خلاصه روش پيشنهادي براي طبقه‌بندي مطالب استخراج شده از كتاب‌ها

16 تير 1395

# موجودي

1. متن كتاب‌هاي مرتبط با تبليغ
2. فهرست كتاب‌ها به سه صورت؛ فايل Word، فايل PDF و فايل تصويري

# نيازمندي

1. مطالب تفكيك‌شده بر اساس موضوع
2. درختواره‌اي از موضوعات طبقه‌بندي شده
3. ارتباط درختواره موضوعي به مطالب تفكيك‌شده
4. ارجاع از مطالب تفكيك‌شده به اصل منبع كتاب

# مسير دستيابي به نيازمندي با استفاده از موجودي

1. يكسان‌‌سازي فهرست‌ها
   1. تايپ فهرست‌هاي تصويري و PDF
   2. انتقال فهرست‌هاي Word به Excel فرمت‌گذاري شده
   3. انتقال فهرست‌هاي تايپ‌شده از فايل‌هاي تصويري و PDF به Excel فرمت‌گذاري شده
2. دسته‌بندي اجمالي فايل Excel و تبديل به پنج يا شش گروه از موضوعات
   1. تجزيه فايل اكسل به چند فايل جدا از هم در كلان‌موضوعات
   2. تحويل اكسل هر كلان‌موضوع به يك محقق يا يك گروه كاري
3. استخراج متن‌ها از طريق نشاني ذكر شده در فايل اكسل براي هر كلان‌موضوع
4. برچسب‌گذاري موضوعي متن‌ها و تقطيع آن‌ها به بخش‌هاي كوچك‌تر
5. استخراج برچسب‌ها و فهرست نمودن آن‌ها در يك فايل اكسل مستقل براي هر كلان‌موضوع
   1. مراقبت براي از بين نرفتن ارتباط برچسب‌ها با متن تقطيع شده، از طريق كدگذاري آن‌ها
6. تجميع تمامي برچسب‌هاي تمامي كلان‌موضوع‌ها در كنار هم و به صورت الفبايي
7. ادغام برچسب‌هاي مترادف و مشابه و انتخاب يك عنوان «نهايي»
   1. مراقبت براي حفظ ارتباط برچسب‌هاي تغيير عنوان داده با برچسب نهايي
8. طبقه‌بندي برچسب‌ها و توليد درختواره

# نتيجه

1. اكنون درختواره‌اي از موضوعات داريم
2. عناوين درختواره به متن‌هاي تقطيع شده متصل هستند
3. متن‌هاي تقطيع شده به منبع كتابي خود مرتبط هستند

پايان