیاهه‌ای از اصطلاحات نمایه‌ای است که معمولاً با نظم الفبایی مرتب شده است و شکل مستندی برای توصیف موضوعات در یک نمایه، فهرست یا پایگاه داده‌ها فراهم می‌کند.

۲- **اصطلاحنامه:** اصطلاحنامه مجموعه‌ای ساخت یافته از کلمات و واژه‌های برگزیده به شمار می‌آید که میان آنها روابط تعریف شده‌ای حاکم است.

۳- **لغت هستی‌شناسی:** در رشته فلسفه اصطلاح لغت هستی‌شناسی سابقه طولانی مدت و قابل تاملی دارد که به معنای ارائه تفسیری سیستماتیک از هستی است. در برهه‌ای از زمان در دهه اخیر این کلمه توسط جامعه اطلاع‌رسانان برای بیان عناصر زیربنایی به کار گرفته شد تا مشابهت دانش انسانها و رایانه‌ها را میسر سازد. وسیع‌ترین کاربرد لغت هستی‌شناسی در هوش مصنوعی برای بازنمون‌های رسمی آن چیزی است که برای ما انسانها به منزله عقل سلیم تعریف می‌شود. در این مفهوم مقولات شامل چیزهایی مانند فضا، زمان و ساختار اشیاء فیزیکی است. چنان بازنمونهای رسمی در خلق عوامل هوشمند که بتواند وظایف خاصی مثل جارو کردن کف اتاق را انجام دهند مفید است. یک چنین هستی‌شناسی یک هستی‌شناسی غیرزبانی است. تلاشهای زیادی برای فرموله کردن مفاهیم انتزاعی صورت گرفته است. یک لغت هستی‌شناسی در فرایند زبان طبیعی باید واقعیت کاربرد زبان برای برقراری ارتباط را فرموله کند. به این گونه هستی‌شناسی‌ها، هستی‌شناسی زبانی می‌گوییم و می‌تواند واقعیت‌هایی چون دستور زبان، معناشناسی، صرف و نحو و ... را در بر گیرد (تیلور، ۱۳۸۱، ۱۹۸-۲۱۲).

ویژگی زبان‌های نمایه‌سازی:

زبان نمایه‌سازی، کنترل شده و زبان نمایه‌سازی طبیعی هر یک دارای ویژگی‌های خاصی است که در ذیل مقایسه ساده‌ای از این دو به دست داده می‌شود.

ویژگی‌های زبان طبیعی:

- ۱- رایانه می‌تواند نمایه کند.
- ۲- مدارک تحت تعداد بیشتری کلید واژه پراکنده می‌شوند.
- ۳- کلیه واژه‌ها از زبان طبیعی مدارکی که نمایه می‌شوند استخراج می‌شوند.
- ۴- رایانه به گونه‌ای با ثبات کلیدواژه‌ها را تعیین می‌کند، بنابر این اگر کلیدواژه‌ای جزو موضوعات مدرک نمایه‌سازی شده باشد به یقین مدرک با آن کلیدواژه بازیابی می‌شود.
- ۵- روابط بین واژه‌های نمایه نامشخص است مگر این که کنکاش ذهنی برای برقراری روابط به کار گرفته شود.
- ۶- زبان نمایه‌سازی می‌تواند با سن مدرک و تحول موضوعی مدارک همخوان باشد. بنابراین زبان پویاست و محتمل است که منعکس کننده کاربردهای جاری باشد.
- ۷- غالباً در پایگاه‌های تمام متن به کار می‌رود.
- ۸- کنکاش ذهنی برای بازیابی اساسی است.
- ۹- هر مدرک احتمالاً با تعداد زیادی کلیدواژه قابل بازیابی است.
- ۱۰- تعداد نسبتاً زیادی مدارک غیر مرتبط بازیابی می‌شود.
- ۱۱- لازم است که تمام مترادفها و تنوع کلمات در جستجو دخالت داده شود تا جستجوی کامل و جامع انجام گیرد.
- ۱۲- به دلیل معانی متفاوت کلمات ریزش کاذب وجود دارد.
- ۱۳- کنترل بر پیوند بین کلمات وجود ندارد مثلاً تمایزی بین "رایانه برای طراحی" و "طراحی رایانه" وجود ندارد.

ویژگی‌های زبان کنترل شده:

- ۱- نمایه‌ساز نمایه‌سازی می‌کند.

- ۲- مدارک با تعداد محدودتری توصیفگر گروه‌بندی می‌شوند.
- ۳- توصیفگرهایی که مدارک با آنها گروه‌بندی می‌شوند قابل تعریف و پیش‌بینی هستند.
- ۴- ممکن است مدارک همیشه با توصیفگرهایی که کاربر متوقع است نمایه نشوند، امکان گواهی کاربر به دلیل اشتباه انسانی وجود دارد.
- ۵- روابط بین اصطلاحات مشخص و قابل استخراج است.
- ۶- زبان نمایه‌سازی تنها با تغییر اصطلاحنامه یا فهرست مستند موضوعی میسر است، این فرایند ممکن است کند باشد و موجب در اختیار داشتن زبان نمایه‌سازی ایستا شود.
- ۷- زبانی رایج در پایگاه‌های اطلاعاتی و فهرست‌های سنتی است.
- ۸- کنکاش ذهنی برای نمایه‌سازی ضروری است.
- ۹- هر مدرک احتمالاً با تعداد محدودی اصطلاح موضوعی قابل بازیابی است.
- ۱۰- بازیابی از دقت بالا برخوردار است.
- ۱۱- لازم نیست که تمام مترادفات در جستجو دخالت داده شود.
- ۱۲- ریزش کاذب در سطح حداقل است (تیلور، ۱۳۸۱، ۱۷۸-۲۱۰).

۲-۲- پیشینه تحقیق:

همان گونه که در صفحات قبل بدان اشاره شد رشد سریع دانش، انتشارات و منابع غیرکتابی پس از جنگ جهانی دوم دست اندر کاران اطلاع‌رسانی را با نیروی "انفجار اطلاعات" مواجه ساخت و ضرورت بهره‌گیری از روش‌هایی را به منظور کنترل اطلاعات، ذخیره و بازیابی آن ایجاد کرد. برخی از اندیشمندان اطلاع‌رسانی چون "مورتیمر تاوب" (Mortimer Taube) واژه‌های به کار رفته توسط پدید آورندگان مدرک را مورد توجه قرار داده و از همین واژه‌ها برای ذخیره و بازیابی مدرک استفاده کردند. برخی نیز چون "مورز" (Calvin Mooers) به دلیل مسائل مربوط به بازیابی اطلاعات از طریق کلید واژه‌های پدید آورندگان (مانند ریزش کاذب، همارای نادرست، وجود مترادف‌ها و ...) به روش‌های کنترل شده در ذخیره و بازیابی اطلاعات روی آوردند.

کتابخانه ها و مراکز اسناد مختلف نیز هر یک بنا به شرایط و امکانات خود یکی از این روش ها را در ذخیره و بازیابی اطلاعات مورد استفاده قرار دادند، علاوه بر این متخصصان اطلاع رسانی به منظور اتخاذ تصمیم های مناسب در امر ذخیره و بازیابی اطلاعات، روش های یاد شده را از دیدگاه های مختلفی چون هزینه و سودمندی، جامعیت و مانعیت بازیافت و میزان همپوشانی این روش ها مورد بررسی و ارزیابی قرار دادند که در ادامه به برخی از آنها اشاره می شود.

۲-۲-۱- مطالعات انجام شده در خارج از ایران :

باتل و پرایبیش (Bottle and Preibish, 1970) در پژوهشی به بررسی سودمندی نمایه سازی کوئیک در مورد مقالات روانشناسی پرداخته و دریافتند در "چکیده نامه روانشناسی"^۱ ۶۰ درصد مقالات با استفاده از کلیدواژه های عنوان قابل بازیابی است. همین طور نتایج حاصل از پژوهش انانسون (Anonson) در سال ۱۹۸۷ _ که به ارزیابی جستجوی کلید واژه ای پرداخت _ نشان داد نه تنها جستجوی کلید واژه ای عنوان باعث بازیابی مدارک مرتبط می شود بلکه از طریق چنین جستجویی رکوردهایی بازیابی می شوند که با جستجوی موضوعی قابل بازیابی نیستند.

شفرد (Shepherd, 1987) در تحقیقی به بررسی میزان همخوانی کلیدواژه های عنوان و توصیفگرهای موضوعی در پایگاه اطلاعاتی (OON)^۲ پرداخت. در این پایگاه که وابسته به موسسه اطلاعات علمی و فنی کانادا (CISTI)^۳ می باشد، گزارش کنفرانسها، تک نگاشتها و گزارش های فنی در حوزه علوم پزشکی و تکنولوژی نگهداری می شود. تمام رکوردهای این پایگاه دارای فیلد عنوان و حداقل یک فیلد توصیفگر موضوعی است. گفتنی است این پایگاه فاقد چکیده بوده و بازیابی در این پایگاه به دو صورت امکان پذیر است: استفاده از واژه های عنوان و استفاده از توصیفگرهای موضوعی. هدف اصلی این پژوهش اندازه گیری میزان همپوشانی

1 - Psychological Abstracts

1 - Object Oriented Networking

2 - Canada Institute for Scientific and Technical Information

موضوعی مدارک بازیابی شده در این دو روش بود. نتایج این بررسی نشان داد که استفاده کنندگان این پایگاه برای دسترسی به بیشترین اطلاعات بایستی از هر دو روش بازیابی استفاده نمایند.

فراست (Frost, 1989) در پژوهشی به بررسی همپوشانی بین سرعنوانهای برگزیده از (LCSH)^۱ و واژه‌های استخراج شده از عنوان در فهرست کتابخانه دانشگاه میشیگان پرداخت. او قصد داشت بررسی کند که چه درجه از همپوشانی بین واژگان کنترل شده موضوعی و واژه‌های عنوان وجود دارد. برخی از سوالات مطرح شده در این پژوهش عبارت بودند از:

۱- چه میزان از واژه‌های عنوان دقیقاً با سر عنوان موضوعی همپوشانی دارند؟

۲- چه میزان از واژه‌های عنوان با سرعنوان موضوعی همراه با تقسیم فرعی آن همپوشانی دارد؟

۳- چه میزان از واژه‌های عنوان حداقل با بخشی از سرعنوان موضوعی اختصاص داده شده همپوشانی دارند؟

۴- چه میزان از واژه‌های عنوان حداقل با بخشی از واژه‌های تشکیل دهنده تقسیمات فرعی همپوشانی دارند؟

یافته‌های این پژوهش حاکی از آن بود که ۷۳ درصد، واژه‌های عنوان دقیقاً یا حداقل با بخشی از سر عنوان یا تقسیمات فرعی همپوشانی دارند. در ۱۱ درصد موارد نیز واژه‌های عنوان، دقیقاً با سر عنوان برابر است و در بیش از ۵۳ درصد موارد واژه‌های عنوان حداقل با یک واژه از سر عنوان همپوشانی دارند. پیترز و کورس (Peters and Kurth, 1991) پژوهشی با عنوان "جستجوی موضوعی با واژگان کنترل شده و کنترل نشده در فهرست پیوسته کتابخانه دانشگاه میسوری کانزاس" انجام دادند که برخی از فرضیه‌های این پژوهش عبارت بودند از:

۱- در هر جلسه جستجو از کلید واژه‌های عنوان، استفاده می‌شود.

۲- در جستجوی موضوعی، بیشتر جستجوها با کمک واژگان کنترل شده آغاز می‌شود.

۳- تفاوت معنی‌داری میان ضرایب بازیافت در این دو روش وجود ندارد.

یافته‌های پژوهشی فرضیه شماره ۱ را تأیید نمود و مشخص شد بیشتر جستجوها حداقل یک جستجوی کلیدواژه‌های عنوان داشتند که به عنوان دسترسی موضوعی از طریق واژگان کنترل نشده به کار رفته است. فرضیه ۲ در این بررسی تأیید نشد و نتایج نشان داد که تقریباً ۵۹ درصد جستجوها با واژگان کنترل نشده آغاز شده است. فرضیه ۳ نیز تأیید نشد و معلوم گشت حاصل بازیابی از طریق واژگان کنترل نشده دارای ضریب بازیافت بالاتری نسبت به واژگان کنترل شده می‌باشد. نتایج این مطالعه نشان داد که کاربران از کلیدواژه‌های عنوان برای دسترسی موضوعی استفاده می‌کنند و دو روش را در طول یک جلسه جستجو ترکیب می‌کنند.

کالر (Keller, 1992) پژوهشی بر روی پایان‌نامه‌های مقطع کارشناسی ارشد کتابخانه دانشگاه ایالتی ایندیانا انجام داد و به مقایسه نخستین سرعنوان موضوعی کنگره- که به پایان‌نامه اختصاص داده شده بود - با واژه‌های عنوان پرداخت تا مشخص کند احتمال وجود کلید واژه‌های عنوان در سرعنوان‌های موضوعی چقدر است. فرضیه پژوهش این بود که هر چه قدر تعداد واژه‌های تشکیل دهنده عنوان بیشتر باشد احتمال بیشتری دارد که اولین سرعنوان موضوعی در عنوان پایان‌نامه وجود داشته باشد. یعنی افزایش تعداد واژه‌های عنوان احتمال همپوشانی کلیدواژه عنوان با سرعنوان موضوعی اختصاص داده شده به پایان‌نامه را افزایش می‌دهد. جامعه این پژوهش شامل پایان‌نامه‌های کارشناسی ارشد بین سالهای ۹۰-۱۹۷۸ بود. پایان‌نامه‌های مربوط به سالهای قبل از ۱۹۷۸ به علت این که عبارات دچار تغییرات زیادی در طی این سالها شده بود، در نظر گرفته نشد. هم چنین پایان‌نامه‌های رشته ادبیات از جامعه پژوهش حذف شد که به این ترتیب جامعه پژوهش شامل ۳۷۱ پایان‌نامه بود. نخستین سرعنوان موضوعی به این جهت انتخاب شد که اصلی‌ترین موضوع پایان‌نامه را در بردارد. نتایج حاصل از این بررسی نشان داد که از ۳۷۱ پایان‌نامه در ۲۳۸ عنوان (۶۴ درصد) بین واژه‌های عنوان و سرعنوان موضوعی همپوشانی وجود دارد. یعنی در پایان‌نامه‌هایی که دارای ۶-۱ واژه در عنوان هستند ۱۳ درصد و پایان‌نامه‌های دارای ۱۲-۷ واژه در عنوان ۴۷ درصد و ۱۸-۱۳ واژه در عنوان ۳۰ درصد و ۲۴-۱۹ واژه

در عنوان ۱۰ درصد همپوشانی میان واژه‌های عنوان و نخستین سرعنوان موضوعی وجود دارد که بدین ترتیب فرضیه پژوهش رد شد.

اسپینک و ساراسویک (Spink and Saracevic, 1993) در پژوهشی قابلیت کلید واژه‌ها را در بازیابی اطلاعات مرتبط مورد بررسی قرار دادند. جامعه مورد بررسی در این پژوهش را، جستجوهای انجام شده در پایگاه دیالوگ (DIALOG) تشکیل می‌داد. ۴۰ استفاده کننده و ۴ کاوشگر حرفه‌ای در این پژوهش شرکت داشتند. همچنین پنج راهی که کلید واژه‌های جستجو از این راهها بدست می آمد عبارت بودند از :

۱- واژه‌های نوشته شده توسط مراجعه کنندگان در درخواستهای جستجو؛

۲- واژه‌هایی که طی هر جستجو از طریق مصاحبه مرجع با استفاده کنندگان به دست می‌آمد؛

۳- واژه‌هایی که از بازیابی مدارک مرتبط دیگری به دست می‌آمد؛

۴- اصطلاحنامه موجود در پایگاه؛

۵- واژه‌های استخراج شده توسط کاوشگران حرفه‌ای، طی هر مرحله جستجو.

نتایج پژوهش نشان داد مورد شماره یک بیشترین کاربرد و سودمندی را در بازیافت داشته و این واژه‌ها نتایج بهتری در بازیابی داشته‌اند. ۴۰ درصد واژه‌هایی که جهت جستجو استفاده شد از نوع اول بودند و در تمام ۴۰ مورد جستجو به کار برده شدند، یعنی در بازیابی موضوعات مرتبط نقش مثبتی داشته‌اند. همچنین ۶۷ درصد از واژه‌هایی که به شیوه دوم استخراج شده بودند باعث بازیافت مدارک مرتبط شده و در واقع این واژه‌ها که ۱۰ درصد از کل واژه‌های استخراج شده به ۵ روش را تشکیل داده بودند گرچه خیلی مفید نبودند ولی هنگامی که استفاده می‌شدند، در به دست آوردن نتایج بهتر موثر بودند.

هنزلر (Henzler, 1978) در پژوهشی کلیدواژه‌های عناوین برخی از رکورد های موجود در پایگاه اطلاعاتی Cancer net را با نمونه هایی از توصیفگرهای موجود در اصطلاحنامه Cancer net مورد مطالعه و بررسی قرار داد و به این نتیجه رسید که ۳۵ درصد از کلید واژه های عناوین فاقد توصیفگرهای مناسبی در اصطلاحنامه هستند. از این بین ۵ درصد اصطلاحات بی معنی، ۲۵ درصد مرتبط به موضوعات خاص و ۵ درصد به اسامی

خاص و شماره ها اختصاص دارند. وی در پایان نتیجه گیری می کند که استفاده هم زمان از کلید واژه ها و توصیفگر ها منجر به بازیابی مطلوب اطلاعات می شود.

لنکستر (Lancaster, 1991) و همکارانش نیز از طریق "فهرست های همگانی پیوسته" اطلاعات مربوط به ۵۱ موضوع را استخراج کردند. آنها از بین عناوین بازیابی شده، ۶۰۷ مدرک را که در کتابخانه دانشگاه ایلینویز موجود بود به عنوان "سیاهه واری" در نظر گرفتند، سپس از جستجوگران با تجربه درخواست کردند که از طریق "فهرست همگانی پیوسته" و تنها با استفاده از سر عنوان های موضوعی کتابخانه کنگره رکورد های موجود در رابطه با ۵۱ موضوع یاد شده را استخراج کنند. بدین طریق تنها ۵۳ درصد از ۶۰۷ مدرک موجود در دانشگاه ایلینویز بازیابی شد.

حمیدی (Hmeidi, 1997) در پژوهشی پیرامون نمایه سازی ماشینی مدارک به زبان عربی، نتایج بازیابی از طریق کلید واژه ها و ریشه واژه ها در زبان عربی را مورد بررسی و ارزیابی قرار داد. وی دریافت که در زبان عربی استفاده از ریشه لغات از نظر میزان بازیافت و دقت نتایج بهتری را در مقایسه با بازیابی از طریق کلید واژه ها به دست می دهد.

ووربیج (Voorbij, 1998) نیز بر روی تک نگاشت های موجود در کتابخانه ملی هلند، در حوزه علوم انسانی و اجتماعی که از طریق فهرست همگانی پیوسته قابل بازیابی بود، دو تحقیق انجام داد. وی در تحقیق نخست از کتابداران متخصص خواست تا در مورد ۴۷۵ رکورد، میزان همخوانی توصیفگرها و کلیدواژه ها را مقایسه کنند. برای نشان دادن میزان همخوانی این دو، هفت گزینه در اختیار کتابداران قرار گرفت که عبارت بودند از:

۱. توصیفگر دقیقاً مشابه کلید واژه است؛

۲. توصیفگر مترادف با کلید واژه عنوان است؛

۳. توصیفگر عام تر از کلید واژه است؛

۴. توصیفگر خاص تر از کلید واژه است؛

۵. توصیفگر با کلید واژه مرتبط است؛

۶. توصیفگر به شکلی با کلید واژه همخوان است ولی اختصاص گزینه های ۲،۳،۴ یا ۵ به آن مشکل است؛

۷. توصیفگر هیچ همخوانی با کلید واژه ندارد.

نتایج این بررسی نشان داد از نظر میزان همخوانی کلید واژه های عنوان و توصیفگرها ۶۰ درصد در گزینه های ۳،۱،۲ قرار گرفته و به ۴۰ درصد دیگر نیز گزینه های ۷،۴،۵،۶ اختصاص داده شده است. در بررسی دوم از کتابداران خواسته شد که با استفاده از کلید واژه های عنوان و توصیفگرهای این رکوردها عملیات بازیابی را انجام دهند، که طی آن مشخص شد ۸۶ درصد منابع با استفاده از توصیفگرها و ۴۸ درصد با استفاده از کلید واژه های عنوان بازیابی شدند. وربیج علت بازیافت کم از طریق کلید واژه در مقایسه با توصیفگرها را محدود بودن دامنه توصیفگرها، تنوع عناوین و گستره موضوعی دانست که مطالب علوم انسانی و اجتماعی در قالب آن عرضه می شود.

۲-۲-۲- مطالعات انجام شده در داخل کشور:

معرف زاده (۱۳۶۶) در پژوهشی با عنوان "بررسی وضعیت ابزار بازیابی اطلاعات پزشکی فارسی و پیشنهادی برای تهیه نمایه پزشکی ایران" به بررسی این پرسش می پردازد که کدام شیوه نمایه سازی گردان برای ذخیره و بازیابی اطلاعات پزشکی بیشتر می تواند مفید واقع شود، تا از آن برای تهیه نمایه پزشکی ایران بتوان بهره گرفت. وی سه روش نمایه سازی گردان شامل: ۱- درون بافتی (کوئیک)^۱، ۲- برون بافتی (کووک)^۲، ۳- افزوده به عنوان (کواک)^۳ را مورد بررسی قرار داده و به این نتیجه رسید که کلید واژه های استخراج شده از عنوان مقالات، کتب و پایان نامه های مورد بررسی ۵۸ درصد از کل واژه های عناوین این مدارک است که این نسبت در پایان نامه های پزشکی به ۶۹ درصد افزایش یافته است. پژوهشگر در پایان روش "نوع افزوده به عنوان"^۴ را برای تهیه نمایه پزشکی ایران پیشنهاد می کند.

1- Key Word In Context (KWIC)

2- Key Word Out of Context (KWOK)

3- Key Word Augmented of Context (KWAK)

۴- در این نوع نمایه گردان علاوه بر کلید واژه های عنوان، از کلید واژه های عنوان های فرعی، چکیده یا متن مدرک نیز استفاده می شود.

بزرگی (۱۳۷۷) در پژوهشی با عنوان "بررسی میزان سازگاری عنوان ها و موضوع های کتاب های فارسی منتشره در ایران در فاصله سالهای ۱۳۶۵-۱۳۷۵" دو فرضیه را مورد آزمون قرار داد :

۱- میان عنوان و موضوع کتابهای فارسی سازگاری وجود دارد.

۲- میزان سازگاری عنوان و موضوع کتابها در همه رده های موضوعی یکسان نیست.

جامعه این پژوهش شامل کتابهای فارسی بود که در فاصله سالهای ۱۳۶۵ تا ۱۳۷۵ در ایران منتشر و در کتابشناسی ملی ایران درج شده بود. وی با استفاده از فهرست برگه های کتابخانه ملی عناوین کتاب ها (اعم از عنوان اصلی یا هر نوع اطلاع و توضیحی که در مورد آن باشد) و موضوعات اختصاص یافته به کتابها را مورد بررسی قرار داد. یافته های این پژوهش نشان داد از مجموع موضوعاتی که به کتابها اختصاص یافته است ۳۶ درصد آنها با عنوان سازگاری کامل دارند. ۳ درصد موضوعها از طریق ارجاع « نیز نگاه کنید» به عنوان کتابها مربوط شده و ۲۰ درصد موضوعها به وسیله ارجاع « نگاه کنید» به عنوان ها مرتبط شده اند. علاوه بر این مشخص شد که میان موضوعات و عناوین کتابهای فارسی بطور میانگین ۷۳ درصد سازگاری وجود دارد. بالاترین میزان سازگاری یعنی ۸۷ درصد به رده کلیات تعلق داشته و پس از آن رده علوم کاربردی (تکنولوژی) با برخورداری از ۸۳ درصد سازگاری در مقاوم دوم قرار دارد. کمترین میزان سازگاری در رده ادبیات (۴۱ درصد) مشاهده شد.

داور پناه (۱۳۷۵) نیز در پژوهشی میزان همخوانی عناوین مقالات فارسی با محتوای آن ها را در گروه های مختلف علمی بررسی نموده و به این نتیجه رسید که عناوین مقالات در حوزه علوم انسانی، در مقایسه با عناوین مقالات علوم پایه و مهندسی و مقالات علوم پزشکی از سازگاری کمتری با محتوا برخوردار هستند.

انصاری (۱۳۷۹) در پایان نامه خود به "بررسی انطباق میان توصیفگرهای نمایه سازی و کلید واژه های عنوان پایان نامه های دکترای تخصصی کودکان، زنان، قلب و عروق و روانپزشکی" پرداخت و نشان داد که در حوزه مورد بررسی به میزان ۷۰/۳ درصد میان توصیفگرها و کلید واژه های عنوان انطباق وجود دارد. علاوه بر این

پایان نامه های رشته روانپزشکی از نظر انواع انطباق در وضعیت متفاوتی از سایر پایان نامه ها قرار دارند و افزایش تعداد کلید واژه های عنوان تأثیری بر میزان انواع انطباق ندارد.

همان گونه که مشاهده می شود هر یک از این پژوهش ها به شکلی بر نقش و اهمیت "عنوان" تاکید دارند. در برخی از این پژوهش ها نشان داده شده که میزان استفاده از کلید واژه های عنوان توسط جستجوگران بیش از توصیفگرها است علاوه بر این درصد بالایی از توصیفگرها با کلید واژه های عنوان همخوان هستند. در تحقیق حاضر نیز مانند پژوهش های دیگر میزان همخوانی یا همپوشانی کلیدواژه های "عنوان" پایان نامه ها با توصیفگرهایی که به منظور ساماندهی این منابع اختصاص داده می شود ارزیابی شده تا میزان استفاده از عنوان در فعالیت های نمایه سازی سنجیده شود. در صورتی که این همخوانی زیاد باشد می توان از کلید واژه های عنوان، که زبان طبیعی یا به عبارتی زبان استفاده کنندگان است در نمایه سازی استفاده کرد و با بهره گیری از شیوه های نمایه سازی ماشینی در وقت و هزینه نمایه سازی دستی، ویراستاری، تصحیح و به طور کلی ذخیره و بازیابی اطلاعات صرفه جویی نمود. در این تحقیق موضوع مورد بررسی محدود نشده و کلیه حوزه هایی که در قالب آنها پایان نامه تهیه می شود مورد بررسی قرار گرفته است تا ویژگی های "عنوان" پایان نامه ها در حوزه های مختلف همچنین شباهت ها و تفاوت های حوزه ها با یکدیگر مشخص شود. علاوه بر این میزان بازیافت از طریق کلید واژه های عنوان و توصیفگرهای نمایه سازان نیز سنجیده شده تا بتوان فرایندهای مناسب تری در ذخیره و بازیابی اطلاعات مربوط به پایان نامه ها اتخاذ نمود.

فصل سوم

روش پژوهش و گردآوری اطلاعات

۳-۱- روش پژوهش:

این تحقیق با روش تحلیل محتوا انجام شده است؛ زیرا طی آن به گونه ای کمی استفاده از کلیدواژه ها در عنوان پایان نامه ها و میزان همخوانی آن ها با توصیفگرها تحلیل و ارزیابی شده است. تحلیل محتوا به زعم کاپلان (Kaplan, 1964, 21) یک فن پژوهشی است که با استفاده از داده ها، استنتاج های معتبر و تکرار پذیری در مورد زمینه وفاد آن به عمل می آید. وبر (Weber, 1990, 9) نیز این فن را یک روش پژوهش می داند که از طریق برخی شیوه ها استنتاج های معتبری از متن به عمل می آورد، این استنتاج ها در مورد فرستنده پیام، خود پیام و مخاطبان پیام است.

۳-۲- جامعه آماری :

در این تحقیق جامعه آماری، ۶۵۵۳۶ عنوان پایان نامه ای است که در پایگاه "چکیده پایان نامه های ایران" تا پایان سال ۱۳۸۱ معرفی شده اند.

۳-۳- روش نمونه گیری :

در اولین مرحله از تحقیق اطلاعات مربوط به عنوان پایان نامه ها، توصیفگرهای اختصاص داده شده به هر عنوان، تاریخ نگارش، حوزه اصلی و رشته فرعی پایان نامه های موجود در "پایگاه چکیده پایان نامه های ایران" از ابتدا تا پایان سال ۱۳۸۱، از محیط نرم افزاری CDS-ISIS به محیط نرم افزاری EXCEL انتقال یافت و هر یک از فیلدهای ذکر شده در ستون جداگانه ای قرار گرفت. بدین ترتیب ۶۵۵۳۶ رکورد به عنوان اولین جامعه تحقیق مورد بررسی قرار گرفت. در ادامه به منظور به دست آوردن جامعه یکدستی برای تحقیق، قبل از تفکیک پایان نامه ها به رشته های اصلی عناوینی که دارای نقایص آشکاری بودند از جامعه حذف شدند. انجام این کار عمدتاً از طریق مرتب (Sort) کردن اطلاعات هر ستون صورت گرفت. به عنوان مثال با مرتب کردن ستون "توصیفگرها" رکورد های فاقد توصیفگر به راحتی قابل تشخیص بود. موارد حذف شده در جدول شماره ۱ آمده است:

جدول شماره ۱- تعداد رکورد های حذف شده قبل از تفکیک حوزه

تعداد	موارد حذف شده
۴۳۳۱	بدون تاریخ تهیه
۱۵۹	بدون رشته اصلی و فرعی
۷۷۳	بدون توصیفگر
۳۳	بدون عنوان
۵۲۹۶	جمع

در ادامه جامعه آماری به رشته های اصلی تقسیم شده و عناوینی که فاقد رشته فرعی یا درج رشته فرعی اشتباه بودند (مثلاً رشته نساجی در حوزه علوم انسانی) از جامعه تحقیق حذف شدند. همچنین مشاهده شد که عناوین برخی پایان نامه ها بدون هیچ گونه تفاوتی دو یا چند بار تکرار شده است. در این خصوص مرتب کردن ستون "عنوان" بر اساس حروف الفبا تنها باعث می شد که عناوین کاملاً مشابه ذیل یکدیگر قرار گیرند و تنها وجود یک یا چند فاصله خالی باعث مرتب شدن دو عنوان مشابه در ردیف های متفاوتی می شد. علاوه بر این یافتن عناوین مشابه در بیش از ۶۵۰۰۰ رکورد کار مشکلی می نمود. لذا نخست با استفاده از دستور TRIM فاصله خالی اضافی (بیش از یک فاصله خالی بین کلمات یا عبارات) حذف شده سپس با استفاده از دستور Pivot table هر عنوان پایان نامه و موارد احتمالی تکرار آن مشخص شد (شکل شماره ۱) .

Microsoft Excel - ویرایش 2 - احتساب تکرار.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF

Type a question for help

Arial 10 B I U % , +.00 +.00

G8

تعداد تکرار

عنوان پایان نامه

Pivot Table دستور

PivotTable and PivotChart Wizard - Step 1 of 3

Where is the data that you want to analyze?

- ☒ Microsoft Excel list or database
- ☐ External data source
- ☐ Multiple consolidation ranges
- ☐ Another PivotTable report or PivotChart report

What kind of report do you want to create?

- ☒ PivotTable
- ☐ PivotChart report (with PivotTable report)

Cancel < Back Next >

ردیف	تعداد تکرار	عنوان پایان نامه
Total		عنوان
5		استرداد مجرمین
5		مهاجرت اعراب به ایران
4		بررسی فرهنگ سازمانی دانشگاه محقق اردبیلی و تأثیر آن در رضایت شغلی کارکنان
4		ارغینهای مطالعه کودکان در ده دستان تهران
4		عوامل اجتماعی جرم جوانان
4		وضع مسکن در ایران
3		ارزیابی اقتصادی تولید انرژی در نیروگاههای برق - آبی کوچک ایران (نیروگاه)
3		آثار اقتصادی سدسازی در ایران
3		بررسی اشتغال مادران کارمند و تأثیر آن بر هرم قدرت و تصمیمگیری در خانواده
3		بررسی صورتی و محتوایی روندهای ترکیب و اشتغال در شاخصنامه فردوسی و ام

Sheet1 علوم انسانی حذف شده 1 علوم پایه حذف شده 2 علا

start 2 Microsoft Word 2 Microsoft Excel EN 18:16

شکل شماره ۱- دستور Pivot table

پس از طبقه بندی پایان نامه ها به رشته های اصلی، عناوین تکراری یا فاقد رشته فرعی نیز حذف شدند که گزارش آن در جدول شماره ۲ آمده است.

جدول شماره ۲- عناوین حذف شده پس از تفکیک رشته های اصلی.

جمع	هنر	کشاورزی	فنی و مهندسی	علوم پزشکی	علوم پایه	علوم انسانی	حوزه
709	۲	۰	۴	۹	۱۳۱	۵۶۳	رشته فرعی اشتباه
3780	۵۵	۱۲۷	۳۰۲	۲۲۶۶	۱۹۳	۸۳۷	عناوین تکراری
۴۴۸۹	۵۷	۱۲۷	۳۰۶	۲۲۷۵	۳۲۴	۱۴۰۰	رکوردهای حذف شده

بدین ترتیب، با توجه به موارد یاد شده، ۹۷۸۵ عنوان پایان نامه از جامعه اولیه تحقیق به دلیل داشتن مشکلات فاحش حذف شد.^۱ در ادامه جامعه اصلی تحقیق در هر طبقه مشخص شده و با استفاده از فرمول $n = \frac{NZ^2 pq}{(d^2(N-1) + Z^2 pq)}$ (سرای، ۱۳۷۲، ۱۲۸) حجم نمونه مربوط به هر طبقه محاسبه شد. چون از تحقیقات صورت گرفته قبلی تخمین مناسبی برای p وجود نداشت بنابراین مقدار p مساوی ۰/۵ در نظر گرفته شد که بدین ترتیب بیشترین حجم نمونه ممکن برای انجام پژوهش به دست آمده است. همچنین مقدار خطای حاصل از برآورد آماری ۰/۵ و در نتیجه Z معادل ۱/۹۶ در نظر گرفته شد مقدار d نیز برابر ۱/۱ در نظر گرفته شد. بر این اساس حجم نمونه در هر طبقه به طور جداگانه محاسبه شده و نتایج زیر به دست آمد:

- علوم انسانی ۹۶
- علوم پایه ۹۳
- علوم پزشکی ۹۵
- فنی و مهندسی ۱۰۰
- کشاورزی ۹۴

^۱ گفتنی است یافته های مربوط به عناوین تکراری، فاقد توصیفگر، فاقد عنوان و ... به ریاست محترم مرکز گزارش شد و ایشان دستورات لازم در خصوص انجام اصلاحات را صادر نمودند.

به نظر بیبی (Babbie, 1979, 274) اگر نمونه احتمالی شامل کمتر از ۱۰۰ مورد باشد، به احتمال زیاد معرف و نماینده واقعی جامعه نخواهد بود. از طرف دیگر به منظور این که در مراحل بعدی تحقیق بتوان یافته‌های مربوط به نمونه‌های هر حوزه را از دیدگاه‌های مختلف با یکدیگر مقایسه نمود، حجم نمونه مربوط به هر طبقه با اندکی تغییر، ۱۰۰ رکورد در نظر گرفته شد. در ادامه با استفاده از دستور گزینش تصادفی (Sampling) در نرم افزار Excel از بین جامعه آماری هر حوزه ۱۰۰ رکورد به شکل کاملاً تصادفی انتخاب شد.

۳-۴- ابزار گردآوری داده‌ها:

داده‌های مورد نیاز پژوهش با استفاده از فهرست کنترلی (چک لیست) محقق ساخته ای گردآوری شد. این فهرست در قالب ۲۳ ستون (پیوست شماره ۲) طراحی گردید که حاوی موارد زیر است:

ردیف، تاریخ تهیه پایان نامه، رشته اصلی، رشته فرعی، عنوان، توصیفگر، تعداد واژه، تعداد کلید واژه‌های عنوان، تعداد کلید واژه مفرد، تعداد کلید واژه جمع، تعداد توصیفگر، تعداد توصیفگر مفرد، تعداد توصیفگر جمع، تعداد همخوانی کامل، تعداد همخوانی نسبی، مجموع همخوانی کامل و نسبی، تعداد عدم همخوانی، تعداد توصیفگر تک واژه، تعداد توصیفگر ترکیب واژه، کلید واژه انتخابی جهت بازیافت، تعداد بازیافت از طریق کلید واژه، توصیفگر انتخابی جهت بازیافت، تعداد بازیافت از طریق توصیفگر.

۳-۵- روش تجزیه و تحلیل اطلاعات:

داده‌های به دست آمده در قالب چک لیست یاد شده در جداول توزیع فراوانی ارائه گردیده، جداول و نمودارهای مربوط به هر یک ترسیم شد. برای متغیرهای کمی پژوهش شاخص‌های آماری میانگین محاسبه گردیده است. جهت پاسخگویی به آن دسته از سوالات تحقیق که رابطه بین متغیرها را مورد پژوهش قرار داده از آزمون "ضریب همبستگی پیرسن" و برای سوالاتی که ناظر بر تفاوت بین متغیرها بوده از آزمون "مجدورکای" (X^2) استفاده شده است.

۳-۶- محدودیت های پژوهش:

از جمله محدودیت های این پژوهش قرار داشتن اطلاعات "پایگاه چکیده پایان نامه های ایران" در محیط نرم افزاری CDS-ISIS بود. این محیط نرم افزاری فاقد امکانات لازم برای انجام بررسی ها و تحلیل های آماری بود. یکی دیگر از محدودیت ها، عدم پیروی از دستورالعمل های مشخص و مدوّن در نگارش فیلدهای مختلف و عدم انجام ویرایش های ادبی و فنی در این پایگاه بود. به عنوان مثال تعدادی از پایان نامه های "تربیت بدنی و علوم ورزشی" در حوزه علوم پزشکی و تعدادی دیگر در حوزه علوم انسانی گردآمده بود، و یا این که پایان نامه های مربوط به یک رشته واحد ذیل عناوین مختلف پراکنده شده بودند. برای مثال پایان نامه های مربوط به رشته "جامعه شناسی" ذیل این عناوین گرد آمده بود: "علوم اجتماعی" ۳۸۷ رکورد، "جامعه شناسی" ۳۲ رکورد و "جامعه شناسی" ۳۸۶ رکورد، و یا این که ۱۵۷ رکورد ذیل عبارت "روانشناسی" و ۱۵۱ رکورد ذیل عبارت "روانشناسی" درج شده بود. در بعضی موارد نیز رشته های فرعی در حوزه های نامربوط حضور یافته بودند. مثلاً تعدادی از پایان نامه های رشته نساجی در حوزه علوم انسانی مشاهده شد و یا برخی از پایان نامه های حوزه کشاورزی در حوزه فنی و مهندسی. جستجو و حذف رکورد های تکراری و یا حذف رکوردهایی که فاقد حداقل ملاک های لازم برای تحقیق بودند و نیز تلاش برای یکدست کردن جامعه تحقیق از مهم ترین مشکلات تحقیق بود.

فصل چهارم

یافته ها و تحلیل اطلاعات

۴-۱- مقدمه:

در این فصل سوالات پژوهش که در فصل نخست بیان شده بود به ترتیب مطرح شده، سپس با استفاده از داده های گردآوری شده در قالب ۳۳ جدول و ۶ نمودار و ۷ شکل، به این سوالات پاسخ داده شده است.

۴-۲- چه تفاوتی بین تعداد متوسط کلید واژه عنوان پایان نامه در حوزه های علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر وجود دارد؟

برای پاسخگویی به این پرسش برابر تعریف عملیاتی کلید واژه، ابتدا تعداد کلید واژه های هر یک از پایان نامه ها به تفکیک حوزه موضوعی شمارش شده سپس برابر جدول شماره ۳ تعداد متوسط کلید واژه ها در حوزه های مختلف شمارش شده است.

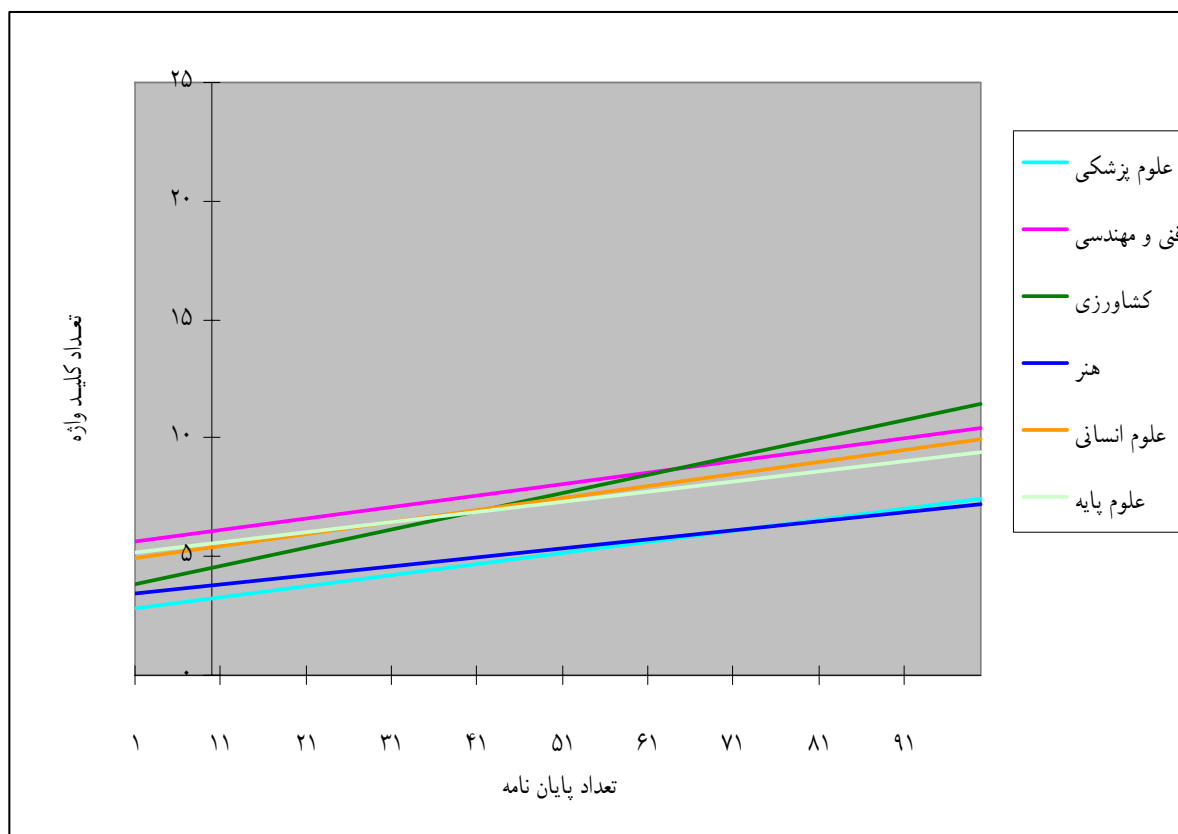
جدول شماره ۳- تعداد متوسط کلید واژه در عناوین پایان نامه های هر یک از حوزه های شش گانه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد پایان نامه	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰۰
تعداد کلید واژه های عنوان	۷۴۴	۷۳۰	۵۱۵	۸۰۴	۷۶۵	۵۳۴	۴۰۸۲
تعداد متوسط کلید واژه های عنوان	۷	۷	۵	۸	۸	۵	۷
نما	۶	۴	۳	۶	۶	۴	۶

در این جدول نشان داده شده که به طور کلی (با در نظر گرفتن کلیه حوزه ها)، هر عنوان پایان نامه دارای ۷ کلید واژه است. این تعداد در عناوین پایان نامه های حوزه "علوم پزشکی" و "هنر" کمتر از حد متوسط و در دیگر حوزه ها تقریباً مساوی با تعداد متوسط کلید واژه های عنوان است. شاید یکی از دلایل کوتاه بودن عنوان در آثار و تالیفات هنری این باشد که معمولاً هنرمندان عنوان را نیز چون رنگ ها، سایه روشن ها، اشکال

و صور مختلف، جزئی از جریان هنری می دانند و با نگرشی فلسفی برای یک عنوان مفاهیم بسیاری را تصور کرده و از مخاطبان نیز می خواهند که با کنکاش ذهنی از یک اثر یا عنوان کوتاه به یافته های بدیعی دست یابند. گاه هنرمندان کلام را نیز برای بیان هنری خویش قاصر می بینند و به همین دلیل در نمایشگاه آثار هنری گاه به تابلوها یا آثاری بر می خوریم که با عبارت "بدون عنوان" مشخص شده اند. از طرف دیگر انتخاب کوتاه ترین و در عین حال موثرترین عنوان جنبه های سمبولیک داشته و خود یک هنر است. در خصوص عنوان پایان نامه های حوزه علوم پزشکی نیز مشاهده شد که عمدتاً مطالعه یا بررسی بیماری خاصی مطرح شده و محقق با استفاده از چند کلید واژه در عنوان پایان نامه، مقصود خود را بیان داشته است. البته این نکته را نیز نباید از نظر دور داشت که تغییر سیاست مرکز در گردآوری پایان نامه های حوزه علوم پزشکی و افت شدید حضور پایان نامه های این حوزه در پایگاه مرکز (نمودار شماره ۳) باعث شد ۷۰ درصد پایان نامه های حضور یافته در نمونه تحقیق، مربوط به قبل از سال های ۱۳۶۰ (۲۰ سال قبل) باشد. چنین به نظر می رسد که با تخصصی تر شدن حوزه های تحقیق (از جمله انجام پایان نامه ها) تعداد کلید واژه ها عنوان نیز افزایش می یابد. به منظور ارزیابی این نکته، رکوردهای نمونه تحقیق به ترتیب تاریخ مرتب شده سپس با استفاده از امکانات برنامه نرم افزاری Excel نمودار سیر خطی^۱ تعداد کلید واژه های هر حوزه ترسیم شد (نمودار شماره ۱).

¹ Trendline



نمودار شماره ۱ - سیر خطی تعداد کلید واژه های عنوان بر حسب تاریخ

همان گونه که در نمودار شماره ۱ مشاهده می شود در کلیه حوزه ها احتمالاً به دلیل گسترش دانش و تخصصی تر شدن علوم متغیر های متنوع تر و بیشتری در عنوان پایان نامه ها مطرح شده و به همین دلیل با گذشت زمان تعداد کلید واژه های عنوان نیز افزایش یافته است.

در ادامه به منظور این که تفاوت متوسط تعداد کلید واژه های عنوان پایان نامه های حوزه های مورد بررسی، از منظر فرمول های آماری نیز سنجیده شود نخست تعداد کلید واژه ها در ستونی به ثبت رسیده و رکوردهای مربوط به هر حوزه با عددهای ۱ تا ۶ (هر عدد مختص به یک حوزه) از یکدیگر متمایز شدند. سپس با استفاده از آزمون ANOVA تفاوت تعداد متوسط کلید واژه های عنوان در حوزه های مختلف بررسی شد که نتیجه زیر به دست آمد:

جدول شماره ۴- آزمون ANOVA برای محاسبه اختلاف تعداد متوسط کلید واژه های عنوان

منبع	جمع مجزورات	درجه آزادی	میانگین مجزورات	نسبت ب	سطح معنی داری
واریانس بین گروه ها	777.140	5	155.428	14.062	.000
واریانس درون گروه ها	6565.420	594	11.053		
جمع	7342.560	599			

نتایج به دست آمده بیانگر معنی داری تفاوت بین تعداد متوسط کلید واژه های عنوان در حوزه های مختلف است.

آخرین ردیف جدول شماره ۳ بیانگر نما (مد^۱) یا فراوانی، تعداد کلید واژه عنوان در حوزه های مختلف است. با در نظر گرفتن کلیه حوزه ها، پایان نامه های با ۶ کلید واژه در عنوان بیشتر تکرار شده بود. نتایج به دست آمده از احتساب نما به تفکیک حوزه ها نیز تا حد زیادی شبیه به نتایج حاصل از محاسبه تعداد متوسط کلید واژه های عنوان بود. بدین معنی که در حوزه های علوم پزشکی و هنر به ترتیب با نما های ۳ و ۴ دارای پایان نامه هایی، با کلید واژه های کمتر از حد متوسط و پایان نامه های حوزه های علوم انسانی، فنی و مهندسی و کشاورزی نیز با نما ۶ دارای تکرار پایان نامه هایی در حد متوسط بودند. تنها نمای حوزه علوم پایه با تعداد متوسط کلید واژه های این حوزه همخوانی نداشت. شاید یکی از دلایل آن متنوع بودن عناوین پایان نامه های حوزه علوم پایه در مقایسه با دیگر حوزه ها است.

¹ Mode

در ادامه توزیع فراوانی پایان نامه ها بر حسب تعداد کلید واژه های عنوان به تفکیک حوزه ها بررسی شد که نتایج آن در جدول شماره ۵^۱ درج شده است.

جدول شماره ۵^۱ - توزیع فراوانی پایان نامه ها بر حسب تعداد کلید واژه های عنوان ، به تفکیک حوزه.

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع	
تعداد کلید واژه	تعداد	تعداد	تعداد	تعداد	تعداد	تعداد	درصد	تعداد
1	۰	۱	۳	۰	۱	۲	۷٪	۷
2	۴	۴	۱۴	۲	۳	۱۱	۳۸٪	۳۸
3	۸	۳	۱۷	۱	۷	۱۶	۵۲٪	۵۲
4	۱۲	۱۷	۱۴	۶	۸	۲۰	۷۷٪	۷۷
5	۱۱	۱۴	۱۶	۸	۱۰	۱۲	۷۱٪	۷۱
6	۱۷	۹	۹	۲۰	۱۳	۱۲	۸۰٪	۸۰
7	۸	۱۳	۹	۱۱	۱۰	۶	۵۷٪	۵۷
8	۹	۶	۲	۱۳	۱۲	۸	۵۰٪	۵۰
9	۷	۹	۶	۱۲	۱۰	۴	۴۸٪	۴۸
10	۳	۸	۴	۱۰	۹	۲	۳۶٪	۳۶
بیشتر از 10	۲۱	۱۶	۶	۱۷	۱۷	۷	۸۴٪	۸۴
جمع	100	100	100	100	100	100	۶۰۰٪	۶۰۰

جدول شماره ۵ نشان می دهد که پایان نامه های حوزه های مختلف از نظر تعداد کلید واژه های عنوان دارای توزیع مشابه ای نیستند. به عنوان مثال از این نظر پایان نامه های حوزه علوم پزشکی و حوزه فنی و مهندسی تقریباً حالتی متضاد دارند. در حوزه علوم پزشکی ۶۴ درصد پایان نامه ها دارای ۱ تا ۵ کلید واژه و ۳۶ درصد بیش از ۶ کلید واژه دارند در صورتی که در حوزه فنی و مهندسی تنها ۱۷ درصد پایان نامه ها دارای ۱ تا

۱- با توجه به این که تعداد نمونه مورد بررسی در هر حوزه ۱۰۰ می باشد، اعداد به دست آمده در واقع نمایانگر درصد نیز خواهد بود. به همین دلیل از ذکر مقدار درصد در ذیل حوزه ها خودداری شده است. به عنوان مثال در حوزه علوم انسانی تعداد پایان نامه با ۳ کلید واژه ۸ رکورد، یا ۸٪ از صد (۸٪) است.

۵ کلید واژه و ۸۳ درصد دارای بیش از ۶ کلید واژه هستند. وجود متغیرهای بیشتر و متنوع تر در حوزه فنی و مهندسی را می توان یکی از دلایل تجمع فراوانی بیشتر با بیش از ۶ کلید واژه در عنوان دانست.

به منظور این که تصویر روشن تری از ترکیب و ساختار واژگانی عناوین پایان نامه ها در حوزه های مختلف به دست داده شود تعداد متوسط "واژه" در عنوان پایان نامه های حوزه های مختلف نیز بررسی شد که داده های به دست آمده در جدول شماره ۶ مشاهده می شود. گفتنی است منظور از "واژه" کلیه کلمات با بار معنایی (Go list) و بدون بار معنایی (Stop list) است. معمولاً در پایگاه های اطلاعاتی صورتی از واژه های بدون بار معنایی (مانند حروف ربط، حروف اضافه، حروف تعریف، ضمائر و ...) تهیه می شود تا با بازداشتن کامپیوتر از انتخاب این واژه ها نتایج زیر حاصل شود:

۱- جلوگیری از افزوده شدن مدخل های بی فایده به پایگاه،

۲- جلوگیری از حجیم شدن بی دلیل نمایه نامه و نیز افزایش قیمت آن،

۳- کاستن از بار زحمتی که در جستجوی موضوعات، به واسطه مدخل های بی ربط بر دوش جستجوگر گذاشته می شود.

جدول شماره ۶- تعداد متوسط واژه ها ، در عناوین پایان نامه های هر یک از حوزه های شش گانه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	تجارت
تعداد پایان نامه	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰۰
تعداد واژه های عنوان	۱۰۰۳	۱۱۲۱	۸۳۵	۱۰۴۱	۱۰۰۱	۶۶۸	۵۶۶۹
تعداد متوسط واژه های عنوان	۱۰	۱۱	۸	۱۰	۱۰	۷	۹

مقایسه جدول فوق با جدول شماره ۳ نشان می دهد که واژه های عنوان نیز مانند کلید واژه های عنوان در حوزه "علوم پزشکی" و "هنر" کمتر از حد متوسط است. در عین حال در حوزه علوم پایه تعداد واژه ها بر خلاف تعداد کلید واژه های این حوزه بیشتر از حد میانگین است. تفاضل تعداد واژه های عنوان و کلید واژه ها نشان دهنده تعداد واژگان بدون بار معنایی خواهد بود. که نتایج آن در جدول شماره ۷ آمده است.

جدول شماره ۷- تعداد متوسط واژه های غیر موضوعی در عناوین پایان نامه های هر یک از حوزه های شش گانه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	سایر
تعداد پایان نامه	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰۰
تعداد واژه های غیر موضوعی	۲۵۹	۳۹۱	۳۲۰	۲۳۷	۲۳۶	۱۳۴	۱۵۷۷
تعداد متوسط واژه های غیر موضوعی	۳	۴	۳	۲	۲	۱	۳

همان گونه که مشاهده می شود به طور کلی هر عنوان پایان نامه دارای ۳ واژه غیر موضوعی است. تعداد این واژه ها در حوزه علوم پایه بیش از دیگر حوزه ها است. شاید یکی از دلایل زیاد بودن واژگان غیر موضوعی در حوزه علوم پایه ضرورت بهره گیری از "حروف ربط" به منظور نشان دادن رابطه بین متغیرهای خاص در هر عنوان تحقیق باشد. به عنوان مثال در پایان نامه ای با عنوان "بررسی و شناخت ساختمان بلوری و خواص مختلف مواد رسی و کاربرد آن در صنایع سرامیک به وسیله دیفراکسیون /شعه ایکس" متغیر های زیادی چون "ساختمان بلوری"، "خواص مواد رسی"، "صنایع سرامیک"، "دیفراکسیون" و "اشعه ایکس" مطرح شده و به منظور نشان دادن نوع رابطه بین این مفاهیم از ۸ واژه غیر موضوعی استفاده شده است. حال سوالی که مطرح می شود این است که در حوزه های مختلف بین میزان عناصر تشکل دهنده عنوان (اعم از کلید واژه یا واژگان

غیر موضوعی) چه تفاوتی وجود دارد؟ جدول شماره ۸ تصویری کلی از تفاوت های تعداد کلید واژه ها و واژه های غیر موضوعی به دست می دهد.

جدول شماره ۸- مقایسه تعداد کلید واژه و واژه های غیر موضوعی در عناوین پایان نامه های حوزه های شش گانه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد واژه	۱۰۰۳	۱۱۲۱	۸۳۵	۱۰۴۱	۱۰۰۱	۶۶۸	۵۶۶۹
تعداد کلید واژه	۷۴۴ (۷۴٪)	۷۳۰ (۶۵٪)	۵۱۵ (۶۲٪)	۸۰۴ (۷۷٪)	۷۶۵ (۷۶٪)	۵۳۴ (۸۰٪)	۴۰۹۲ (۷۲٪)
تعداد واژه غیر موضوعی	۲۵۹ (۲۶٪)	۳۹۱ (۳۵٪)	۳۲۰ (۳۸٪)	۲۳۷ (۲۳٪)	۲۳۶ (۲۴٪)	۱۳۴ (۲۰٪)	۱۵۷۷ (۲۸٪)

علاوه بر یافته فوق بر اساس آزمون مجذور کای، که خلاصه آن در جدول شماره ۹ آمده مشخص شد

که بین تعداد کلید واژه های عنوان حوزه های مختلف تفاوت معنی داری وجود دارد.

جدول شماره ۹- محاسبه مجذور کای در خصوص تفاوت تعداد کلید واژه ها و واژه های غیر موضوعی عنوان

کلید واژه ها			واژه های غیر موضوعی			محاسبه مجذور کای
سطح معنی داری	درجه آزادی	آماره آزمون	سطح معنی داری	درجه آزادی	آماره آزمون	
/۰۰۰	۵	۱۱۳/۹۵۰	/۰۰۰	۵	۱۴۳/۴۱۷	

همچنین تفاوت تعداد واژه های غیر موضوعی حوزه های مختلف نیز براساس آزمون مجذور کای

محاسبه شد که نتایج به دست آمده از این محاسبه در جدول شماره ۱۰ بیانگر تفاوت معنی دار بین تعداد واژه های غیر موضوعی عناوین مختلف است.

جدول شماره ۱۰- محاسبه مجذور کای در خصوص تفاوت تعداد واژه های غیر موضوعی.

	VAR00001
Chi-Square(a)	143.417
df	5
Asymp. Sig.	.000

با توجه به مطالب فوق در پاسخ به سوال نخست تحقیق باید اذعان داشت که عناوین پایان نامه های حوزه های مختلف نه تنها از نظر تعداد کلید واژه های عنوان بلکه از نظر دیگر عناصر تشکیل دهنده عنوان (مانند حروف ربط، حروف اضافه) نیز تفاوت دارند.

۳-۴- بین تعداد کلید واژه های عنوان پایان نامه ها با تعداد توصیفگرهای اختصاص یافته چه رابطه ای وجود دارد؟

پاسخگویی به این سوال مستلزم این بود که توزیع فراوانی تعداد کلید واژه ها و میزان تبعیت توصیفگرها از این توزیع مشخص شود. بدین منظور پایان نامه ها بر اساس تعداد کلیدواژه های عنوان به ۶ طبقه دسته بندی شده و پس از آن مجموع کلید واژه ها و مجموع توصیفگرهای هر طبقه شمارش شد تا توزیع فراوانی هر یک مشخص شود. جدول شماره ۱۱ نمایانگر نتایج حاصل از این بررسی است.

جدول شماره ۱۱- توزیع فراوانی تعداد کلید واژه های عنوان و تعداد توصیفگرها به تفکیک حوزه.

حوزه	علوم انسانی		علوم پایه		علوم پزشکی		فنی و مهندسی		کشاورزی		هنر		جمع															
تعداد کلید واژه	کلید واژه	توصیفگر	کلید واژه	توصیفگر	کلید واژه	توصیفگر	کلید واژه	توصیفگر	کلید واژه	توصیفگر	کلید واژه	توصیفگر	کلید واژه	توصیفگر														
۱-۳	(۴٪)	۳۲	(۷٪)	۲۹	(۲٪)	۱۸	(۳٪)	۱۴	(۱۶٪)	۸۲	(۲۵٪)	۸۱	(۱٪)	۷	(۳٪)	۱۸	(۴٪)	۲۸	(۴٪)	۱۸	(۱۳٪)	۷۲	(۲۵٪)	۹۰	(۶٪)	۲۳۹	(۱۰٪)	۲۵۰
۴-۶	(۲۸٪)	۲۰۵	(۳۵٪)	۱۵۱	(۲۶٪)	۱۸۸	(۳۳٪)	۱۶۵	(۳۷٪)	۱۹۰	(۳۶٪)	۱۱۴	(۲۳٪)	۱۸۴	(۲۷٪)	۱۵۱	(۲۱٪)	۱۶۰	(۲۲٪)	۹۳	(۴۰٪)	۲۱۲	(۴۲٪)	۱۵۴	(۲۸٪)	۱۱۳۹	(۳۲٪)	۸۲۸
۷-۹	(۲۶٪)	۱۹۱	(۲۶٪)	۱۰۹	(۳۰٪)	۲۲۰	(۳۴٪)	۱۷۲	(۲۶٪)	۱۳۳	(۲۴٪)	۷۷	(۳۶٪)	۲۸۹	(۳۷٪)	۲۰۷	(۳۳٪)	۲۵۶	(۳۸٪)	۱۶۳	(۲۷٪)	۱۴۲	(۱۹٪)	۷۱	(۳۰٪)	۱۲۳۱	(۳۱٪)	۷۹۹
۱۰-۱۲	(۱۶٪)	۱۲۲	(۱۳٪)	۵۷	(۲۰٪)	۱۴۹	(۱۴٪)	۶۹	(۱۶٪)	۸۴	(۱۳٪)	۴۰	(۲۵٪)	۲۰۳	(۲۳٪)	۱۳۰	(۱۸٪)	۱۳۷	(۲۰٪)	۸۴	(۱۷٪)	۸۹	(۱۰٪)	۳۶	(۱۹٪)	۷۸۴	(۱۶٪)	۴۱۶
۱۳-۱۵	(۱۷٪)	۱۲۳	(۱۳٪)	۵۷	(۱۹٪)	۱۳۶	(۱۵٪)	۷۴	(۵٪)	۲۶	(۳٪)	۸	(۱۰٪)	۸۳	(۸٪)	۴۷	(۱۹٪)	۱۴۷	(۱۳٪)	۵۴	(۰٪)	۰	(۰٪)	۰	(۱۳٪)	۵۱۵	(۹٪)	۲۴۰
بیش از ۱۵	(۱۰٪)	۷۱	(۶٪)	۲۴	(۳٪)	۱۹	(۲٪)	۱۱	(۰٪)	۰	(۰٪)	۰	(۵٪)	۳۸	(۲٪)	۱۲	(۵٪)	۳۷	(۴٪)	۱۵	(۴٪)	۱۹	(۴٪)	۱۴	(۴٪)	۱۸۴	(۳٪)	۷۶
جمع	۷۴۴	۴۲۷	۷۳۰	۵۰۵	۵۱۵	۳۲۰	۸۰۴	۵۶۵	۷۶۵	۴۲۷	۵۳۴	۳۶۵	۴۰۹۲	۲۶۰۹	جمع	۲۶۰۹	۴۰۹۲	۳۶۵	۵۳۴	۴۲۷	۵۳۴	۳۶۵	۴۰۹۲	۲۶۰۹	جمع	۲۶۰۹	۴۰۹۲	

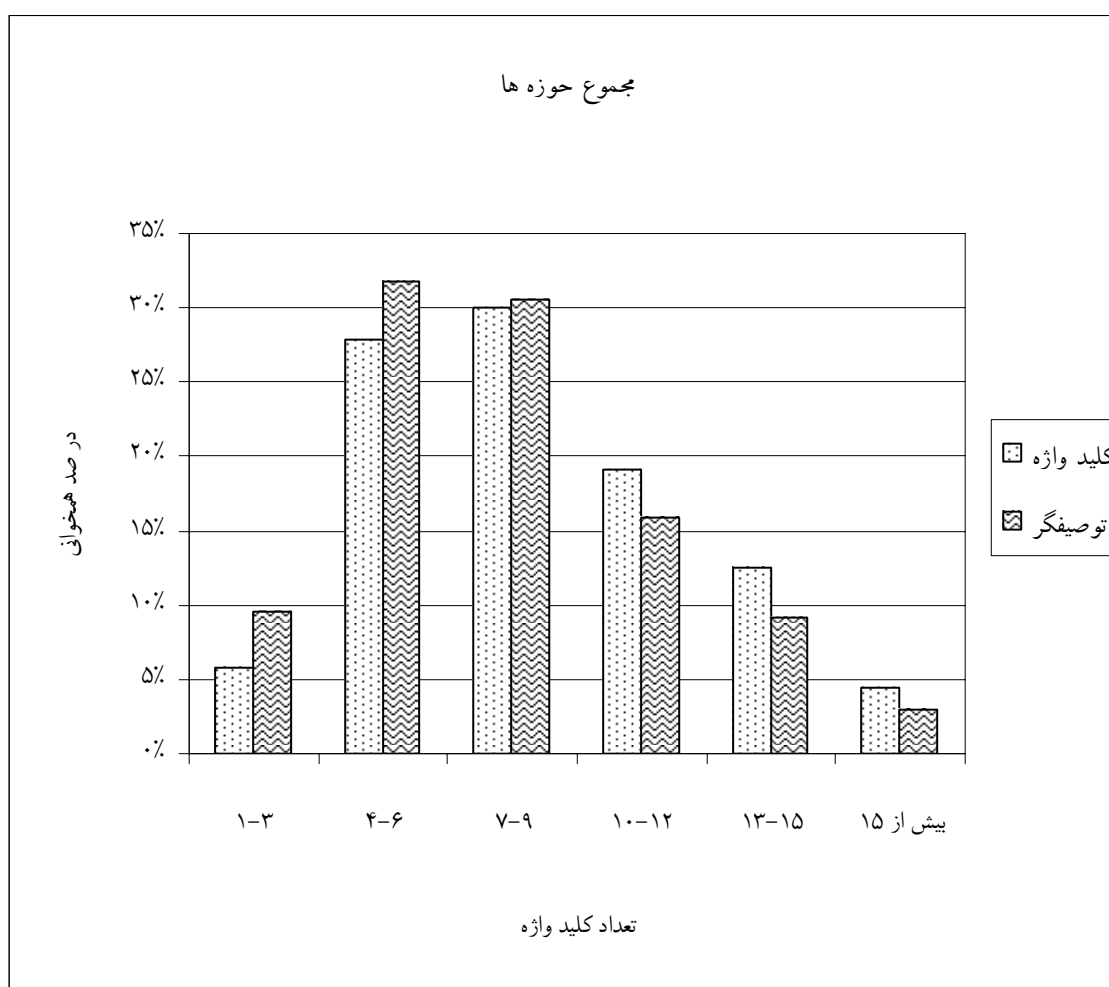
همان طور که مشاهده می شود به طور کلی (با احتساب کلیه حوزه ها) تعداد توصیفگرهای اختصاص یافته

از تعداد کلید واژه های عنوان پایان نامه ها تاثیر می پذیرد و در مواقعی که عنوان پایان نامه ها دارای تعداد

کلید واژه هایی در حد متوسط^۱ (۹ کلید واژه در عنوان) است بیشترین میزان همخوانی با تعداد توصیفگرها وجود دارد. کلر (Keller, 1992) نیز در پژوهشی که بر روی پایان نامه های مقطع کارشناسی ارشد کتابخانه دانشگاه ایالتی ایندیانا انجام داد و به مقایسه نخستین سرعنوان موضوعی کنگره- که به پایان نامه اختصاص داده شده بود - با واژه های عنوان پرداخت به این نتیجه رسید که بیشترین میزان همپوشانی بین کلید واژه های عنوان و سرعنوان های موضوعی زمانی است که پایان نامه ها دارای ۷-۱۲ کلید واژه در عنوان هستند. ولی افزایش تعداد کلید واژه های عنوان به بیش از حد متوسط، به همان نسبت منجر به افزایش تعداد توصیفگرها

^۱ جدول شماره ۵

نخواهد شد. در نمودار شماره ۲ با در نظر گرفتن کلیه حوزه ها تصویر روشن تری از تعداد کلید واژه های عنوان و تعداد توصیفگرها به دست داده شده است.



نمودار شماره ۲- توزیع فراوانی تعداد کلید واژه ها و توصیفگرها

۴-۴- چند درصدِ توصیفگرها، از کلید واژه های عنوان گزینش شده اند؟ به عبارت دیگر همخوانی کامل

کلید واژه های عنوان با توصیفگرها چقدر است؟

۴-۵- در حوزه های علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر چه تفاوتی بین

میزان همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان وجود دارد ؟

به منظور پاسخگویی به دو سوال فوق وضعیت همخوانی کلید واژه های عنوان با توصیفگرها، در قالب ۳

حالت زیر بررسی شد:

۱- همخوانی کامل، مانند: توصیفگر "جرائم مطبوعاتی"، برای پایان نامه ای با عنوان "تحقیقی پیرامون

جرائم مطبوعاتی". در این حالت توصیفگر دقیقاً با یکی از کلید واژه های عنوان همخوان است.

۲- همخوانی نسبی، مانند: توصیفگر "مطبوعات"، برای پایان نامه ای با عنوان "تحقیقی پیرامون جرائم

مطبوعاتی". در این حالت توصیفگر با یکی از کلید واژه های عنوان هم ریشه است و تنها حالت وصفی

آن تغییر کرده است.

۳- عدم همخوانی، مانند: توصیفگر "نظام های حقوقی"، برای پایان نامه ای با عنوان "تحقیقی پیرامون

جرائم مطبوعاتی". در این حالت عبارت "نظام های حقوقی" به عنوان یک توصیفگر به شکلی با عنوان

فوق مربوط است ولی به هیچ شکلی با کلید واژه های این عنوان همخوانی ندارد.

پس از بررسی حالت های ذکر شده در نمونه تحقیق، نتایج به دست آمده در جدول شماره ۱۳ گرد آمده

است.

جدول شماره ۱۳- وضعیت همخوانی کلید واژه های عنوان با توصیفگرها ، به تفکیک حوزه.

حوزه		علوم انسانی		علوم پایه		علوم پزشکی		فنی و مهندسی		کشاورزی		هنر		جمع	
وضعیت همخوانی	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد
همخوانی کامل	۲۱۰	۴۹٪	۲۰۹	۴۱٪	۱۶۰	۵۰٪	۲۰۰	۳۵٪	۲۱۵	۵۰٪	۱۹۰	۵۲٪	۱۱۸۴	۴۵٪	
همخوانی نسبی	۹۲	۲۲٪	۱۲۰	۲۴٪	۶۱	۱۹٪	۱۳۰	۲۳٪	۹۰	۲۱٪	۸۱	۲۲٪	۵۷۴	۲۲٪	
عدم همخوانی	۱۲۵	۲۹٪	۱۷۵	۳۵٪	۹۹	۳۱٪	۲۳۵	۴۲٪	۱۲۲	۲۹٪	۹۴	۲۶٪	۸۵۰	۳۳٪	
جمع	۴۲۷	۱۰۰٪	۵۰۴	۱۰۰٪	۳۲۰	۱۰۰٪	۵۶۵	۱۰۰٪	۴۲۷	۱۰۰٪	۳۶۵	۱۰۰٪	۲۶۰۸	۱۰۰٪	

داده های این جدول نشان می دهد که به طور کلی، ۴۵ درصد از توصیفگرهای نمایه سازان را، کلید واژه های عنوان تشکیل می دهد. به عبارت دیگر در ۴۵ درصد موارد، میان کلید واژه های عنوان و توصیفگرهای انتخاب شده توسط نمایه سازان همخوانی کامل وجود دارد. بیشترین میزان همخوانی کامل به ترتیب در حوزه های هنر (۵۲٪)، کشاورزی (۵۰٪)، علوم پزشکی (۵۰٪) و علوم انسانی (۴۹٪) و کمترین میزان همخوانی کامل کلید واژه های عنوان با توصیفگرهای نمایه سازان در حوزه علوم پایه (۴۱٪) و فنی و مهندسی (۳۵٪) است.

با توجه به این که توصیفگرهای دارای همخوانی نسبی با کلید واژه های عنوان نیز همانند موارد همخوانی کامل، به شکلی از عنوان پایان نامه ها استخراج شده بودند، موارد همخوانی (اعم از همخوانی کامل و نسبی) با موارد عدم همخوانی مقایسه شد که نتایج آن در جدول شماره ۱۴ مشاهده می شود.

جدول شماره ۱۴- مقایسه همخوانی و عدم همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان به تفکیک حوزه.

حوزه	علوم انسانی		علوم پایه		علوم پزشکی		فنی و مهندسی		کشاورزی		هنر		جمع	
	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد
وضعیت همخوانی														
همخوانی کامل و نسبی	۳۰۲	۷۱٪	۳۲۹	۶۵٪	۲۲۱	۶۹٪	۳۳۰	۵۸٪	۳۰۵	۷۱٪	۲۷۱	۷۴٪	۱۷۵۸	۶۷٪
عدم همخوانی	۱۲۵	۲۹٪	۱۷۵	۳۵٪	۹۹	۳۱٪	۲۳۵	۴۲٪	۱۲۲	۲۹٪	۹۴	۲۶٪	۸۵۰	۳۳٪
جمع	۴۲۷	۱۰۰٪	۵۰۴	۱۰۰٪	۳۲۰	۱۰۰٪	۵۶۵	۱۰۰٪	۴۲۷	۱۰۰٪	۳۶۵	۱۰۰٪	۲۶۰۸	۱۰۰٪

همان گونه که در این جدول نیز مشاهده می شود ۶۷ درصد توصیفگرهای نمایه سازان- به طور کامل یا نسبی- با کلید واژه های عنوان همخوانی دارند. در عین حال با اندکی تغییر در مقایسه با جدول شماره ۱۳ مشاهده می شود که همچنان حوزه های هنر (۷۴٪)، کشاورزی (۷۱٪)، علوم انسانی (۷۱٪) و علوم پزشکی (۶۹٪) دارای همخوانی بیشتری در مقایسه با حوزه های علوم پایه (۶۵٪) و فنی و مهندسی (۵۸٪) هستند. یافته های حاصل از این تحقیق تا حد زیادی با یافته های پژوهش های دیگر که در فصل دوم به آنها اشاره شد نزدیک است. به عنوان مثال بزرگی (۱۳۷۷) نیز در تحقیق خود به این نتیجه رسید که ۷۳ درصد بین موضوعات و عنوان کتاب های فارسی همخوانی وجود دارد. و یا این که انصاری (۱۳۷۹) نیز ۷۰ درصد همخوانی را بین توصیفگرها و کلید واژه های عنوان پایان نامه های پزشکی نشان می دهد. همچنین فراست (Frost, 1989) در

تحقیقی بر روی فهرست کتابخانه دانشگاه میشیگان، به نتیجه همخوانی ۷۳ درصدی بین واژه های عنوان با سرعنوان های LCSH^۱ دست یافت.

در عین حال بر خلاف تصور محقق میزان همخوانی کامل در حوزه های فنی و مهندسی و علوم پایه کمتر از دیگر حوزه ها بود. لذا به منظور این که بتوان تأثیر خطاهای انسانی را به حداقل کاهش داده و به شکلی صحت یافته های تحقیق حاضر را ارزیابی نمود، و نیز زمینه بهره گیری از امکانات نرم افزارهای رایج را برای انجام تحقیقات مشابه فراهم ساخت، سعی شد تا میزان همخوانی کامل توصیفگرها و کلید واژه های عنوان به شیوه ماشینی محاسبه شود. برای انجام این کار از امکانات برنامه نرم افزاری Excel استفاده شد. در ابتدا محاسبه ماشینی میزان همخوانی کامل در نمونه تحقیق صورت گرفت تا دقت یا خطای محاسبه ماشینی برآورد شود. شیوه انجام کار بدین شکل بود که در مرحله اول به هنگام انتقال توصیفگرهای اختصاص داده شده به هر عنوان از محیط نرم افزاری CDS ISIS به محیط نرم افزاری Excel با استفاده از علامت اسلش " / " توصیفگرها از یکدیگر جدا شدند، سپس با استفاده از دستور Find^۲ مجموعه کاراکترهای قبل از هر اسلش در ستون مربوط به توصیفگرها (ستون C) شمارش شده و در سلول های جداگانه ای درج گردید. برای شمارش تعداد توصیفگرها ۱۹ سلول به عنوان پیش فرض در نظر گرفته شد. اگر دستور Find به کاراکتری برخورد نمی کرد در سلول مربوطه علامت # درج می شد (شکل شماره ۲).

¹ Library of Congress Subject Heading

^۲ دستور Find بر اساس شمارش کاراکتر عمل می کند.

Microsoft Excel - Book1.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF Type a question for help

Badr 12 B I U % , +.00 +.00 100%

G2 =FIND("/",C2)

	H	G	F	E	D	C	B	
1						توصیفگر	عنوان	
2	۷	۲۸	۹	۱	۲۰٪	حاکمیت / انقلاب ارتباطاتی / ارتباط / رسانه گروهی / حکومت	انقلاب ارتباطات و تحول مفهوم حاکمیت	
3	۹	۲۱	۶	۲	۵۰٪	بررسی علل چندشغلی بودن معلمان مقطع متوسطه در معلم / آموزش متوسطه / کرمان / چندشغلی شهر کرمان		
4	۵	۲۳	۱۱	۱	۲۵٪	ساختمان حکومتی ایالت کالیفرنیا و اثرگذاری هوای قوه مجریه / کالیفرنیا / قوه مقننه / قوه قضائیه	سه گانه بر یکدیگر	
5	:	۲۰	۷	۲	۱۰۰٪	الگوهای تصمیم گیری در تولید یک بررسی نمونه ای تولید / تصمیم گیری	در ایران	
6	۹	۲۱	۷	۳	۵۰٪	بررسی عوامل تعیین کننده مالیات بر واردات در ایران درآمد / مالیه عمومی / ایران / اقتصاد / مالیات / واردات		
7	:	۲۰	۱۲	۲	۱۰۰٪	توسعه صنعت سیمان در ایران صنعت سیمان / سیمان		

et2 Sheet1 10:09

start نمونه 2 چهارم ویرایش... Book1.xls EN

شکل شماره ۲ - استفاده از دستور Find به منظور شناسایی توصیفگرها

در ادامه با استفاده از دستور Count تعداد سلول های حاوی عدد، در ۱۹ سلول از پیش تعیین شده (سلول های G تا Y) شمارش شده و در سلول جداگانه ای به عنوان تعداد توصیفگرهای یک عنوان پایان نامه ثبت گردید. بدین ترتیب تعداد توصیفگرهای هر عنوان پایان نامه مشخص شد (شکل شماره ۳).

1	عنوان	توصیفگر							
2	انقلاب ارتباطات و تحول مفهوم حاکمیت	حاکمیت / انقلاب ارتباطاتی / ارتباط / رسانه / گروهی / حکومت	۵	۲۰٪	۱	۹	۲۸	۷	
3	بررسی علل چندشغلی بودن معلمان مقطع متوسطه در معلم / آموزش متوسطه / کرمان / چندشغلی شهر کرمان		۴	۵۰٪	۲	۶	۲۱	۹	
4	ساختار حکومتی ایالت کالیفرنیا و اثرگذاری قوای قوه مجریه / کالیفرنیا / قوه مقننه / قوه قضائیه سه گانه بر یکدیگر		۴	۲۵٪	۱	۱۱	۲۳	۵	
5	الگوهای تصمیم گیری در تولید یک بررسی نمونه ای تولید / تصمیم گیری در ایران		۲	۱۰۰٪	۲	۷	۲۰		
6	بررسی عوامل تعیین کننده مالیات بر واردات در ایران درآمد / مالیه عمومی / ایران / اقتصاد / مالیات / واردات		۶	۵۰٪	۳	۷	۲۱	۹	
7	توسعه صنعت سیمان در ایران صنعت سیمان / سیمان		۲	۱۰۰٪	۲	۱۲	۲۰		

شکل شماره ۳- استفاده از دستور Count برای شمارش تعداد توصیفگرها.

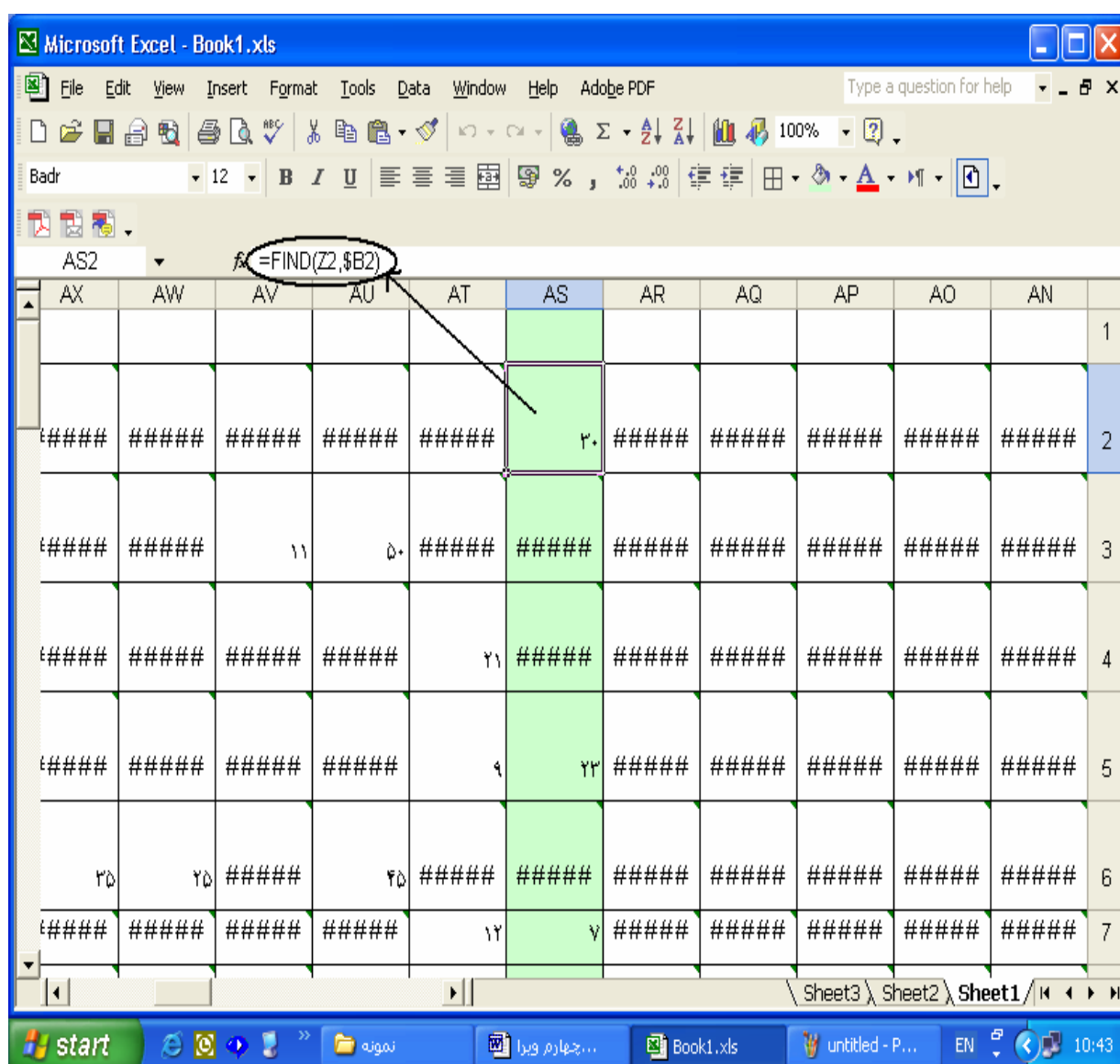
در ادامه با استفاده از دستور Mid^۱ مجموعه کاراکترهای قبل از هر " / " (یا در واقع یک توصیفگر) در سلول جداگانه ای درج شد. به منظور درج توصیفگرها نیز ۱۹ سلول به عنوان پیش فرض در نظر گرفته شد (شکل شماره ۴).

	AC	AB	AA	Z	Y	X	W	V	U
1									
2	حکومت	رسانه گروهی	ارتباط	انقلاب ارتباطاتی	حاکمیت	#####	#####	#####	#####
3	چندشغلی	کرمان	آموزش متوسطه	معلم	#####	#####	#####	#####	#####
4	قوه قضائیه	قوه مقننه	کالیفرنیا	قوه مجریه	#####	#####	#####	#####	#####
5	#VALUE!	#VALUE!	تصمیم گیری	تولید	#####	#####	#####	#####	#####
6	مالیات	اقتصاد	ایران	مالیه عمومی	درآمد	#####	#####	#####	#####
7	#VALUE!	#VALUE!	سیمان	صنعت سیمان	#####	#####	#####	#####	#####

شکل شماره ۴ - دستور Mid برای ثبت توصیفگرها

^۱ دستور Mid کاراکترهای شمارش شده را در محل تعیین شده به شکل متن درج می کند.

پس از آن دوباره از فرمان Find استفاده شد تا توصیفگرهای یافت شده در مرحله قبل را با ستونی که عناوین در آن درج شده بود (ستون B) مقایسه کند. با استفاده از این دستور در صورت همخوانی کامل کاراکترها با هر یک از کلید واژه های عنوان، عدد یا همان تعداد کاراکترها به ثبت می رسید و در صورت عدم همخوانی علامت # درج می شد (شکل شماره ۵).



شکل شماره ۵ - دستور Find برای یافتن موارد همخوانی کامل

در پایان با استفاده از دستور Count تعداد سلول های حاوی عدد(از سلول AS تا سلول BK) شمارش شده و در سلول مجاور تعداد توصیفگرها درج شد. بدین ترتیب تعدادهمخوانی کامل نیز مشخص شده بود (شکل شماره ۶).

ردیف	عنوان	توصیفگر	D	E	F	G
1						
2	۶۸	انقلاب ارتباطات و تحول مفهوم حاکمیت	حاکمیت / انقلاب ارتباطاتی / ارتباط ط / رسانه	۵	۲۰٪	۷
3	۲۰۱	بررسی علل چندشغلی بودن معلمان مقطع متوسطه در	معلم / آموزش متوسطه / کرمان / چندشغلی	۴	۵۰٪	۶
4	۴۷۲	ساختار حکومتی ایالت کالیفرنیا و اثرگذاری قوای	قوه مجریه / کالیفرنیا / قوه مقننه / قوه قضائیه	۴	۲۵٪	۱۱
5	۸۸۷	الگوهای تصمیم گیری در تولید یک بررسی نمونه ای	تولید / تصمیم گیری	۲	۱۰۰٪	۷
6	۱۰۸۵	بررسی عوامل تعیین کننده مالیات بر واردات در ایران	درآمد / مالیات عمومی / ایران / اقتصاد / مالیات / واردات	۶	۵۰٪	۷
7	۱۲۹۷	توسعه صنعت سیمان در ایران	صنعت سیمان / سیمان	۲	۱۰۰٪	۱۲

شکل شماره ۶ - دستور Count برای شمارش تعداد همخوانی کامل

با تقسیم سلول مربوط به تعداد توصیفگرها (سلول F2)، با سلول مربوط به تعداد همخوانی کامل (سلول D2)، درصد همخوانی کامل در خصوص هر عنوان پایان نامه به دست می‌آید (شکل شماره ۷).

ردیف	عنوان	توصیفگر
1	انقلاب ارتباطات و تحول مفهوم حاکمیت	حاکمیت / انقلاب ارتباطاتی / ارتباط ط / رسانه
2	بررسی علل چندشغلی بودن معلمان مقطع متوسطه در	آموزش متوسطه / کرمان / چندشغلی
3	ساختار حکومتی ایالت کالیفرنیا و اثرگذاری قوای	قوه مجریه / کالیفرنیا / قوه مقننه / قوه قضائیه
4	الگوهای تصمیم‌گیری در تولید یک بررسی نمونه‌ای	تولید / تصمیم‌گیری
5	بررسی عوامل تعیین‌کننده مالیات بر واردات در ایران	درآمد / مالیه عمومی / ایران / اقتصاد / مالیات
6	توسعه صنعت سیمان در ایران	صنعت سیمان / سیمان
7	سده گانه بر یکدیگر	سده گانه بر یکدیگر

شکل شماره ۷ - درصد همخوانی کامل در هر عنوان.

با انتخاب دریف نخست که فرمول های یاد شده به طور کامل در سلول های آن درج شده بود (ردیف شماره ۲) و کشیدن آن به سمت پایین، به شکل خودکار فرمول ها در ردیف های دیگر نیز کپی شده و محاسبات انجام شد. در جدول شماره ۱۵ نتایج حاصل از محاسبه ماشینی با محاسبه دستی در نمونه تحقیق مقایسه شده است.

جدول شماره ۱۵- مقایسه بررسی همخوانی کامل کلید واژه های عنوان با توصیفگرها، به شیوه دستی و ماشینی

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
محاسبه دستی تعداد توصیفگر	۴۲۷	۵۰۵	۳۲۰	۵۶۵	۴۲۷	۳۶۵	2609
محاسبه ماشینی تعداد توصیفگر	۴۱۹	۴۹۰	۳۰۱	۵۵۷	۴۱۰	۳۵۸	2535
محاسبه دستی تعداد همخوانی کامل	(%۴۹)	(%۴۱)	(%۵۰)	(%۳۵)	(%۵۰)	(%۵۲)	(%45)
محاسبه ماشینی تعداد همخوانی کامل	۲۱۰	۲۰۹	۱۶۰	۲۰۰	۲۱۵	۱۹۰	1184
محاسبه دستی تعداد همخوانی کامل	(%۴۴)	(%۳۹)	(%۴۲)	(%۳۳)	(%۴۹)	(%۴۷)	(%۴۱)
	۱۸۳	۱۹۳	۱۲۵	۱۸۳	۱۹۹	۱۶۹	۱۶۹

جدول شماره ۱۵ به طور کلی ۴ درصد اختلاف را در محاسبه ماشینی و محاسبه دستی مربوط به میزان همخوانی کامل توصیفگرها و کلید واژه های عنوان نشان می دهد. علت اصلی این اختلاف نیز عدم درج علامت اسلش " / " در خصوص عناوینی بود که تنها دارای یک توصیفگر بودند. علاوه بر این با استفاده از آزمون مجذور کای^۱ نیز اختلاف بین محاسبه دستی و محاسبه ماشینی تعداد همخوانی کامل ارزیابی شد که نتایج به دست آمده در جدول زیر مشاهده می شود:

جدول شماره ۱۶- آزمون مجذور کای برای سنجش تفاوت محاسبه تعداد توصیفگرها به شیوه دستی و ماشینی

	VAR00001	VAR00002
Chi-Square(a,b)	.667	.000
df	4	5
Asymp. Sig.	.955	1.000

داده های این جدول نشان می دهد که در خصوص محاسبه تعداد توصیفگرها، تفاوت معنی داری بین شیوه دستی و ماشینی وجود ندارد. با استفاده از همین آزمون موارد محاسبه تعداد همخوانی کامل در شیوه دستی و ماشینی نیز با یکدیگر مقایسه شد که نتایج به دست آمده در جدول شماره ۱۷ نشان می دهد که در خصوص محاسبه همخوانی کامل به شیوه دستی و ماشینی تفاوت معنی داری وجود ندارد.

جدول شماره ۱۷- آزمون مجذور کای برای سنجش تفاوت محاسبه تعداد همخوانی کامل، به شیوه دستی و ماشینی

	VAR00001	VAR00002
Chi-Square(a,b)	.000	.667
df	5	4
Asymp. Sig.	1.000	.955

بدین ترتیب با اختلاف یا خطای کمی، به شیوه ماشینی محاسبه همخوانی کامل کلید واژه های عنوان با توصیفگرها، در کل جامعه تحقیق امکان پذیر بود. با استفاده از شیوه محاسبه یاد شده، کل جامعه تحقیق به تفکیک حوزه بررسی شد که نتایج به دست آمده در جدول شماره ۱۸ مشاهده می شود.

جدول شماره ۱۸- محاسبه ماشینی تعداد توصیفگرها و همخوانی کامل آنها با کلید واژه های عنوان در کل جامعه تحقیق.

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد عناوین جامعه تحقیق	۱۰۴۴۰	۴۲۹۰	۲۹۹۳۵	۵۲۶۷	۴۷۲۹	۱۰۹۰	۵۵۷۵۱
شمارش ماشینی توصیفگرهای جامعه تحقیق	۵۱۶۹۵	۲۲۲۲۰	۸۶۳۳۷	۲۹۶۱۶	۱۸۹۳۸	۳۸۶۲	۲۱۲۶۶۸
تعداد همخوانی کامل	(۳۸٪)	(۳۴٪)	(۴۰٪)	(۳۴٪)	(۶۱٪)	(۴۲٪)	(۴۰٪)
	۱۹۵۸۶	۷۴۵۴	۳۴۸۸۹	۹۹۸۲	۱۱۵۲۶	۱۶۲۵	۸۵۰۶۲

داده های این جدول تاییدی است بر یافته های حاصل از بررسی نمونه تحقیق که در جدول شماره ۱۴ درج شده است. همان گونه که مشاهده می شود در این بررسی نیز ۴۰ درصد توصیفگرهای نمایه سازان را کلید واژه های عنوان تشکیل می دهد و همانند نتایج حاصل از بررسی غیر ماشینی نمونه تحقیق، پایان نامه های حوزه علوم کشاورزی (۶۱٪)، هنر (۴۲٪)، علوم پزشکی (۴۰٪) و علوم انسانی (۳۸٪) دارای بیشترین همخوانی کامل بین کلید واژه های عنوان با توصیفگرها بوده و پایان نامه های حوزه فنی و مهندسی و علوم پایه هر دو با (۳۴٪) کمترین میزان همخوانی را دارند.

۴-۶- میزان همخوانی کلید واژه های عنوان با توصیفگرهای نمایه سازان در سال های مختلف چه تفاوتی دارد؟

به منظور پاسخگویی به این سوال ستون مربوط به تاریخ تهیه پایان نامه ها در محیط نرم افزاری Excel از کوچک به بزرگ مرتب شده سپس نتایج حاصل از شمارش پایان نامه های نگارش یافته در سال های مختلف

در جدول شماره ۱۹ گرد آمد. در این جدول پراکندگی پایان نامه های مورد بررسی بر حسب سال های مختلف نشان داده شده است.

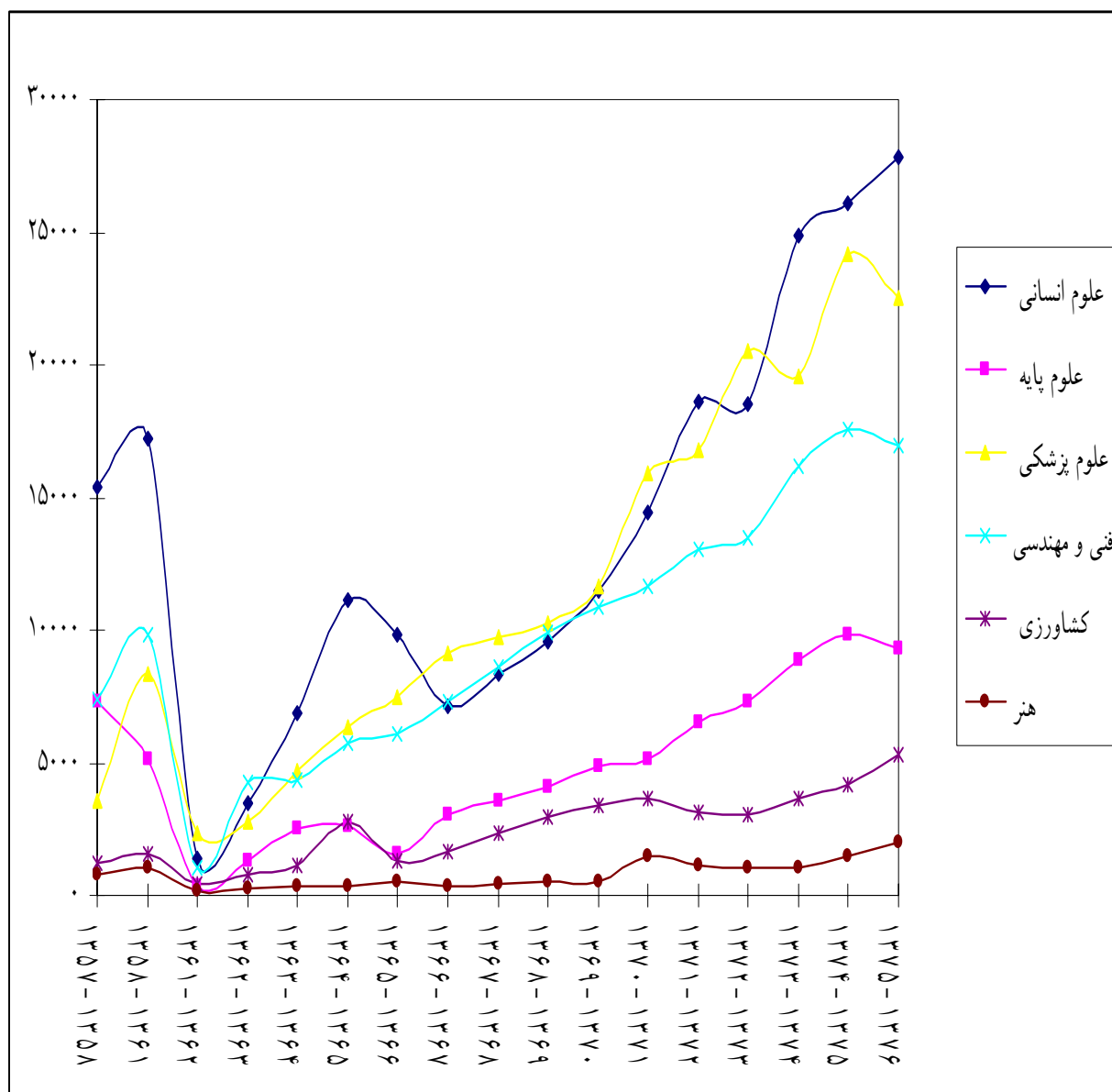
جدول شماره ۱۹- توزیع فراوانی پایان نامه ها بر حسب سال نگارش ، به تفکیک حوزه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
سال نگارش	تعداد	تعداد	تعداد	تعداد	تعداد	تعداد	تعداد
۱۳۱۴-۱۳۲۰	۰	۰	۲	۰	۲	۰	۴ (۱٪)
۱۳۲۱-۱۳۲۷	۰	۰	۳	۰	۳	۰	۶ (۱٪)
۱۳۲۸-۱۳۳۴	۰	۰	۶	۰	۲	۰	۸ (۱٪)
۱۳۳۵-۱۳۴۱	۱	۱	۱۰	۰	۱۱	۰	۲۳ (۴٪)
۱۳۴۲-۱۳۴۸	۳	۹	۱۱	۱	۱۹	۲۰	۶۳ (۱۱٪)
۱۳۴۹-۱۳۵۵	۲۲	۵	۱۸	۳	۸	۲۱	۷۷ (۱۳٪)
۱۳۵۶-۱۳۶۲	۱۴	۲	۲۴	۰	۷	۲۲	۶۹ (۱۲٪)
۱۳۶۳-۱۳۶۹	۱۲	۶	۱۵	۱۶	۱۱	۱۷	۷۷ (۱۳٪)
۱۳۷۰-۱۳۷۶	۴۳	۵۶	۱۰	۵۸	۳۰	۱۳	۲۱۰ (۳۵٪)
۱۳۷۷-۱۳۸۱	۵	۲۱	۱	۲۲	۷	۷	۶۳ (۱۱٪)
جمع	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰۰ (۱۰۰٪)

”وزارت بهداشت درمان و آموزش پزشکی” متصدی گردآوری پایان نامه های حوزه علوم پزشکی است، باعث شد که از سال ۱۳۶۸ به بعد تلاشی در خصوص گردآوری پایان نامه های این حوزه صورت نگیرد، هر چند برخی از دانشجویان یا دانشکده های پزشکی به منظور اشاعه اطلاعات مربوط به پایان نامه های خود نسخه هایی را به مرکز ارسال داشته اند. تا قبل از سال ۱۳۶۳ پایان نامه های حوزه علوم پایه و فنی و مهندسی در مقایسه با دیگر حوزه ها کمترین میزان حضور در پایگاه را داشته و پس از آن روند رو به رشدی داشته اند. ملزم شدن دانشجویان برخی از دانشکده های فنی و علوم پایه (مانند دانشکده فنی و دانشکده علوم پایه دانشگاه تهران و دانشگاه تربیت مدرس) مبنی بر ارائه یک نسخه از پایان نامه خود به مرکز اطلاعات و مدارک علمی ایران را می توان عاملی در راستای افزایش حضور پایان نامه های حوزه فنی و مهندسی و علوم پایه در این پایگاه دانست. گردآوری پایان نامه های حوزه علوم انسانی و علوم کشاورزی از فراز و نشیب بیشتری در مقایسه با دیگر حوزه ها برخوردار بوده و پایان نامه های حوزه هنر دارای روند گردآوری یکنواختی هستند. از سال ۱۳۷۱ به بعد عواملی چون مصوبه ستاد تحصیلات تکمیلی و بخشنامه وزیر فرهنگ و آموزش عالی مبنی بر ارسال اطلاعات مربوط به پایان نامه ها، به مرکز اطلاعات و مدارک علمی ایران؛ همچنین انتصاب رئیس جدید مرکز اطلاعات و مدارک علمی ایران (جناب آقای دکتر غریبی)، افزایش بودجه مرکز، امکانات، نیروی انسانی و انتشار فصلنامه ”چکیده پایان نامه های ایران” از بهار ۱۳۷۲؛ به غیر از حوزه های علوم پزشکی و هنر، موجب رشد چشمگیر گردآوری پایان نامه ها در حوزه های دیگر شد. متأسفانه این روند صعودی نه تنها ادامه پیدا نکرد بلکه از سال ۱۳۷۶ به بعد، دچار افت محسوسی نیز شد.

پس از انقلاب فرهنگی، از سال ۱۳۶۳ به بعد در کلیه حوزه ها روند رو به رشدی در خصوص تعداد فارغ التحصیلان مقاطع تحصیلات تکمیلی مشاهده می شود (نمودار شماره ۴).^۱

^۱ وزارت فرهنگ و آموزش عالی، ۱۳۷۷، ۴۰.



نمودار شماره ۴ فارغ التحصیلان دانشگاه ها و مؤسسات آموزش عالی از سال ۱۳۵۷ تا سال ۱۳۷۶

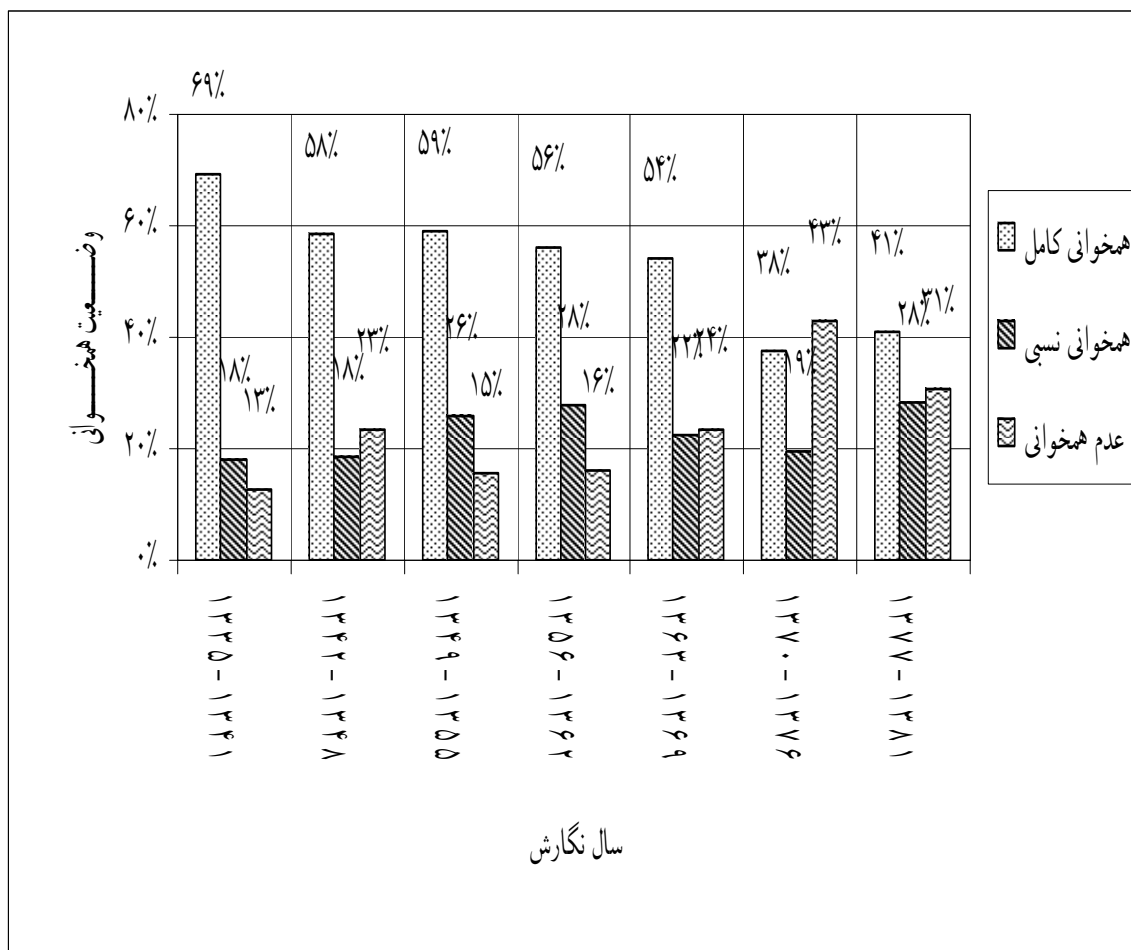
مقایسه نمودار شماره ۴ و نمودار شماره ۳ نشان می دهد که روند گردآوری پایان نامه ها در مرکز هیچ گونه همخوانی با روند تهیه پایان نامه ها در کشور ندارد. تصمیم گیری های مقطعی، فقدان کارشناسان حرفه ای، عدم همخوانی تعداد رکوردهای فصلنامه "چکیده پایان نامه های ایران" با تعداد پایان نامه هایی که در کشور تهیه می شود و نیز عدم توجه یا تمایل مراکز تحصیلات تکمیلی در ارسال پایان نامه های خود به این مرکز از جمله دلایل کاهش روند تهیه پایان نامه ها به شمار می روند. به عنوان مثال هر چند گردآوری پایان نامه های حوزه علوم پزشکی از دستور کار مرکز خارج شده بود ولی در سال جاری (۱۳۸۲) نزدیک به ۷۰۰۰ پایان نامه از دانشگاه علوم پزشکی دانشگاه تهران به مرکز ارائه شد که مراحل مقدماتی نمایه سازی را می گذرانند، علاوه بر این تصمیم یک رئیس دانشکده یا دانشگاه برای ارسال پایان نامه به مرکز باعث افزایش آمار گردآوری پایان نامه در حوزه یا مقطع زمانی خاص خواهد شد.

در ادامه وضعیت همخوانی کلید واژه های عنوان با توصیفگرها در سال های مختلف بررسی شد که نتایج حاصل از آن در جدول شماره ۲۰ درج شده است. در این بررسی با توجه به این که مجموع پایان نامه های حضور یافته در نمونه تحقیق، مربوط به سال های قبل از ۱۳۳۵ کمتر از ۵۰ رکورد بود، پایان نامه های مربوط به این سال ها از نمونه تحقیق حذف شدند.

جدول شماره ۲۰- توزیع فراوانی همخوانی کلید واژه های عنوان با توصیفگرها به تفکیک سال نگارش

سال نگارش همخوانی	۱۳۳۵- ۱۳۴۱	۱۳۴۲- ۱۳۴۸	۱۳۴۹- ۱۳۵۵	۱۳۵۶- ۱۳۶۲	۱۳۶۳- ۱۳۶۹	۱۳۷۰- ۱۳۷۶	۱۳۷۷- ۱۳۸۱
همخوانی کامل	۳۸ (٪۶۹)	۱۰۸ (٪۵۸)	۱۲۶ (٪۵۹)	۱۱۲ (٪۵۶)	۱۴۰ (٪۵۴)	۴۹۸ (٪۳۸)	۱۳۴ (٪۴۱)
همخوانی نسبی	۱۰ (٪۱۸)	۳۴ (٪۱۸)	۵۵ (٪۲۶)	۵۶ (٪۲۸)	۵۸ (٪۲۲)	۲۵۷ (٪۱۹)	۹۲ (٪۲۸)
عدم همخوانی	۷ (٪۱۳)	۴۳ (٪۲۳)	۳۳ (٪۱۵)	۳۲ (٪۱۶)	۶۱ (٪۲۴)	۵۶۷ (٪۴۳)	۱۰۱ (٪۳۱)
جمع	۵۵	۱۸۵	۲۱۴	۲۰۰	۲۵۹	۱۳۲۲	۳۲۷

در نمودار شماره ۵ به شکل دیگری روند کاهش میزان همخوانی کامل، افزایش همخوانی نسبی و نیز افزایش عدم همخوانی کلید واژه های عنوان با توصیفگرها نشان داده شده است.



نمودار شماره ۵ - مقایسه موارد همخوانی و عدم همخوانی توصیفگرها و کلید واژه های عنوان به تفکیک سال

این نمودار نشان می دهد که در سال های گذشته میزان همخوانی کامل کلید واژه های عنوان، بیش از سال های اخیر بوده و با گذشت زمان درصد همخوانی کامل کاهش یافته است. تدوین سر عنوان های موضوعی و اصطلاحنامه های عمومی و تخصصی و نیز به کارگیری نمایه سازان آموزش دیده از جمله دلایل کاهش وابستگی به کلید واژه های عنوان و بهره گیری از اصطلاحنامه ها به شمار می روند.

در ادامه این بررسی از منظر آزمون مجذور کای نیز تفاوت موارد همخوانی کامل، همخوانی نسبی و عدم همخوانی در سال های مختلف سنجیده شد که نتایج به دست آمده در ذیل آمده است.

جدول شماره ۲۱- نتیجه آزمون مجذور کای در خصوص تفاوت موارد همخوانی در سال های مختلف

محاسبه مجذور کای	Asymp.Sig	df.	Chi.
همخوانی کامل	824.637	6	.000
همخوانی نسبی	500.381	6	.000
عدم همخوانی	1971.111	6	.000

داده های این جدول نشان می دهد که موارد همخوانی کامل، همخوانی نسبی و عدم همخوانی در سال های مختلف دارای اختلاف معنی داری است.

۴-۷- آیا با افزایش تعداد کلید واژه های عنوان، میزان همخوانی آنها با توصیفگرها نیز افزایش می یابد؟

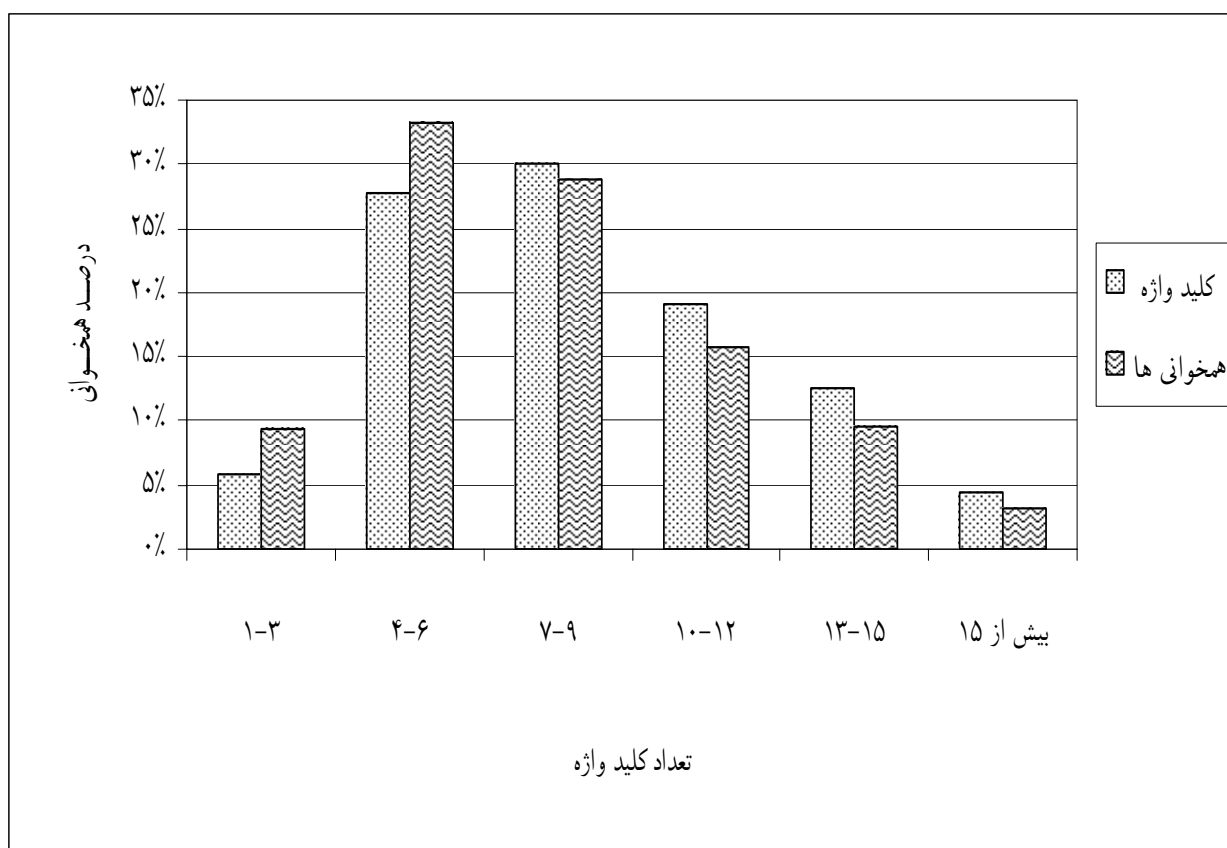
برای پاسخگویی به این سوال در ۶ طبقه توزیع فراوانی کلید واژه های عنوان و موارد همخوانی (اعم از همخوانی کامل و نسبی) مشخص شد تا رابطه این دو متغیر با یکدیگر سنجیده شود. نتایج حاصل از این بررسی در جدول شماره ۲۲ آمده است.

جدول شماره ۲۲- توزیع فراوانی تعداد کلید واژه و همخوانی ها (همخوانی کامل و نسبی) به تفکیک حوزه.

حوزه	علوم انسانی		علوم پایه		علوم پزشکی		فنی و مهندسی		کشاورزی		هنر		جمع	
کلید واژه	همخوانی ه	کلید واژه	همخوانی ه	کلید واژه	همخوانی ه	کلید واژه	همخوانی ه	کلید واژه	همخوانی ه	کلید واژه	همخوانی ه	کلید واژه	همخوانی ه	کلید واژه
1-3	(%4)	(%7)	(%2)	(%3)	(%16)	(%24)	(%1)	(%2)	(%4)	(%6)	(%13)	(%21)	(%6)	(%9)
	32	22	18	10	82	52	7	7	28	17	72	58	239	166
4-6	(%28)	(%35)	(%26)	(%32)	(%37)	(%41)	(%23)	(%27)	(%21)	(%24)	(%40)	(%44)	(%28)	(%33)
	205	106	188	104	190	90	184	90	160	73	212	120	1139	583
7-9	(%26)	(%24)	(%30)	(%31)	(%26)	(%21)	(%36)	(%35)	(%33)	(%36)	(%27)	(%23)	(%30)	(%29)
	191	72	220	103	133	47	289	116	256	108	142	61	1231	507
10-12	(%16)	(%13)	(%20)	(%16)	(%16)	(%12)	(%25)	(%24)	(%18)	(%16)	(%17)	(%10)	(%19)	(%16)
	122	40	149	54	84	26	203	80	137	47	89	28	784	275
13-15	(%17)	(%14)	(%19)	(%15)	(%5)	(%3)	(%10)	(%8)	(%19)	(%15)	(%0)	(%0)	(%13)	(%10)
	123	41	136	51	26	6	83	26	147	45	0	0	515	169
بیش از 15	(%10)	(%7)	(%3)	(%2)	(%0)	(%0)	(%5)	(%3)	(%5)	(%4)	(%4)	(%1)	(%4)	(%3)
	71	21	19	8	0	0	38	11	37	12	19	4	184	56
جمع	744	302	730	330	515	221	804	330	765	302	534	271	4092	1756

داده های این جدول نشان می دهد که در کلیه حوزه ها، تجمع فراوانی همخوانی ها در مورد پایان نامه هایی که کمتر از ۷-۹ کلید واژه در عنوان دارند بیش از مواردی است که پایان نامه ها، دارای ۷-۹ کلید واژه در عنوان هستند. به عبارت دیگر در کلیه حوزه ها با افزایش کلید واژه های عنوان به بیش از حد متوسط به همان

نسبت منجر به افزایش همخوانی نخواهد شد و افزایش تعداد کلید واژه های عنوان میزان همخوانی آنها با توصیفگرها را افزایش نمی دهد. این حالت با احتساب کلیه حوزه ها در نمودار شماره ۶ نشان داده شده است.



نمودار شماره ۶ - توزیع فراوانی کلید واژه های عنوان و همخوانی ها.

در عین حال بر اساس آزمون ضریب همبستگی پیرسون نیز ارتباط تعداد کلید واژه های عنوان و موارد همخوانی کامل مورد ارزیابی قرار گرفت که نتایج به دست آمده در جدول شماره ۲۳ مشاهده می شود.

احتمال عدم همخوانی واژه های از پیش تعریف شده با واژه های کاربر، گند بودن فرایند تغییر واژه ها و فرایند نمایه سازی و اختلال در بازیابی به واسطه وجود فاصله خالی اضافی بین واژه ها، از جمله معایب این شیوه نمایه سازی به شمار می روند. اکبری نژاد (۱۳۷۶، ۴۹) در مقاله ای فاصله خالی بین واژه های ترکیبی را در پایگاه های اطلاعاتی به زبان فارسی مورد توجه قرار داده و به مسائل ناشی از آن در ذخیره و بازیابی اطلاعات اشاره می کند. بدین ترتیب در پایگاه های اطلاعاتی توجه به واژه های ترکیبی در ذخیره و بازیابی اطلاعات، امری ضروری به شمار می رود. در ادامه این تحقیق سعی شد تا تصویری از وضعیت توصیفگرهای تک واژه و ترکیب واژه به دست داده شود. نتایج حاصل از این بررسی در جدول شماره ۲۴ آمده است.

جدول شماره ۲۴- نسبت تعداد توصیفگرهای تک واژه به تعداد واژه های ترکیبی، به تفکیک حوزه.

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
مجموع توصیفگرها	۴۲۷	۵۰۵	۳۲۰	۵۶۵	۴۲۷	۳۶۵	2609
تعداد توصیفگر تک واژه	(۴۶٪)	(۴۲٪)	(۴۷٪)	(۴۲٪)	(۵۹٪)	(۵۹٪)	(48٪)
	۱۹۶	۲۱۱	۱۵۰	۲۳۶	۲۵۴	۲۱۴	1261
تعداد توصیفگرهای ترکیبی	(۵۴٪)	(۵۸٪)	(۵۳٪)	(۵۸٪)	(۴۱٪)	(۴۱٪)	(52٪)
	۲۳۱	۲۹۴	۱۷۰	۳۲۹	۱۷۳	۱۵۱	1348

در جدول شماره ۲۴ نشان داده شده که به طور کلی نسبت تقریباً یکسانی در اختصاص تک واژه ها و ترکیب واژه ها رعایت شده است. در عین حال حوزه های فنی و مهندسی و علوم پایه (هر دو با ۵۸ درصد) دارای تعداد ترکیب واژه های بیشتری در مقایسه با دیگر حوزه ها از جمله حوزه کشاورزی و هنر (هر دو با ۴۱ درصد) هستند. یکی از دلایل این اختلاف احتمالاً به ساختمان زبان شناختی موضوعات مربوط می شود که حاصل اختلاف ساختمان اطلاعاتی آنها است. مثلاً زبان یک شیمیدان در زمینه شیمی آلی کاملاً با زبان مورد استفاده یک جانور شناس متفاوت است زیرا اطلاعاتی که از نظر یک متخصص شیمی حائز اهمیت است نام

ترکیبات آلی، آرایش ساختمانی و خواص فیزیکی و شیمیایی است و حال آنکه از دیدگاه یک جانورشناس توصیف و تعیین جنس و نوع حیوانات مطرح است (بهاتاچاریا، ۱۳۵۶، ۱۴). از طرف دیگر با توجه به این که مفاهیم بازنمودهای اشیاء منفرد هستند (هلموت فلبر، ۱۳۸۱، ۱۵۵) و محققین حوزه های فنی و مهندسی و علوم پایه معمولاً با ترکیب اشیاء، مواد و موضوعات مختلف سروکار دارند، به ناگزیر عبارات نیز به عنوان نمادهای زبانی این اشیاء به شکل ترکیبی ظاهر می شوند.

یکی دیگر از مواردی که در خصوص شیوه نگارش توصیفگرها مورد بررسی قرار گرفت شکل "مفرد" یا "جمع" آنها است. معمولاً مساله جمع و مفرد در سرعنوانهای موضوعی همه زبان ها مشکلاتی را ایجاد می کند. همیشه این سوال پیش می آید که چه چیزهایی را جمع به کار بریم و چه چیزهایی را مفرد (پوری سلطانی، ۱۳۷۳، ۲۹). در زبان انگلیسی معمولاً استفاده از شکل جمع کلمات (تیلور، ۱۳۸۱، ۱۸۷)؛ (Rydings, 1999) و در زبان فارسی استفاده از شکل مفرد در نمایه سازی پیشنهاد می شود (سمایی، ۱۳۷۹، ۳۱). در تحقیق حاضر کلماتی که با "ان"، "ها"، "ات"، "جات" و کلماتی که طبق قاعده زبان عربی به شکل مکسر جمع بسته شده بودند بررسی شد که نتایج حاصل از آن برابر داده های جدول شماره ۲۵ است.

جدول شماره ۲۵- نسبت توصیفگرهای مفرد به توصیفگرهای جمع، به تفکیک حوزه.

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
مجموع توصیفگرها	۴۲۷	۵۰۵	۳۲۰	۵۶۵	۴۲۷	۳۶۵	2609
تعداد توصیفگر مفرد	۳۵۶ (٪۸۳)	۴۴۰ (٪۸۷)	۲۶۴ (٪۸۳)	۵۲۲ (٪۹۲)	۳۸۶ (٪۹۰)	۳۱۰ (٪۸۵)	2278 (87%)
تعداد توصیفگر جمع	۷۱ (٪۱۷)	۶۵ (٪۱۳)	۵۶ (٪۱۸)	۴۳ (٪۸)	۴۱ (٪۱۰)	۵۵ (٪۱۵)	331 (13%)

به طور کلی ۸۷ درصد توصیفگرها به شکل مفرد و ۱۳ درصد به شکل جمع به کار رفته اند. در ادامه به منظور این که میزان استفاده از کلید واژه های مفرد و جمع در عنوان پایان نامه ها مشخص شده و مقایسه ای با توصیفگرهای مفرد و جمع در حوزه های شش گانه به دست داده شود، کلید واژه های عنوان پایان نامه ها از دیدگاه مفرد و جمع نیز مورد بررسی قرار گرفت که نتایج حاصله در جدول شماره ۲۶ آمده است.

جدول شماره ۲۶- نسبت کلید واژه های مفرد به کلید واژه های جمع ، به تفکیک حوزه.

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد کلید واژه	۷۴۴	۷۳۰	۵۱۵	۸۰۴	۷۶۵	۵۳۴	۴۰۹۲
تعداد کلید واژه مفرد	(۸۲٪)	(۸۲٪)	(۸۲٪)	(۸۴٪)	(۸۷٪)	(۸۸٪)	(۸۴٪)
	۶۱۳	۶۰۰	۴۲۳	۶۷۹	۶۶۶	۴۷۰	۳۴۵۱
تعداد کلید واژه جمع	(۱۸٪)	(۱۸٪)	(۱۸٪)	(۱۶٪)	(۱۳٪)	(۱۲٪)	(۱۶٪)
	۱۳۱	۱۳۱	۹۲	۱۲۶	۹۸	۶۲	۶۴۰

همان گونه که مشاهده می شود همانند توصیفگر ها، به طور کلی تعداد کلید واژه های مفرد (۸۴٪) بیش از کلید واژه های جمع (۱۶٪) است. بدین ترتیب کلید واژه ها و توصیفگرها، از نظر حالت های مفرد و جمع بودن نیز در وضعیت مشابهی هستند.

۴-۹- در "پایگاه چکیده پایان نامه های ایران"، بین میزان بازیافت با استفاده از کلید واژه های عنوان و توصیفگرهای نمایه سازان چه تفاوتی وجود دارد؟

برای دریافت اطلاعات از یک نظام، استفاده کننده بایستی با مکانیسم های مربوط به نظام نمایه ارتباط برقرار کند (ویسمان، ۱۳۷۳، ۱۰۶). هر چه این ارتباط سهل تر و عاری از پیچیدگی باشد و بین زبان متقاضی اطلاعات، و زبان نمایه ساز، هماهنگی بیشتری پدید آید، ضریب بازیافت و دقت بالاتر خواهد بود. پیوستگی بین

نظام نمایه سازی و بازیابی اطلاعات به حدی است که اساساً یکی از معیارهای ارزشیابی فعالیت های نمایه سازی مدارک را، بازیابی جامع اطلاعات مورد نظر در جستجو محسوب نموده اند (پائو، ۱۳۷۸، ۱۱۵).

مورتیمر تاب (Mortimer taube) یکی از نمایه سازانی است که نقش به سزایی در روش های جدید ذخیره و بازیابی اطلاعات دارد. به خاطر سهولت ظاهری روش تک واژه ای، تأثیر تاب در حوزه بازیابی اطلاعات بیش از تاثیر باتن (Batten)، کوردونیه (Cordonnier) یا مورز (Mooers) بوده است. باتن، نظام خود را بر پایه واژگان کنترل شده به شکل طرح رده بندی استوار ساخت. مورز اهمیت واژگانی را که توصیفگرهای آن بسیار محدود و به دقت انتخاب شده باشند، مورد تاکید قرار می دهد. با این وجود، تاب حداقل در نوشته های اولیه خود پیرامون تک واژه، واژگان کنترل شده را مورد تاکید قرار نداد. این عدم تاکید بر کنترل واژگان سر و صدای فراوانی را به وجود آورد و به صورت یک اسطوره درآمد، بدین معنی که نظام تک واژه ای، فرایند نمایه سازی موضوعی را از یک کار تقریباً تخصصی و فکری به یک عمل دفتری عادی به کلی عوض کرد. اگر نمایه سازی موضوعی چیزی بیش از برچسب زدن مدارک با اصطلاحات تک واژه ای نیست، آنگاه این فرایند به طور قطع می تواند به وسیله کارکنان دفتری که تعدادی کلید واژه را از متن مدارک استخراج می کنند انجام گیرد. این اسطوره که در دهه ۱۹۵۰ رایج و متداول بود، فرایند نمایه سازی موضوعی را مورد بحث قرار داد و تاثیر و کارایی بسیاری از نمایه هایی که به تازگی ایجاد شده بودند را تضعیف کرد (لانکاستر، ۱۳۷۹، ۴۷). این روش به شکل اولیه خود فقط مدت کوتاهی استمرار داشت. نخستین تغییر، پذیرش تک واژه هایی بود که بیش از یک واژه مفرد را شامل می گردید. برای مثال، اگر سازمانی درباره "آموزش کتابداری" مجموعه عظیمی داشته باشد حفظ "آموزش کتابداری" به عنوان پیش همارایی صحیح به نظر می رسد تا این که بیائیم اصطلاح "آموزش" و "کتابداری" را به صورت جداگانه به کار ببریم. وقتی که حجم مجموعه رشد پیدا می کند مسائلی که در نبود واژگان کنترل شده ایجاد می شود بی درنگ ظاهر می گردد. نمایه سازان مختلف، یاحتی یک نمایه ساز در زمان های گوناگون، برای نشان دادن یک مفهوم کاملاً شبیه هم، از اصطلاحات متفاوتی استفاده می کنند. بدین ترتیب جستجوگر مجبور است از شیوه های گوناگونی برای نشان دادن یک موضوع خاص استفاده کند تا یک

جستجوی معقول و کاملی را انجام دهد. به زودی در نمایه سازی انتخاب یک اصطلاح از میان اصطلاح های مترادف و قرار دادن ارجاعات See مورد پذیرش قرار گرفت. مسأله دیگر در پرتو امکان همارایی کاذب^۱ ایجاد گردید. فرض کنید در جستجوی مدارکی درباره "فلسفه علم" دو اصطلاح "علم" و "فلسفه" را همارایی می کنیم و مدرکی را با عنوان "علم فلسفه" بازیابی می کنیم که مدرکی ناخواسته است. نخستین گام برای اجتناب از این نوع بازیابی ناخواسته استفاده از توصیفگرها بود، که با توسعه فهرست مستندی از توصیفگرهای تایید شده و تعریف روابط بین توصیفگرها "سر عنوان های موضوعی" و "اصطلاحنامه ها" پدید آمدند.

استفاده از کلید واژه ها و توصیفگرها در نمایه سازی هر یک اثر خاصی بر نتایج بازیابی خواهد داشت. در هر پایگاه اطلاعاتی مدارکی وجود دارد که تعدادی از آنها مرتبط با موضوع مورد جستجو و بسیاری نیز نامرتب هستند. از میان مدارک مرتبط نیز برخی با نیاز جستجوگر مغایرت دارند. توانایی نظام در بازیابی مدارک مرتبط را جامعیت و توانایی نظام در کنار گذاشتن مدارک نامرتب را مانعیت گویند که هر یک به شیوه زیر محاسبه می شود (شیری، ۱۳۷۲، ۳۸):

$$\text{میزان جامعیت} = \frac{\text{تعداد مدارک مرتبط بازیابی شده}}{\text{تعداد کل مدارک مرتبط در سیستم}}$$

$$\text{میزان مانعیت} = \frac{\text{تعداد مدارک مرتبط بازیابی شده}}{\text{تعداد کل مدارک بازیابی شده}}$$

در ادامه بررسی حاضر، به منظور این که پایگاه "چکیده پایان نامه های ایران" از منظر جامعیت نیز ارزیابی شود، طی سه مرحله موارد بازیافت از طریق کلید واژه ها و توصیفگرها مقایسه شد. در مرحله نخست جستجو از طریق کلید واژه ها و توصیفگرهایی که دارای همخوانی نسبی بودند صورت گرفت (مواردی چون:

کلید واژه "شیعه"؛ با توصیفگر "تشیع" در یک رکورد، یا کلید واژه "هیئت علمی" با توصیفگر "هیات علمی" در یک رکورد). با توجه به این نکته، تعداد رکوردهایی که دارای موارد همخوانی نسبی بودند در جدول زیر آمده است.

جدول شماره ۲۷- تعداد رکوردهای دارای موارد همخوانی نسبی.

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد رکورد ها با موارد همخوانی نسبی	۵۱	۵۱	۵۴	۵۰	۴۷	۳۹	292

پس از انتخاب کلید واژه ها و توصیفگرهای یاد شده، از طریق هر یک عملیات جستجو در پایگاه "چکیده پایان نامه های ایران" صورت گرفت که نتایج حاصل از آن در جدول شماره ۲۸ آمده است.

جدول شماره ۲۸ - تعداد بازیافت از طریق کلید واژه ها و توصیفگرهای دارای همخوانی نسبی، به تفکیک حوزه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد بازیابی با استفاده از کلید واژه	(۴۷٪) ۶۳۶۱	(۳۹٪) ۲۱۶۲	(۴۴٪) ۷۲۶۳	(۳۷٪) ۲۶۳۳	(۳۸٪) ۲۳۴۸	(۴۶٪) ۱۳۱۹۱	(۴۴٪) ۳۳۹۵۸
تعداد بازیابی با استفاده از توصیفگر	(۵۳٪) ۷۲۷۶	(۶۱٪) ۳۳۲۶	(۵۶٪) ۹۰۷۸	(۶۳٪) ۴۵۴۰	(۶۲٪) ۳۸۵۹	(۵۴٪) ۱۵۶۸۵	(۵۶٪) ۴۳۷۶۴
جمع بازیافت	۱۳۶۳۷	۵۴۸۸	۱۶۳۴۱	۷۱۷۳	۶۲۰۷	۲۸۸۷۶	۷۷۷۲۲

بدین ترتیب در پاسخ به سوال اخیر باید اذعان داشت که بازیابی با استفاده از توصیفگرها (۵۶٪) بیش از موارد بازیابی شده با استفاده از کلید واژه های عنوان (۴۴٪) است. به منظور محاسبه تفاوت معنی داری بازیافت با استفاده از کلید واژه های عنوان و توصیفگرها از آزمون مجذور کای استفاده شد که یافته حاصل از این محاسبه

(جدول شماره ۲۹) تفاوت معنی داری را در بازیافت از طریق کلید واژه های عنوان و توصیفگرهای نمایه سازان نشان می دهد.

جدول شماره ۲۹- آزمون مجذور کای در خصوص میزان بازیافت از طریق کلید واژه های عنوان و توصیفگرهای نمایه سازان

	VAR00001	VAR00002
Chi-Square(a)	.000	.000
df	5	5
Asymp. Sig.	1.000	1.000

این در حالی است که اکثر تحقیقات انجام شده میزان جامعیت از طریق واژگان آزاد و کنترل نشده را بیش از جامعیت بازیابی از طریق توصیفگرها و واژگان کنترل شده نشان می دهد. به عنوان مثال چشمه سهرابی (۱۳۷۸) در پایان نامه خود پس از بررسی ۷ پایگاه اطلاعاتی در ۶ گروه علوم انسانی، علوم پایه، علوم پزشکی، فنی و مهندسی، کشاورزی و هنر به این نتیجه رسید که میزان جامعیت نتایج بازیابی در پایگاه های اطلاعاتی به هنگام استفاده از اصطلاحنامه ۳۹ درصد و به هنگام استفاده از زبان طبیعی ۵۶ درصد است. مارتین (Martin) نیز در جریان تحقیقات خود به این نتیجه رسید که جامعیت از طریق واژگان آزاد (۷۸ درصد) بیش از واژگان کنترل شده (۵۶ درصد) است (شیری، ۱۳۷۲، ۳۸). دلیل بازیابی و جامعیت بیشتر از طریق توصیفگرها انجام طرح پژوهشی "ساماندهی توصیفگرهای پایگاه های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران (جستجوی هوشمند)"^۱ است که توسط واحد خدمات ماشینی مرکز در سال ۱۳۸۰ صورت پذیرفت. قبل از انجام این طرح به دلایلی چون فقدان دستورالعمل مشخص و مدون برای نمایه سازی، تغییر کارشناسان و سلیقه های مختلف در امر نمایه سازی، استفاده از اصطلاحنامه ها، واژگان و واژه نامه های مختلف و... باعث ناهمگونی و نابسامانی بسیار توصیفگرها شده بود و پس از هر جستجو رکوردهای مرتبط بسیاری بازیابی نمی شد. به عنوان مثال اگر پژوهشگری با توصیفگر "آب شناسی" به جستجو در پایگاه مرکز می پرداخت، موفق به بازیابی رکوردهایی که

^۱ پارسى اصفهانی، امیر. ۱۳۸۰. ساماندهی توصیفگرهای پایگاه های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران (جستجوی هوشمند). طرح پژوهشی. تهران: مرکز اطلاعات و مدارک علمی ایران.

ذیل توصیفگرهای "آب شناسی"، "آب شناسی"، "هیدرولوژی"، "هیدروژئولوژی"، "آب شناسی زمین" و "زمین شناسی آب" ذخیره شده بودند، نمی شد. علاوه بر این غلط های املایی، آیین های مختلف نگارش، وجود مترادف ها، واژه های ساده و مرکب، کلمات هم ریشه و ... عواملی بود که به این پراکندگی و نابسامانی می افزود. واحد خدمات ماشینی با توجه به این نکات و از طرق مختلف، مانند معادل لاتین توصیفگرهای فارسی و دسته بندی توصیفگرهای هم مفهوم و پس از یک سال کار گروهی با واحد اصطلاح شناسی مرکز اطلاعات و مدارک علمی ایران امکانی را فراهم آورد که اکنون جستجوگران می توانند از طریق یک توصیفگر، رکوردهایی را که ذیل توصیفگرهای هم مفهوم دیگر پراکنده شده اند را بازیابی کنند. به عنوان مثال با جستجو از طریق توصیفگر "اشعه" علاوه بر بازیابی رکوردهای ذیل این توصیفگر، رکوردهای سامان یافته ذیل توصیفگرهای "پرتو تابشی"، "پرتوها"، "اشعه (پرتو)"، "اشعه ها"، "تابش ها"، "تشعشع" و "تشعشعات" نیز بازیابی می شود. بدین ترتیب انجام این طرح منجر به ایجاد همبستگی بین توصیفگرها، افزایش ضریب بازیافت در پایگاه های مرکز از جمله پایگاه چکیده پایان نامه های ایران شد. در صورتی که چنین طرحی نیز در ارتباط با کلید واژه های عنوان صورت پذیرد می توان جامعیت بازیابی بیشتری از پایگاه "چکیده پایان نامه های ایران" انتظار داشت.

در مرحله بعد بازیافت از طریق کلید واژه ها و توصیفگرهایی مقایسه شد که دارای همخوانی کامل بودند. در حوزه های مورد بررسی تعداد رکوردهایی که حداقل دارای یک کلید واژه و توصیفگر با همخوانی کامل بودند در جدول شماره ۳۰ ذیل آمده است :

جدول شماره ۳۰- تعداد رکوردها دارای حداقل یک مورد همخوانی کامل

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد رکورد ها با موارد همخوانی کامل	۹۳	۶۶	۸۴	۸۸	۹۳	۹۷	۵۲۱

پس از انتخاب کلید واژه ها و توصیفگرهایی که دارای موارد همخوانی کامل بودند از طریق هر یک جستجو در "پایگاه چکیده پایان نامه های ایران" صورت پذیرفت که مجموع نتایج به دست آمده در جدول زیر آمده است:

جدول شماره ۳۱- تعداد بازیافت از طریق کلید واژه ها و توصیفگرهایی که دارای همخوانی کامل، به تفکیک حوزه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد بازیابی با استفاده از کلید واژه	(/۵۰) ۵۱۹۹	(/۵۰) ۴۲۹۵	(/۵۰) ۹۴۵۲	(/۵۰) ۳۱۳۴	(/۵۰) ۵۸۴۹	(/۵۰) ۱۰۱۶۸	(/۵۰) ۳۸۰۹۷
تعداد بازیابی با استفاده از توصیفگر	(/۵۰) ۵۱۹۹	(/۵۰) ۴۲۹۵	(/۵۰) ۹۴۵۲	(/۵۰) ۳۱۳۴	(/۵۰) ۵۸۴۹	(/۵۰) ۱۰۱۶۸	(/۵۰) ۳۸۰۹۷
جمع بازیافت	۱۰۳۹۸	۸۵۹۰	۱۸۹۰۴	۶۲۶۸	۱۱۶۹۸	۲۰۳۳۶	۷۶۱۹۴

داده های این جدول نشان میدهد میزان بازیافت از طریق کلید واژه ها و توصیفگرهایی که دارای همخوانی کامل هستند تفاوتی ندارد، هر چند انتظار می رفت به دلایلی چون فاصله خالی اضافی بین کلید واژه

ها یا توصیفگرهای مرکب، میزان بازیافت بین کلید واژه ها و توصیفگرهایی که همخوانی کامل نیز تفاوت داشته باشد. مثلاً تعداد بازیافت از طریق کلید واژه "قارچکشاها" با تعداد بازیافت از طریق توصیفگر "قارچ کش ها" تفاوت کند. ولی در یکی از بخش های طرح "ساماندهی توصیفگرهای پایگاه های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران (جستجوی هوشمند)" امکاناتی در نظر گرفته شد که فاصله خالی اضافی بین واژه های مرکب حذف شود.

در مرحله پایان سنجش میزان بازیافت از کلید واژه ها و توصیفگرهایی صورت گرفت که در ظاهر هیچ گونه همخوانی با یکدیگر نداشتند ولی مترادف بودند. مانند کلید واژه "سوبسید" با توصیفگر "یارانه" یا کلید واژه "آبستنی" با توصیفگر "حاملگی". رکورد هایی که در حوزه های مورد بررسی دارای کلید واژه ها و توصیفگرهای مترادف بودند در جدول شماره ۳۲ درج شده است:

جدول شماره ۳۲- تعداد رکوردهای دارای کلید واژه و توصیفگر مترادف به تفکیک حوزه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد رکورد ها با موارد همخوانی کامل	۲۰	۱۸	۱۶	۲۷	۲۱	۱۵	۱۰۲

نتایج حاصل از بازیابی از طریق کلیدواژه ها و توصیفگرهای مترادف در جدول شماره ۳۳ آورده شده است.

جدول شماره ۳۳- تعداد بازیافت از طریق کلید واژه ها و توصیفگرهایی مترادف، به تفکیک حوزه

حوزه	علوم انسانی	علوم پایه	علوم پزشکی	فنی و مهندسی	کشاورزی	هنر	جمع
تعداد بازیابی با استفاده از کلید واژه	(/۴۴) ۱۱۸۴	(/۴۱) ۸۸۹	(/۴۷) ۸۷۷	(/۳۸) ۲۰۰۷	(/۴۲) ۱۷۲۳	(/۲۵) ۶۴۳	(/۳۹) ۷۳۲۵
تعداد بازیابی با استفاده از توصیفگر	(/۵۶) ۱۵۱۴	(/۵۹) ۱۲۸۸	(/۵۳) ۹۹۶	(/۶۲) ۳۳۱۲	(/۵۸) ۲۳۵۵	(/۷۵) ۱۹۷۱	(/۶۱) ۱۱۴۳۹
جمع بازیافت	۲۶۹۸	۲۱۷۷	۱۸۷۳	۵۳۱۹	۴۰۷۸	۲۶۱۴	۱۸۷۶۸

داده های این جدول نشان می دهد که بازیافت از طریق توصیفگرهای مترادف بیش از موارد بازیافت از طریق کلید واژه ها است. صورت پذیرفتن طرح "ساماندهی توصیفگرهای پایگاه های اطلاعاتی مرکز اطلاعات و مدارک علمی ایران (جستجوی هوشمند)"، اجتناب نمایه سازان از اختصاص واژه های مهجور و نامتداول (مانند: "کتابخانه همگانی" به جای "کتابخانه عمومی" یا "بندر پهلوی" به جای "بندر انزلی")، اجتناب از اختصاص معادل های لاتین (مانند: "آنالیز" به جای "تجزیه و تحلیل") و مانند آن را می توان از جمله دلایل ایجاد یکپارچگی بیشتر بین توصیفگرها و افزایش میزان بازیافت از طریق توصیفگرهای مترادف با کلید واژه ها دانست.

۴-۱۰- توصیفگرهای نمایه سازان تا چه حد با توصیفگرهای اصطلاحنامه های فارسی موجود همخوانی

دارد؟

سابقه تدوین اصطلاحنامه در ایران به چند سال اخیر باز می گردد و تعداد اصطلاحنامه هایی که تدوین شده انگشت شمار است. به منظور پاسخگویی به سوال اخیر به طور تصادفی از هر حوزه ۱۰۰ توصیفگر انتخاب شد. با توجه به این که در حوزه های علوم کشاورزی و علوم پایه اصطلاحنامه ای به زبان فارسی وجود نداشت

توصیفگرهای انتخاب شده مربوط به این دو حوزه حذف شد و ۴۰۰ توصیفگر دیگر از نظر میزان همخوانی با اصطلاحنامه ها زیر مورد بررسی قرار گرفت:

۱- اصطلاحنامه پزشکی فارسی / تالیف فاطمه رها دوست. - تهران: کتابخانه ملی جمهوری اسلامی ایران، ۱۳۷۶.

۲- اصطلاحنامه فرهنگی فارسی "اصفا" / تدوین فریبرز خسروی. - تهران: کتابخانه ملی جمهوری اسلامی ایران، ۱۳۸۰.

۳- اصطلاحنامه "نما" (نظام مبادله اطلاعات علمی و فنی) / مجری ملوک السادات بهشتی. - تهران مرکز اطلاعات و مدارک علمی ایران، ۱۳۷۵.

جدول شماره ۳۴- وضعیت همخوانی توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه های فارسی

حوزه		علوم انسا		علوم پزشکی		ف و مهندسی		هنر		ع
وضعیت خوا		تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد	درصد
خوا کامل		45	۴۵٪	34	۳۴٪	24	۲۴٪	46	۴۶٪	149 ۳۷٪
شی خوا		16	۱۶٪	20	۲۰٪	26	۲۶٪	19	۱۹٪	81 ۲۰٪
عدم خوا		39	۳۹٪	46	۴۶٪	50	۵۰٪	35	۳۵٪	170 ۴۳٪
جمع		۱۰۰	۱۰۰٪	۱۰۰	۱۰۰٪	۱۰۰	۱۰۰٪	۱۰۰	۱۰۰٪	۴۰۰ ۴۰۰٪

جدول فوق نشان می دهد به طور کلی ۳۷ درصد توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه های فارسی منطبق است که این رقم در حوزه علوم نظری مانند حوزه علوم انسانی و هنر بیش از دیگر حوزه ها است. فقدان اصطلاحنامه ها و نیز متخصصان موضوعی نمایه سازی در حوزه های فنی و مهندسی و علوم پزشکی را می توان از جمله دلایل همخوانی کمتر توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه ها دانست.

۴- ۱۱- توصیفگرهای نمایه سازان تا چه حد با رسم الخط فرهنگستان زبان و ادب فارسی همخوانی

دارد؟

به منظور پاسخگویی به این سوال نیز از هر حوزه ۱۰۰ توصیفگر به شکل تصادفی گزینش شده و هر یک با قواعد مندرج در کتابچه "دستور خط فارسی"^۱ منطبق داده شد که نتایج حاصل از این بررسی در جدول ذیل آمده است.

جدول شماره ۳۵- میزان همخوانی توصیفگرها با رسم الخط فارسی.

حوزه	علوم انسا	علوم پایه	علوم پزشکی	ف و مهندسي	کشاورزي	هنر	ع
وضعیت خوا	تعداد	تعداد	تعداد	تعداد	تعداد	تعداد	درصد
خوا با رسم ا ط	99	۹۹	۹۹	۹۸	۱۰۰	100	595 ۹۹%
عدم خوا با رسم ا ط	1	۱	۱	۲	۰	0	5 ۱%
جمع	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۶۰۰ ۱۰۰%

همان گونه که مشاهده می شود ۹۹ درصد توصیفگرهای اختصاص یافته توسط نمایه سازان با قواعد رسم الخط فارسی همخوان است و از این نظر مشکل خاصی در شکل نگارش توصیفگرها وجود ندارد و بیشتر رابطه موضوعی و مفهومی توصیفگرها با یکدیگر و پایان نامه ها است که نمایه سازان و کاربران را در امر ذخیره و بازیابی اطلاعات یاری می دهد.

^۱ فرهنگستان زبان و ادب فارسی. ۱۳۸۱. دستور خط فارسی- تهران: فرهنگستان زبان و ادب فارسی، نشر آثار.

فصل پنجم

نتیجه گیری و پیشنهادها

نتیجه گیری:

در این تحقیق اشاره شد که عنوان پایان نامه ها و کلید واژه های تشکیل دهنده آن از منظر خوانندگان و پژوهشگران دارای اهمیت است و ایشان را در اولین برخورد در تشخیص مدرک مربوط یا نامربوط یاری می دهد. از طرف دیگر نمایه سازان نیز از نقش و اهمیت عناوین در ساماندهی اطلاعات و غافل نبوده و از دیر باز تاکنون به عنوان یکی از مهم ترین ملاک های توصیف کننده مدارک در ایجاد پایگاه های اطلاعاتی از آن بهره جسته اند. با توجه به این نکته در تحقیق حاضر نخست ساختار و اجزاء تشکیل دهنده عنوان پایان نامه ها در حوزه های مختلف بررسی و این نتیجه حاصل شد که به طور کلی عنوان پایان نامه ها دارای ۷ کلید واژه است که می توان در ذخیره و بازیابی اطلاعات مربوط به پایان نامه ها از آن بهره جست هر چند در این خصوص بین تعداد متوسط کلید واژه های عنوان پایان نامه ها در حوزه های مختلف تفاوت معنی داری وجود دارد (جدول شماره ۴). همچنین نشان داده شد که در حوزه های مختلف توزیع فراوانی پایان نامه ها از نظر تعداد کلید واژه های عنوان یکسان نیست. به عنوان مثال در حوزه های هنر و علوم پزشکی بر خلاف حوزه های فنی و مهندسی و علوم پایه بیشتر پایان نامه ها دارای ۱ تا ۵ کلید واژه هستند. از طرف دیگر به طور متوسط در هر عنوان پایان نامه ۲ تا ۳ واژه بدون بار معنایی وجود دارد که به منظور جلوگیری از حجیم شدن پایگاه اطلاعاتی و فراهم آوردن زمینه نمایه سازی ماشینی باید انواع آنها شناسایی و به عنوان مجموعه ای از واژگان بدون بار معنایی در امر ذخیره و بازیابی اطلاعات مورد توجه قرار گیرند (جدول شماره ۷). ۴۵ درصد همخوانی کامل بین کلید واژه های عنوان و توصیفگرهای نمایه سازان یا به عبارت دیگر اتخاذ نیمی از کلید واژه های عنوان، به منزله توصیفگر و به طور کلی وابستگی ۶۷ درصدی توصیفگرهای نمایه سازان (همخوانی کامل و نسبی) به کلید واژه های عنوان (جدول شماره ۱۳) تأیید دیگری بر نقش و اهمیت عنوان مدارک در ساماندهی اطلاعات مربوط به آنها است. محققان دیگری چون فراست (Frost, 1989) و بزرگی (۱۳۷۷) نیز هر دو در بررسی خود نشان دادند که ۷۳ درصد واژه های عنوان دقیقاً یا حداقل با بخشی از واژه های سرعنوان های موضوعی همپوشانی دارند. از

طرف دیگر نشان داده شد که بین تعداد کلید واژه های عنوان و تعداد توصیفگرها نیز همبستگی بالایی وجود دارد و این همبستگی افزایش یا کاهش موارد همخوانی کلید واژه های عنوان و توصیفگرها را نیز تحت تاثیر قرار داده است (نمودار شماره ۲). در عین حال مشاهده شد که با گذشت زمان و به دلایلی چون تدوین اصطلاحنامه ها و آموزش نمایه سازان میزان وابستگی به کلید واژه های عنوان کاهش یافته و از این مراجع در اختصاص توصیفگرها استفاده شده است (نمودار شماره ۵). البته استفاده از این منابع در نمایه سازی از نقش و اهمیت عناوین نمی کاهد، بلکه استفاده توانمند از کلید واژه های عنوان و اصطلاحنامه ها منجر به افزایش جامعیت و مانعیت ذخیره و بازیابی اطلاعات خواهد شد.

بررسی کلید واژه های عنوان و توصیفگرها از نظر شیوه نگارش نیز یکی دیگر از اجزاء این تحقیق بود که طی آن مشخص شد به طور کلی با در نظر گرفتن کلیه حوزه ها نسبت تقریباً یکسانی بین توصیفگرهای تک واژه (۴۸ درصد) و ترکیب واژه (۵۲ درصد) وجود دارد هر چند درصد ترکیب واژه ها در حوزه های فنی و مهندسی و علوم پایه بیش از دیگر حوزه ها است. علاوه بر این مشخص شد توصیفگرها از منظر مفرد (۸۷ درصد) و جمع (۱۳ درصد) بودن نیز تا حد زیادی با مفرد (۸۴ درصد) و جمع (۱۶ درصد) کلید واژه ها همخوان است و شاید به همین دلیل در زبان فارسی بر خلاف زبان انگلیسی بیشتر از واژه های مفرد در امر نمایه سازی استفاده می شود (جدول شماره ۲۵). در بخش پایانی این تحقیق میزان بازیافت از طریق کلید واژه های عنوان و توصیفگرهای نمایه سازان در سه حالت همخوانی نسبی، همخوانی کامل و عدم همخوانی، به عنوان یکی از ملاک های ارزیابی پایگاه های اطلاعاتی بررسی شد. در حالت همخوانی نسبی، فزونی بازیافت از طریق توصیفگرها (۵۶ درصد) در مقایسه با کلید واژه ها و نیز بازیافت مساوی در موارد همخوانی کامل نقش و اهمیت انجام فرایندهایی به منظور ایجاد ارتباط منطقی بین توصیفگرها یا کلید واژه ها را آشکار ساخت. در خصوص کلید واژه ها و توصیفگرهای مترادف نیز به دلیل یکپارچگی بیشتر و برقرار شدن ارتباط منطقی بین توصیفگرهای هم مفهوم (در پی اجرای طرح ساماندهی توصیفگرها) موارد بازیافت از طریق توصیفگرها بیش از کلید واژه ها بود.

همچنین مشخص شد که ۳۷ درصد توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه های فارسی منطبق است که این رقم در حوزه علوم نظری مانند حوزه علوم انسانی و هنر بیش از دیگر حوزه ها است. فقدان اصطلاحنامه ها و نیز متخصصان موضوعی نمایه سازی در حوزه های فنی و مهندسی و علوم پزشکی را می توان از جمله دلایل همخوانی کمتر توصیفگرهای نمایه سازان با توصیفگرهای اصطلاحنامه ها دانست.

در پاسخ به سوال پایانی نیز مشخص شد که هر چند در سال های اخیر اختلاف زیادی بر سر موضوع جدا نویسی یا پیوسته نویسی کلمات مرکب بوده است ولی ۹۹ درصد توصیفگرها با رسم الخط مدون فرهنگستان زبان و ادب فارسی همخوان است.

پیشنهادات :

- در این تحقیق نشان داده شد که گردآوری داده های "پایگاه چکیده پایان نامه های ایران" (پایان نامه ها) از روال منظم و دقیقی پیروی نمی کند و با میزان پایان نامه های تولید شده در کشور همخوان نیست (نمودارهای شماره ۳ و ۴). پیشنهاد می شود مرکز اطلاعات و مدارک علمی ایران با واحد آمار و پژوهش های انفورماتیک وزارت علوم، تحقیقات و فناوری ارتباط نزدیک و مستمری برقرار کرده و آمار دقیقی از تعداد پذیرفته شده گان مقاطع تحصیلات تکمیلی داشته باشد. علاوه بر این با توجیه مسئولان معاونت های پژوهشی دانشگاه ها، یا رؤسای کتابخانه های دانشگاهی و پی گیری جدی تر می توان روند گردآوری پایان نامه ها را سرعت بخشیده و پایگاه مرکز را به عنوان یک مرجع مناسب غنی ساخت، به شکلی که پژوهشگران یا مراکز فعال در زمینه های موضوعی مختلف ضرورت ایجاد مراجع مشابهی را احساس نکنند.
- همان گونه که در مراحل نخست تحقیق اشاره شد در "پایگاه چکیده پایان نامه های ایران" رکوردهای تکراری بسیاری وجود دارد (جدول شماره ۲). کنترل ورود اطلاعات به پایگاه و طراحی سیستمی که مانع ورود رکوردهای تکراری شود یکی از ضرورت های این پایگاه به شمار می رود. علاوه بر این انجام تحقیق های مشابهی بر روی دیگر پایگاه های مرکز و حذف رکوردهای تکراری و ناقص از ابعاد مختلف صرفه جویی های بسیاری را برای طراحان پایگاه های مرکز و استفاده کنندگان آن در پی خواهد داشت.
- در ثبت و ضبط و تقسیم پایان نامه ها ذیل رشته های فرعی نیز ناهماهنگی و عدم یکپارچگی بسیاری مشاهده شد که ضرورت دارد دستورالعمل های مشخصی به منظور ثبت رشته های فرعی تدوین شود.

- طی سالیان متمادی نمایه سازی در مرکز اطلاعات و مدارک علمی ایران نزدیک به ۳۰۰۰۰۰ توصیفگر در پایگاه های این مرکز ثبت شده است. به چاپ رسیدن این توصیفگرها و نظامند کردن آن همانند یک اصطلاحنامه، نه تنها مانع پراگندگی اختصاص توصیفگرها می شود بلکه می تواند به عنوان مرجع مناسبی توسط نمایه سازان داخل و خارج از مرکز مورد استفاده قرار گیرد.
- یافته های این بررسی همانند تحقیق های مشابه نشان داد که ۶۷ درصد توصیفگرها _ به طور کامل یا نسبی _ با کلید واژه هایی که نویسندگان به منظور توصیف محتوی پایان نامه های خود، در عناوین به کار می برند همخوانی دارد. بدین ترتیب با اتکاء به عناوین پایان نامه ها، تا درصد بالایی امکان ذخیره و بازیابی مطلوب اطلاعات امکان پذیر است. نمایه سازی بر اساس کلید واژه های عنوان نه تنها ساده، سریع و ارزان صورت می پذیرد بلکه منعکس کننده واژگان رایج در حوزه مربوطه نیز هست. پژوهشگران معمولاً در تخصیص عناوین از کلماتی برای نمایاندن محتوای مدرک استفاده می کنند که دقیق و روزآمد بوده و منعکس کننده واژه های رایج در حوزه مربوطه هستند. آشنا کردن دانشجویان مقاطع تحصیلات تکمیلی، با نقش کلید واژه های عنوان در نمایه سازی و ذخیره و بازیابی اطلاعات، می تواند تأثیر به سزایی در انتخاب واژه های مناسب برای توصیف محتوای پایان نامه و نیز پرهیز از عنوان های استعاری، تمثیلی، قیاسی و... داشته باشد. از طرف دیگر در صورتی که طرح هایی، همانند طرح "ساماندهی توصیفگرهای پایگاه های اطلاعاتی (جستجوی هوشمند)" در خصوص کلید واژه های عنوان نیز انجام گیرد، باعث برقراری ارتباط بین واژه های پراکنده ای خواهد شد که بر یک مفهوم دلالت دارند.

پیشنهادهایی برای تحقیقات آینده:

- ۱- مرکز اطلاعات و مدارک علمی ایران دارای ۶ پایگاه اطلاعاتی دیگر با عناوین "پایگاه چکیده پایان نامه های دانشجویان ایرانی خارج از کشور"، "پایگاه همایش های ایران"، "پایگاه طرح های پژوهشی ایران"، "پایگاه گزارش های دولتی ایران"، "پایگاه محققین و متخصصین کشور" و "پایگاه چکیده مقالات علمی و فنی" است. پیشنهاد می شود تحقیق مشابهی بر روی این پایگاه های صورت گرفته و میزان همخوانی کلید واژه های عنوان و توصیفگرهای نمایه سازان در این پایگاه های مورد ارزیابی قرار گیرد.
- ۲- بررسی میزان استفاده از کلید واژه های عنوان و توصیفگرهای نمایه سازان در جستجو و بازیابی اطلاعات، توسط استفاده کنندگان از پایگاه های مرکز.
- ۳- بررسی میزان همخوانی کلیدواژه های موجود در چکیده پایان نامه ها، با توصیفگرهای نمایه سازان.
- ۴- بررسی میزان همخوانی کلید واژه های عنوان با توصیفگرهای اصطلاحنامه های فارسی.
- ۵- بررسی میزان همخوانی کلید واژه های عنوان پایان نامه ها به زبان انگلیسی (که در صفحه آخر پایان نامه ها درج می شود) با توصیفگرهای اصطلاحنامه های موجود به زبان انگلیسی.
- ۶- سنجش تعداد معادل های مناسب کلید واژه های عنوان، در اصطلاحنامه ها.

منابع

فهرست منابع و مآخذ

فارسی :

۱. آزادیان، آنوش. ۱۳۵۲. *نمایه سازی گردان و برخی از مشکلات آن در زبان فارسی*. پایان نامه کارشناسی ارشد علوم کتابداری و اطلاع رسانی، دانشگاه تهران، دانشکده علوم تربیتی.
۲. آقا بخشی، علی. ۱۳۷۶. *نمایه سازی همارا*. تهران : مرکز اطلاعات و مدارک علمی ایران.
۳. اکبری نژاد، سعید. ۱۳۷۶. *فاصله خالی میان واژه ها در ذخیره و بازیابی اطلاعات*. فصلنامه کتاب. ۸ (۱): ۴۹-۵۶.
۴. انصاری، مریم. ۱۳۷۹. *بررسی انطباق میان توصیفگرهای نمایه سازی و کلید واژه های عنوان پایان نامه های دکترای تخصصی کودکان، زنان، قلب و عروق و روانپزشکی*. پایان نامه کارشناسی ارشد کتابداری و اطلاع رسانی پزشکی، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی- درمانی ایران.
۵. بزرگی، اشرف السادات. ۱۳۷۷. *بررسی میزان سازگاری عنوان ها و موضوع های کتاب های فارسی منتشر شده در ایران در فاصله سال های ۱۳۶۵ - ۱۳۷۵*. پایان نامه دکترای، واحد علوم و تحقیقات دانشگاه آزاد اسلامی.
۶. بهاتاچاریا. ۱۹۷۴. *تاثیر زبان طبیعی در نمایه سازی و بازیابی علوم*. ترجمه مصطفی سید صادق. ۱۳۵۶. *فصلنامه اطلاع رسانی*. ۴ (۲۱): ۱۶-۲۰.
۷. پائو، میراندا لی. ۱۹۸۹. *مفاهیم بازیابی اطلاعات*. ترجمه اسدالله آزاد و رحمت الله فتاحی. ۱۳۷۹. مشهد: دانشگاه فردوسی.
۸. تیلور، آرلین جی. ۱۹۹۶. *سازماندهی اطلاعات*. ترجمه محمد حسین دیانی. ۱۳۸۱ مشهد: کتابخانه رایانه ای .
۹. چشمه سهرابی، مظفر. ۱۳۷۸. *تأثیر استفاده از اصطلاحنامه در بانک های اطلاعاتی کتابشناختی بر میزان جامعیت، مانعیت و مدت زمان جستجوی اطلاعات بازیابی شده*. پایان نامه کارشناسی ارشد علوم کتابداری و اطلاع رسانی. دانشکده علوم انسانی، دانشگاه تربیت مدرس.
۱۰. حدادیان مقدم، اسماعیل. ۱۳۷۷. *عنوان و ترجمه آن*. فصلنامه مترجم، ۲۶ (۲۱): ۱۰-۱۴.
۱۱. حرّی، عباس. ۱۳۷۲. *مروری بر اطلاعات و اطلاع رسانی*. تهران : دبیرخانه هیات امنای کتابخانه های کشور، نشر کتابخانه .
۱۲. حریری، مهر انگیز. ۱۳۶۱. *اصطلاحنامه در نظام ذخیره و بازیابی اطلاعات*. اطلاع رسانی، نشریه فنی مرکز اسناد و مدارک علمی. ۶ (۲۱): ۶۰-۶۷.

۱۳. حریری، مهرانگیز، مسعوده توفیق. ۱۳۶۲. نمایه گردان. اطلاع رسانی، نشریه فنی مرکز اسناد و مدارک علمی. ۸ (۲): ۴-۵.
۱۴. داور پناه، محمدرضا. ۱۳۷۵. بررسی میزان سازگاری عناوین مقالات فارسی با محتوای آنها. فصلنامه اطلاع رسانی، ۱۳ (۲): ۱-۱۲.
۱۵. دیانی، محمدحسین. ۱۳۷۸. پایان نامه، گنج ناشناخته رها شده. کتابداری و اطلاع رسانی، فصلنامه کتابخانه مرکزی و مرکز اسناد آستان قدس رضوی. ۲ (۴): ۱-۱۶.
۱۶. راولی، جنیفر. ۱۹۸۸. نمایه سازی و چکیده نویسی. ترجمه جعفر مهاد. ۱۳۷۴. تهران: سازمان مدارک فرهنگی و انقلاب اسلامی.
۱۷. سرایی، حسن. ۱۳۷۲. مقدمه ای بر نمونه گیری در تحقیق. تهران: سازمان مطالعه و تدوین کتب علوم انسانی.
۱۸. سلطانی، پوری. ۱۳۷۹. دانشنامه کتابداری و اطلاع رسانی. تهران: فرهنگ معاصر.
۱۹. سلطانی، پوری، کامران فانی. ۱۳۷۳. سرعنوان های موضوعی فارسی. تهران: کتابخانه ملی جمهوری اسلامی ایران.
۲۰. سمایی، سید مهدی. ۱۳۷۹. مفرد و جمع در نمایه سازی. فصلنامه اطلاع رسانی. ۱۶ (۱ و ۲)، ۲۷-۳۱.
۲۱. سینایی، علی. ۱۳۵۱. نمایه سازی همارا. اطلاع رسانی، نشریه فنی مرکز مدارک علمی. ۱ (۲): ۱۴-۲۰.
۲۲. شیر، علی اصغر. ۱۳۷۲. ارزیابی جامعیت و مانعیت نظام های بازیابی اطلاعات. پیام کتابخانه، (۱ و ۲): ۳۸.
۲۳. فلبر، هلموت. ۱۹۸۴. مبانی اصطلاح شناسی. ترجمه محسن عزیزی. ۱۳۸۱. تهران: مرکز اطلاعات و مدارک علمی ایران.
۲۴. کینن، استلا. ۱۹۹۶. فرهنگ فشرده کتابداری و اطلاع رسانی. ترجمه و تدوین فاطمه اسدی گرکانی. ۱۳۷۸. تهران: نشر کتابدار.
۲۵. لنکستر، فردریک ویلفرید. ۱۹۶۸. نظام های بازیابی اطلاعات، ویژگی ها، آزمون و ارزیابی. ترجمه جعفر مهاد. ۱۳۷۹. شیراز: نوید.
۲۶. محمدی فر، محمد رضا. ۱۳۸۰. آشنایی با مدرک شناسی. تهران: وزارت فرهنگ و ارشاد اسلامی.
۲۷. محمدی فر، محمد رضا. ۱۳۸۱. شیوه نامه ویرایش. تهران: وزارت فرهنگ و ارشاد اسلامی.
۲۸. محمدی فر، محمد رضا. ۱۳۸۱. مبانی نمایه سازی. تهران: وزارت فرهنگ و ارشاد اسلامی.
۲۹. معترف، مهوش. ۱۳۷۷. بررسی فعالیت های اطلاع رسانی مرکز اطلاعات و مدارک علمی ایران از ابتدا تاکنون. طرح پژوهشی. تهران: مرکز اطلاعات و مدارک علمی ایران.
۳۰. معرف زاده، عبدالحمید. ۱۳۶۶. بررسی وضعیت ابزار بازیابی اطلاعات پزشکی فارسی و پیشنهادی برای تهیه نمایه پزشکی ایران. پایان نامه کارشناسی ارشد کتابداری و اطلاع رسانی پزشکی، دانشکده مدیریت و اطلاع رسانی پزشکی، دانشگاه علوم پزشکی و خدمات بهداشتی - درمانی ایران.

۳۱. معمار طلوعی، محبوبه. ۱۳۸۱. نحوه نمایه سازی پایان نامه ها در مرکز اطلاعات و مدارک علمی ایران. فصلنامه کتاب. ۱۳ (۲): ۵۱-۵۷.
۳۲. وزارت فرهنگ و آموزش عالی. ۱۳۷۷. آمار عملکرد ۲۰ ساله آموزش عالی ایران. ۲۶.
۳۳. ویسمان، هرمان. بی تا. سیستم ها و خدمات اطلاع رسانی کامپیوتر. ترجمه جعفر مهران. ۱۳۷۳. شیراز: انتشارات نوید.
۳۴. یوسفی، احمد. ۱۳۷۶. ریزش کاذب در ذخیره و بازیابی اطلاعات. فصلنامه اطلاع رسانی، ۱۳ (۱): ۹-۱۴.

- 1- Aanonson, Jonh.1987. Precision and Recall in Title Keyword searchers. *Information technology and libraries*, 14(3):162-170.
- 2- Babbie, Earl R.1979. *The practice of Social Research*. 2nded. Belmont, CA: Wadsworth. P.274.
- 3- Bernier, Charles.1961. The Indexing Problem. *Journal of Chemical Documentation*. 1(3): PP. 25-27.
- 4- Bottle, Robert; Preibish.1970. Proposed KWIC index for Psychology. *Journal of the American Society for information science*,(11,12): 427-8.
- 5- Frost, Carolyn. 1989. Title word as entry vocabulary to LCSH: Correlation between Assigned LCSH term and derived terms from title. *Cataloging & Classification Quarterly*, 10(1, 2): 165-179.
- 6- Henzler, R.G. 1978. Free or controlled vocabularies. *International Classification*, 5.21-26.
- 7- Hmeidi, Ismail. 1997. Design and Implementation of Automatic Indexing for Information Retrieval with Arabic Documentation. *Journal of the American Society for Information Science*, 48(10):867-88.
- 8- Kaplan, Abraham. 1964. *The conduct of inquiry: Methodology for behavioral science*. San Francisco: Chandler publishing.
- 9- Keller, Barbara.1992.Subject Content through Title: A Master Matching study in Indian State University. *Cataloging & Classification Quarterly*, 15(3):69-80.
- 10- Lancaster, F.W. et al.1991.Identifying barriers to effective subject access in Library catalogs. *Library Resources and Technical Services*, 35.377-391.
- 11- Markey, Karen; Pauline Atherton; Newtor Cluadia.1980. An Analysis of Controlled Vocabulary and free text Search Statement in On-line Searches. *On line Review*, 4: 225-236.
- 12- Newmark, peter. 1988. *Approaches to translation*, Prentice Hall International.
- 13- Rydings, Tony. 1999. Indexing of periodicals- format of entries. *Society of indexers Genealogical Group newsletter*. No.20: p.5-8. [Online].<http://www.sigg.org.uk/arts/arts2001.htm>. [22 Sep 2003]
- 14- Shepherd, Michael; A., Phillips; wJ. Lui.1983. *Overlap of title keyword and Assigned Descriptors in the OON Database*. *Canadian Journal of Information Service*, 12(1):58-66.
- 15- Spink, Amanda; Saracevic, Tefko.1993. Interaction in Information Retrival: Selection & Effectiveness of Search Terms. *Journal of the American Society for Information Science*, 48(8): 741-761.

- 16-**Voorbij, Henk. 1998. Title Keywords and subject descriptors A Comparison of subject search entries of books in the humanities and social sciences. *Journal of Documentation*. 54(4):466-476.
- 17-**Weber, Robert Philip. 1990. *Basic Content Analysis*. London: Sage.