19_10_2022

Eligibility criteria - Finding Venues (revisited)

Selecting a set of candidate venues (Journals and Conferences) that will establish the basis of the research process represents the first step in defining the eligibility criteria of the covered literature.

In order to perform a grounded selection, the repositories of a set of five major scholarly research publishers were queried and the resulting papers aggregated by venue and counted.
The consulted **repositories** were:

- IEEEXplore
- ACM DL
- SpringerLink
- ScienceDirect
- ACL Anthology

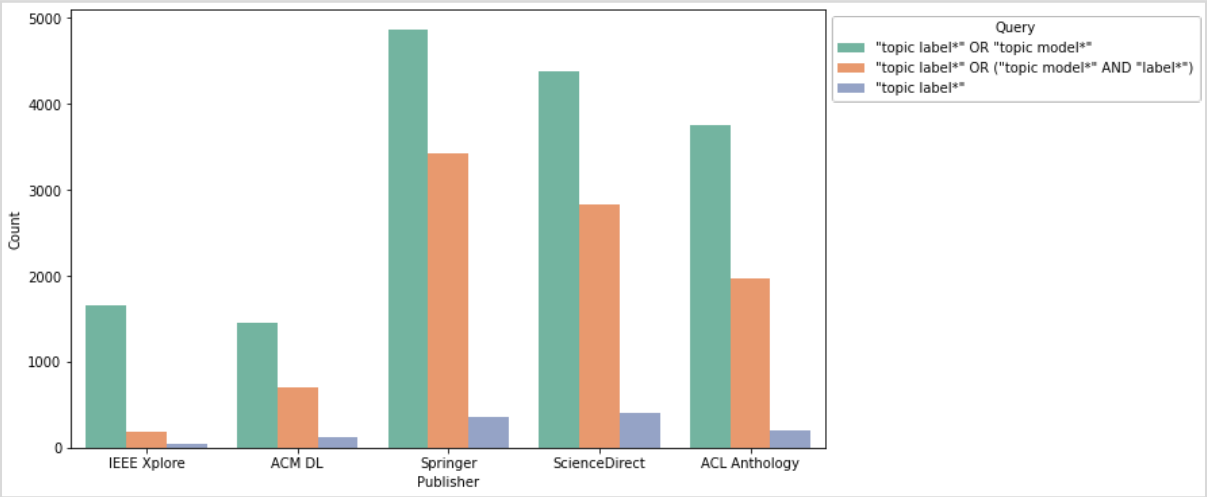
In order to perform this initial exploratory search, **three different queries** were taken into consideration.
The three queries (together with the number of results that they generated) look as follows:

- `"topic label*" OR "topic model*"`
  - IEEE Xplore: 1649
  - ACM DL: 1447
  - Springer: 4863
  - ScienceDirect: 4385
  - ACL: 3750 *

- `"topic label*"`
  - IEEE Xplore: 39
  - ACM DL: 128
  - Springer: 364
  - ScienceDirect: 397
  - ACL: 206 *

- `"topic label*" OR ("topic model*" AND "label*")`
  - IEEE Xplore: 179
  - ACM DL: 708

- Springer: 3427
- ScienceDirect: 2833
- ACL: 1970 *



Ultimately, the venue selection was performed based on the aggregated results obtained using the third query (`"topic label*" OR ("topic model*" AND "label*")`).
The choice of using this query stems from the fact that it represents a **middle ground** between the narrow query that selects only papers that explicitly mention topic labeling (or topic labels, topic label, etc…) and the broad query that also retrieves all work containing terms related to "topic model".
The chosen query ensures that the collected papers that do not explicitly mention topic labeling have at least some mention of terms related to the root word "label".
An example of this scenario would be a paper containing the term "Topic modeling" that mentions the labeling activity as "Labeling of the generated topics".
At the same time, the selected query avoids retrieving a significant amount of potentially irrelevant papers lacking any mentions of any labeling activities.

The obtained results were evaluated and the venues selected by taking into account both the **nr. of retrieved papers** and the **rating** issued by the chosen reference bodies.
In this regard, the SCImago Journal Rank (**SJR** score averaged in the time span 2017-2021) was considered for journals and the Computing Research and Education Association of Australasia (**CORE** Inc.) and the GII-GRIN-SCIE (**GGS**) ratings were considered for conferences.

All repositories were queried considering **all metadata** available for a given candidate result (title, abstract, content, tags, etc…).

The only notable exception to this fact is represented by **ACL Anthology**. Since this

repository does not directly allow to obtain aggregated data from query results, the aggregated data was obtained from the full bibliography with abstract.
Unlike other repositories, this complete bibliography only offered two useful fields (title and abstract) that could be used in order to obtain the aggregated results.
This limitation was taken into account when analysing the result values related to the candidate venues published in ACL Anthology.

All the selected journals publish original research.

Side note:
Some repositories support the operator `NEAR`, which can be used to search for terms that appear in the vicinity of one another.
During the search process, this can be used to create a refined version of the query that has been utilised to select the venues:

- IEEEXplore: `"topic label*" OR ("topic model*" AND ("topic*" NEAR/10 "label*"))`
- Springer: `"topic label*" OR ("topic model*" AND ("topic*" NEAR "label*"))`

Pajek / PajekXXL / Pajek3XL

Citation network

#Thesis/Temporary notes#