Crowdsourcing Terms for Thematic Exploration in the Catholic Portal

Ingrid Hsieh-Yee, Catholic University of America
Pat Lawton, University of Notre Dame

The Catholic Portal of the Catholic Research Resources Alliance (CRRA) facilitates access to resources related to studies of the Catholic experience and supports CRRA’s mission to provide enduring global access to Catholic research resources. The Portal builds its collections on 12 themes, and CRRA has experimented with portal theme enhancements to support exploration and use of its collections. The crowdsourcing project is part of the subject access enhancement effort.

Project objectives
1. To analyze topical terms from users for patterns and subject coverage
2. To identify highly relevant expressions for inclusion in the Portal

Topical terms are compared with the Library of Congress subject headings and analyzed for possible linking to Portal subject headings. Findings will have implications for enhancing subject access to Catholic resources in local catalogs and the CRRA’s Catholic Portal.

Methodology
Crowdsourcing through an online survey

Survey Design
The study focuses on eight of the 12 CRRA Portal Themes. To make the tasks more manageable for participants, two online surveys were designed. The surveys are identical in all aspects, except one version covers Catholic Education, Liturgy, Peace-building, and Catholicism and Citizenship, while the other covers Diocesan Collections, Catholic Intellectual Life, Catholic Missions, and Catholic Social Action.

Data Collection
Fifty-eight librarians from members of the Catholic Research Resources Alliance were invited to participate in an online survey to provide topical terms for selected Portal themes. The survey has a 71% response rate, with 41 of the 58 invitees participating in the study. However, not all of them provided topical terms for the four portal themes presented in the survey.

Participant Background (n=41)
(up to 3 areas named)

Terms Collected: 404 terms

<table>
<thead>
<tr>
<th>Topic</th>
<th>Education</th>
<th>Liturgy</th>
<th>Peace-building</th>
<th>Catholicism &amp; Citizenship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of terms collected</td>
<td>42</td>
<td>47</td>
<td>40</td>
<td>38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic</th>
<th>Diocesan</th>
<th>Intellectual life</th>
<th>Missions</th>
<th>Social actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of terms collected</td>
<td>63</td>
<td>55</td>
<td>60</td>
<td>59</td>
</tr>
</tbody>
</table>

Project Status
Data analysis of Catholic Education, Catholic Social Action, Liturgy, and Peace-building has been completed, and preliminary findings are reported in this poster.

Next Steps
- Topical terms for the remaining four themes will be analyzed.
- Findings of this project will be compared with findings from a portal heading analysis project to assess the merits of each approach for portal enhancement.

Term Analysis
Step 1. Sort terms by level of relevance reported by each participant, resulting in groups of terms marked as HR (highly relevant), R (relevant), and N (neutral)
Step 2. Start with the HR terms and search against LCSH in the authority file at http://authorities.loc.gov to identify user terms that match LCSH or cross references.
Step 3. Check R terms and N terms against LCSH authority file to identify user terms that match LCSH or cross references.
Step 4. Results of Step 2 and Step 3 identify LC subject headings and cross references matched by user terms. Use the matched LC subject headings as the pool to code user terms that do not match LC subject headings. Data values for coding term relationships to identified LCSH include

Sample codes for term relationship

VA: Variant expression
“Catholic education” for “Catholic Church—Education” (an LCSH)
NA: New concept, associated with an identified subject heading
“Seminary” to “Catholic Church—Education”
NB: New concept, broader than an identified subject heading
No: New concept, narrower than an identified subject heading
“Catholic teachers” to “Catholic Church—Education”
NC: New concept. No hierarchical relations with headings identified

Pattern of Findings
Besides LCSH matches and cross references already in LCSH, these three types of term relationships, if accommodated, will increase success in retrieving items from the Catholic Portal:

- VA: variant expression to a named subject heading
- NA: new concept, associated with a named subject heading
- NB: new concept, narrower than a named subject heading