Introduction to the starting question

Features of Colon Classification

CC is an analytical-synthetic scheme for classification, as it allows the classifier to build any class number necessary to express composed subjects.

Nevertheless, Part 3 of CC contains “Schedules of classics and sacred books with special names”.

It is a long list of ready-made class numbers for classic works in many main classes such as L: Medicine, Δ: Spiritual Experience and mysticism, N: Fine arts, O: Literature, P: Linguistics, Q: Religion and R6: Indian Philosophy.

What if a Part 4 Western Classics and Western Literature Classics?

Creation of CC class numbers

Prospective Part 4 of CC would require a considerable amount of time and hard work, so:

1. Is the, quick, cost-effective on big scale, automatic production of a Schedule of Colon class numbers for Western literary classics possible?
2. Which are the requirements for a large scale automatic production of class numbers of a faceted scheme of classification?
Scope of the research

- To investigate a prospective positive answer to one or both questions, two prerequisites must be complied with:
  1. creation of class numbers must be possible through a regular, exception-free and predictable process, so to be transformed in an algorithm;
  2. modular, sharable, reusable and preferably free data needed to give value to the facets must be available

Scope of the research (2)

- The former prerequisite is granted by CC, as it is an analytical-synthetic scheme of classification
- More exactly, CC presents two very useful specific characteristics: Classic device and the facet formula in Main Class O Literature.
- The latter prerequisite should be theoretically granted by Linked Open Data available in the Semantic Web.

Classic works and classic device

- In CC, a classic work is "a book stimulating other books and literature on itself";
- In CC two types of classic devices are provided;
  1) Classic in main classes O Literature and Q Religion (chronological principle)
  2) Classic in any other Main Class (subject approach, and bibliographical-literary approach).

Class O Literature and its facet formula

- Facet formula of Main Class O Literature is the following:
  O [P], [P2] [P3], [P4]
- So that a class number for a literary work can be obtained by the combined notation of this elements
  O + Language +, + Lit. Form + Author +, + No. Work
  Notice: Facets are to be used when apply

Class O Literature. Example

O + Language +, + Lit. Form + Author +, + No. Work

- A book of Russian Poetry
- Literature = O
- Russian language = 142
- Poetry = 1
  \( \rightarrow O_{142,1} \)
- English (111) poetry \( \rightarrow O_{111,1} \)
- Polish (145) Drama (2) \( \rightarrow O_{145,2} \)

Class O Literature. Example (2)

O + Language +, + Lit. Form + Author +, + No. Work

- Shakespeare's Romeo and Juliet
  Literature = O; English = 111; Drama 2
- Shakespeare (born in 1564) = J54
- Romeo and Juliet = (id. number for W) 1
  \( \rightarrow O_{111,2J54,1} \)
- Hamlet = 2
  \( \rightarrow O_{111,2J54,2} \)
- And so on …
To sum up

To complete a CC class number for an author or a work of an author the following data are needed
- Language (i.e. Author’s nationality of lang.)
- Literary Form mostly associated with the Author
- Author’s year of birth
- Date of the Author’s Work (or arbitrary no.)

Are they available?

Literary data in the semantic web

Wikidata: very rich amount of data, but lacking of consistency, regularity and completeness as to the literary universe

Specialised sources of LOD:
- "datos.bne.es" (LOD from Spanish National library)
- "data.bnf.fr" (LOD from the French National Library)

SPARQL Query on data.bnf.fr

PREFIX foaf: <http://xmlns.com/foaf/0.1/>     # [try it]
PREFIX bnf-onto: <http://data.bnf.fr/ontology/bnf-onto/>
PREFIX rdagroup1elements: <http://rdvocab.info/Elements/>
PREFIX rdagroup2elements: <http://rdvocab.info/ElementsGr2/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
WHERE {
  ?opera dcterms:title ?titolo_opera ;
  rdfs:label ?altro_titolo_opera ;
  rdagroup1elements:dateOfWork ?data_opera ;
  bnf-onto:firstYear ?anno_opera.
}

The sample and its data

- Italian authors born between 1800 and 1900 were chosen, according to the following requirements:
  - Birth year between 1800 and 1900;
  - "Italy" as country associated to the person;
  - "Italian" as language associated to the person;
  - "Literature" as field of activity (!)

Results: 33 authors and 99 works

Processing of the data sample

Data sample was processed by OpenRefine:
- Deduplication of records
- Arrangement by author, date of the work and title of the work
- Progressive numbering of author’s works
- Substitution of values (terms) with required CC notation (e.g. poetry=1, drama=2, 1836= M36 etc.)
- Export of the table in a Part 4 sample Schedule
Discussion

Question 1?
The quick, large-scale, automatic creation of class numbers of the main class O Literature relating to classics (single works of specific authors) resulted workable
But, the following issues emerged:
1. The extraction of linked open data about literary authors resulted partial and lacking;
2. The identification of the relevant literary class for each author was not completely automated

Discussion (2)

Question 2?
The case study highlighted the following requirements for an automatic production of class numbers of a faceted scheme of classification:
1. the classification scheme to be based on regular semantics and syntaxes – that is to say on facets and facet formula;
2. the availability of semantic data with specific characteristics:
   1. Semantic data are not sufficient (may be inconsistent)
   2. Consistent semantic data from one repository/source are not sufficient (may be incomplete)
   3. Complete and consistent semantic data are not sufficient (maybe partial with respect to their domain)

Five lines for further research

1. Investigate the easy and quick identification of authors’ literary form
2. Identify other CC facet formulas, to investigate more case studies (e.g. classics in other Main Classes, or Music resources)
3. Investigate the use of LOD for the production of book numbers (identifiers for books with the same class number)

Questions?

1. Investigate the easy and quick identification of authors’ literary form
2. Identify other CC facet formulas, to investigate more case studies (e.g. classics in other Main Classes, or Music resources)
3. Investigate the use of LOD for the production of book numbers (identifiers for books with the same class number)

Five lines for further research

4. Investigate and detect facet formulas in other classifications (e.g. 400 Linguistics, 780 Music, or 800 Literature in DDC)
5. Design one purely faceted classification scheme mainly based on data easily available as LOD