UNIVERSITÀ DI PAVIA

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION

JAGIELLONIAN UNIVERSITY IN KRAKOW

From semantic web to faceted classification:

A case study and five lines of future research

A joint event of ISKO Poland and ISKO Italy
Krakow, 16-17 January 2019

1

Summary

- The starting question
- CC Classic device and Basic Class O Literature
- Literary data in the Semantic Web
- The data sample
- Data sample processing
- Discussion
- Five lines of future research

2

Introduction to the starting question

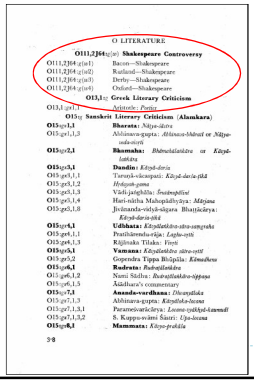
3

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*.

4

Features of Colon Classification


O LITERATURE

O11,2[64:(g1e) **Shakespeare Controversy**
O11,2[64:(g1e1) Bacon—Shakespeare
O11,2[64:(g1e2) Rutland—Shakespeare
O11,2[64:(g1e3) Derby—Shakespeare
O11,2[64:(g1e4) Oxford—Shakespeare

O13,1 **Greek Literary Criticism**
O13,1:(p1) *Aristotle: Poetics*

O15:g **Sanskrit Literary Criticism (Alaṃkāra)**
O15:g1,1 **Bharata:** *Nāṭya-śāstra*
O15:g1,1,3 *Abhinava-gupta: Abhinava-bhāratī* or *Nāṭya-veda-vivṛti*

O15:g2,1 **Bhāmaha:** *Bhāmahālaṅkāra* or *Kāvyā-laṅkāra*

O15:g3,1 **Daṇḍin:** *Kāvyā-darśa*
O15:g3,1,1 Taruṇa-vācaspati: *Kāvyā-darśa-ṭīkā*
O15:g3,1,2 *Hṛdaya-gama*
O15:g3,1,3 Vādi-jaṅghāla: *Śrutabodhinī*
O15:g3,1,4 Hari-nātha Mahopādhyāya: *Mādana*
O15:g3,1,8 Jīvānanda-vidyā-sāgara Bhaṭṭācārya: *Kāvyā-darśa-ṭīkā*

O15:g4,1 **Udbhata:** *Kāvyālaṅkāra-sāra-saṃgraha*
O15:g4,1,1 Pratīhāreṇdu-rāja: *Laghu-vṛtti*
O15:g4,1,3 Rājānaka Tilaka: *Vivṛti*
O15:g5,1 **Vamana:** *Kāvyālaṅkāra sūra-vṛtti*
O15:g5,2 Gopendra Tippa Bhūpāla: *Kāmadhenu*
O15:g6,1 **Rudraṭa:** *Rudrayālaṅkāra*
O15:g6,1,2 Namī Sādha: *Rudrayālaṅkāra-ṭippaṇa*
O15:g6,1,5 Āśādhara's commentary
O15:g7,1 **Ānandavardhana:** *Dhvanyāloka*
O15:g7,1,3 Abhinava-gupta: *Kāvyāloka-locana*
O15:g7,1,3,1 Parameśvarācārya: *Locana-vyākhyā-kaumudī*
O15:g7,1,3,2 S. Kuppu-swāmi Śāstri: *Ujjvala-locana*
O15:g8,1 **Mammaṭa:** *Kāvya-prakāśa*

38

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

5

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?

6

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

7

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.

8

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).

9

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

10

Class O Literature. Example

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

- A book of Russian Poetry
- Literature = O
- Russian language = 142
- Poetry = 1

→ O142,1

- English (111) poetry → O111,1
- Polish (145) Drama (2)→ O145,2

11

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

→ O111,2J54,1

- Hamlet = 2

→ O111,2J54,2

- And so on …

12

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?

13

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)

14

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 rdagroup2elements:
<http://rdvocab.info/ElementsGr2/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdagroup1elements: <http://rdvocab.info/Elements/>
SELECT DISTINCT ?opera ?titolo_opera lang(?titolo_opera)
?data_opera ?anno_opera ?autore ?NomeCompleto
?NomeDiBattesimo ?Cognome ?nato ?morto ?tipo_autore
WHERE {
  ?opera dcterms:creator ?autore.
  ?opera dcterms:title ?titolo_opera ;
        rdfs:label ?altro_titolo_opera ;
   rdagroup1elements:dateOfWork ?data_opera ;
bnf-onto:firstYear ?anno_opera.
```

15

The sample and its data

- Italian authors born between 1800 and 1900 were chosen, according to the following requirements:
- Birth year between1800 and 1900;
- "Italy" as country associated to the person;
- "Italian" as language associated to the person;
- "Literature" as field of activity (!)

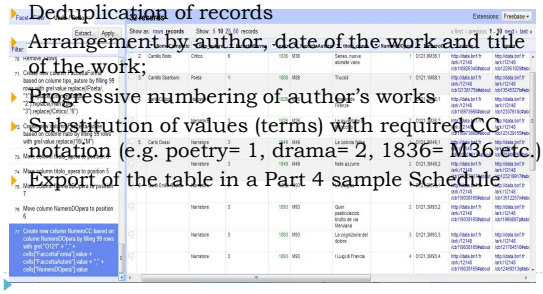
Results: 33 authors and 99 works

16

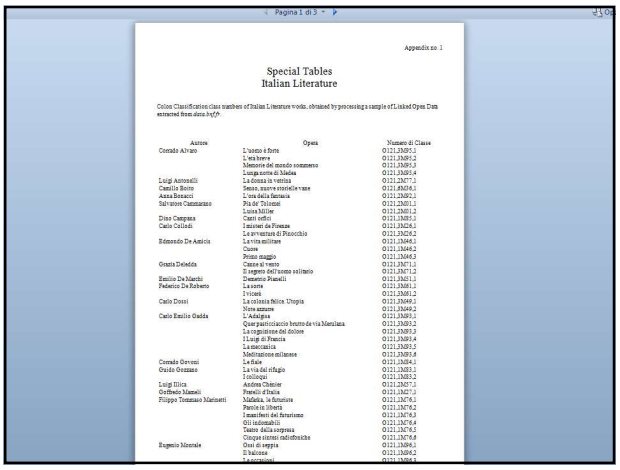
Processing of the data sample

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



17



Appendix no. 1

Special Tables
Italian Literature

Colon Classification class numbers of Italian Literature works, obtained by processing a sample of Linked Open Data extracted from data.bnf.fr

Autore	Opera	Numero di Classe
Corrado Alvaro	L'uomo è forte	0121,2M95,1
	L'età breve	0121,3M95,2
	Memorie del mondo sommerso	0121,3M95,3
	Lunga notte di Medea	0121,3M95,4
Luigi Antonelli	La donna in vertina	0121,2M77,1
Camillo Boito	Senza, nuove storielle vane	0121,4M28,1
Anna Bonacci	L'ora della farmacia	0121,2M92,1
Salvatore Cammarano	Fra de' Talentati	0121,2M81,1
	Luisa Miller	0121,2M81,2
Dino Campana	Canti orfici	0121,3M85,1
Carlo Collodi	I misteri di Firenze	0121,3M28,1
	Le avventure di Pinocchio	0121,2M28,2
Edmondo De Amicis	La vita militare	0121,1M46,1
	Cuore	0121,1M46,2
	Primo maggio	0121,1M46,3
Grazia Deledda	Canne al vento	0121,2M71,1
	Il segreto dell'uomo solitario	0121,3M71,2
Emilio De Marchi	Demetrio Pianelli	0121,2M51,1
Federico De Roberto	La sorte	0121,3M61,1
	I viceré	0121,3M61,2
Carlo Dossi	La colonia felice: Utopia	0121,3M49,1
	Note azzurre	0121,3M49,2
Carlo Emilio Gadda	L'Adalgisa	0121,3M93,1
	Quer pasticciaccio brutto de via Merulana	0121,3M93,2
	La cognizione del dolore	0121,3M93,3
	I Luigi di Francia	0121,3M93,4
	La meccanica	0121,2M93,5
	Meditazione milanese	0121,3M93,6
Corrado Govoni	Le fiale	0121,3M54,1
Guido Gozzano	La via del rifugio	0121,3M83,1
	I colloqui	0121,3M83,2
Luigi Illica	Andrea Chénier	0121,2M27,1
Goffredo Mameli	Fratelli d'Italia	0121,1M27,1
Filippo Tommaso Marinetti	Mafarka, le futurista	0121,1M76,1
	Parole in libertà	0121,1M76,2
	I manifesti del futurismo	0121,1M76,3
	Gli indomabili	0121,1M76,4
	Teatro della sorpresa	0121,1M76,5
	Cinque sensi radiofonici	0121,1M76,6
	Gusi di sergia	0121,1M96,1
	Il balcone	0121,1M96,2
	La cucina	0121,1M96,3

18

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

19

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 semantica data are not sufficient (maybe partial with respect to their domain)

20

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)

21

22

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

23

Questions?

24