

خلاصه روش پیشنهادی برای طبقه‌بندی مطالب استخراج شده از کتاب‌ها

موجودی

۱. متن کتاب‌های مرتبط با تبلیغ
۲. فهرست کتاب‌ها به سه صورت؛ فایل Word، فایل PDF و فایل تصویری

نیازمندی

۱. مطالب تفکیک‌شده بر اساس موضوع
۲. درختواره‌ای از موضوعات طبقه‌بندی شده
۳. ارتباط درختواره موضوعی به مطالب تفکیک‌شده
۴. ارجاع از مطالب تفکیک‌شده به اصل منبع کتاب

مسیر دستیابی به نیازمندی با استفاده از موجودی

۱. یکسان‌سازی فهرست‌ها
 - ۱/۱. تایپ فهرست‌های تصویری و PDF
 - ۲/۱. انتقال فهرست‌های Word به Excel فرمت‌گذاری شده
 - ۳/۱. انتقال فهرست‌های تایپ‌شده از فایل‌های تصویری و PDF به Excel فرمت‌گذاری شده
۲. دسته‌بندی اجمالی فایل Excel و تبدیل به پنج یا شش گروه از موضوعات
 - ۱/۲. تجزیه فایل اکسل به چند فایل جدا از هم در کلان‌موضوعات
 - ۲/۲. تحویل اکسل هر کلان‌موضوع به یک محقق یا یک گروه کاری
۳. استخراج متن‌ها از طریق نشانی ذکر شده در فایل اکسل برای هر کلان‌موضوع
۴. برچسب‌گذاری موضوعی متن‌ها و تقطیع آن‌ها به بخش‌های کوچک‌تر
۵. استخراج برچسب‌ها و فهرست نمودن آن‌ها در یک فایل اکسل مستقل برای هر کلان‌موضوع
 - ۱/۵. مراقبت برای از بین نرفتن ارتباط برچسب‌ها با متن تقطیع شده، از طریق کدگذاری آن‌ها
۶. تجمیع تمامی برچسب‌های تمامی کلان‌موضوع‌ها در کنار هم و به صورت الفبایی
۷. ادغام برچسب‌های مترادف و مشابه و انتخاب یک عنوان «نهایی»
 - ۱/۷. مراقبت برای حفظ ارتباط برچسب‌های تغییر عنوان داده با برچسب نهایی
۸. طبقه‌بندی برچسب‌ها و تولید درختواره

نتیجه

۱. اکنون درختواره‌ای از موضوعات داریم
۲. عناوین درختواره به متن‌های تقطیع شده متصل هستند
۳. متن‌های تقطیع شده به منبع کتابی خود مرتبط هستند

پایان