Making a bibliography using metadata

The case of the Norwegian photo book

Lars G Johnsen
Arthur Tennøe

In this presentation we will discuss how one might create a bibliography using metadata taken from libraries in conjunction with other sources external to library information. We limit our scope to enumerative bibliographies as opposed to annotated (or descriptive) bibliographies (Carter and Barker (2010).

A bibliography is in general a list of books that satisfy some description or idea, which in this work is specified to as the concept photo book. taken as an artistic book containing photos that are not made for any . Here we are after a special kind of photo book, the book as an art form, distinguishing it from other books containing photos for special (scientific) purposes, while being open towards the possibility that the book might have a narrative, however the story is mainly conveyed via the photographic images.

We believe that the method developed here can be generalized to other topics, especially the use of library metadata, and as such should be of interest to scholars and librarians that work on other topics.

Metadata for books are provided as MARC-posts (see e.g. Library of Congress (2018)) in a national bibliography, which contains information on all books within a national library system. Some of these books are equipped with classificatory information like subject headings and Dewey decimal classification (see e.g. Majumder and Sarma (2007) on WebDewey and references therein), although all books have information about publication date and author, further classifications varies. In the approximately 500 000 books in the Norwegian National Library, a third is equipped with a Dewey decimal classification. However, the number almost doubles to 60 %, if only newer additions (books after 1970) are considered.

For our purposes, we also consulted an external source of Norwegian photographers listed in Wikipedia, who have published photo books among their works. This list is checked against library metadata to extract titles and classification information, like Dewey and subject keywords.

The Dewey number of most interest is the decimal 779 which is about photo books. In contrast to keywords, the Dewey system comes with relatively detailed instructions for applying its codes. Topic words rely more on the subjective interpretation of the librarian in charge of the classification.

Now, any classification (subject heading or Dewey number) and combinations of these will give rise to a list of books: simply select the books that satisfy the combination. The challenge when making a bibliography is finding these combinations. What subject headings will single out a list of photo books?
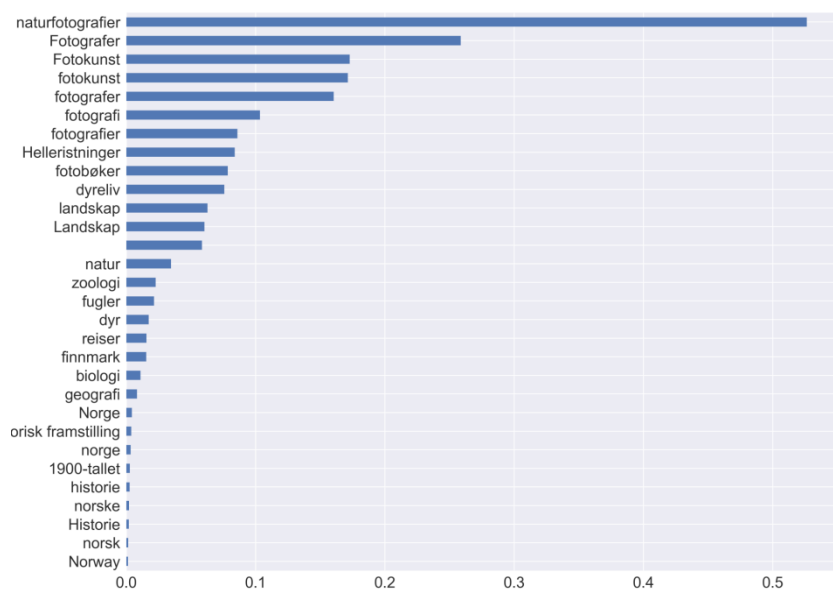
In order to browse the set of topic words (i.e. subject headings) we use the notion of a sister topic. Every book is classified in more than one way, while the Dewey number is unique (up to which library responsible for classification), there are a number of other topics occurring together with a particular topic keyword. The co-occurrence and the frequency will give hints as to which other topic words that may be worth pursuing. Some may be relevant while some others are not.

In the following figure we show a heat map illustrating the topics for keywords relate to photography. This particular table is sorted according to the keyword "Fotokunst" (photographic art), where the capitalized key words signalizes that it belongs to a restricted keyword thesaurus within the Norwegian library system, while unrestricted keywords are used with lower case initial letters. The heat map highlights keywords within each column and the value in each cell shows how many books the row keyword has in common with the columnar keyword. The cell with identical row and column keyword contains the number of titles for that particular keyword.

	Fotohistorie	Fotokunst	foto	fotobok	fotobøker	fotografi	fotografier	fotohistorie	fotokunst	kunstfotografi
Fotokunst	0	280	5	4	31	65	68	8	51	2
Norge	3	110	15	10	33	33	98	17	29	3
Fotografer	2	86	1	0	3	34	30	7	9	2
fotografier	1	68	16	2	52	49	448	9	29	0
fotografi	2	65	10	0	4	292	49	9	52	4
fotokunst	0	51	3	0	7	52	29	0	111	1
norge	2	33	8	10	37	69	111	10	13	2
fotobøker	0	31	1	25	435	4	52	2	7	0
fotografer	0	28	1	0	0	45	25	5	43	0
natur	0	28	1	2	46	24	18	1	5	0
Motiver (Bildekunst)	0	23	0	0	1	10	7	0	8	0
Fotografisamlinger	1	22	0	0	1	4	5	1	9	0
Lofoten	0	20	0	0	7	16	1	0	2	0
Fotografi	3	19	8	0	1	44	28	10	3	1
Fotografering	0	18	8	2	1	31	20	3	3	1
dikt	0	18	0	1	3	18	9	1	4	0
Portrett	0	18	1	0	0	3	5	1	3	0

Such tables can be used to study the relationship between keywords, and how they can be used to refine sets of books. For instance

For the list of Norwegian photographers, we also conducted a study of the topic words associated with their books. These are shown in the following graph, which is adjusted so that typical keywords appear first. Note that while nature photography (naturfotografi) is at the top on this list, photographic art is high (fotokuns).

Using the above information, together with Dewey classifications and keywords we were able to construct a list of about 100 titles.

In addition to the above information we plan also to use the full text information. Photographic books tend to have a small text to photo ratio, which can be used in order to get candidates for books, particularly for those that are not classified with respect to Dewey or topic word. This will perhaps also have a wider interest for projects trying to build bibliographies in general.

References

John Carter; Nicolas Barker (2016). "Bibliography". ABC for Book Collectors (9th ed.). Oak Knoll Press and British Library

Library of Congress (2018) (https://www.loc.gov/marc/)

Majumder, Apurba Jyoti; Gautam Sarma. "Webdewey: The Dewey Decimal Classification in The Web" INFLIBNET Centre, Ahmedabad, Planner 2007. As PDF:
http://ir.inflibnet.ac.in:8080/ir/ViewerJS/#../bitstream/1944/1047/1/16.pdf