

Publindex: A Web Service to Automatically Evaluate Research Publications According to Customized Criteria

Eduardo Mena
 emena@unizar.es
 University of Zaragoza
 Zaragoza, Spain

Carlos Bobed
 cbobed@unizar.es
 everis / NTT Data
 Zaragoza, Spain

Álvaro Juan
 alvarojc95@gmail.com
 University of Zaragoza
 Zaragoza, Spain

ABSTRACT

We introduce Publindex, a system that retrieves, classifies, and returns research publications of a given researcher according to the criteria and in the format predefined by the user.

KEYWORDS

Research Publications Evaluation

ACM Reference Format:

Eduardo Mena, Carlos Bobed, and Álvaro Juan. 2020. Publindex: A Web Service to Automatically Evaluate Research Publications According to Customized Criteria. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20)*, August 1–5, 2020, Virtual Event, China. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3383583.3398630>

1 MOTIVATION

Evaluating the research activity of researchers is a very common task in many situations. Reading several tens of papers per candidate is an unaffordable goal, especially when dealing with many people with a high number of publications. Therefore we estimate the quality of such publications by analyzing different publication features, such as the quality of the publication venue, number of pages, list of coauthors, or number of citations, among other parameters. In any case, gathering all this information from a high number of publications is indeed a very time-consuming task. Ideally, we would like to have a system where, just by entering the id of a researcher, you get his/her best research contributions.

2 PUBLINDEX

Publindex (see Figure 1) is a system that 1) automatically retrieves the publication references of a researcher (by accessing different data sources), and filters them by extracting and evaluating their relevant attributes, such as authors, date, publication venue and its correlated historical quality data coming from different quality indexes (index, year, impact factor, categories, quality information), information about citations, etc.; 2) instead of providing yet another ranking, it allows users to predefine their own classification criteria

to cluster publications fulfilling the defined conditions; and 3) it returns the results in the customized way predefined by the user.

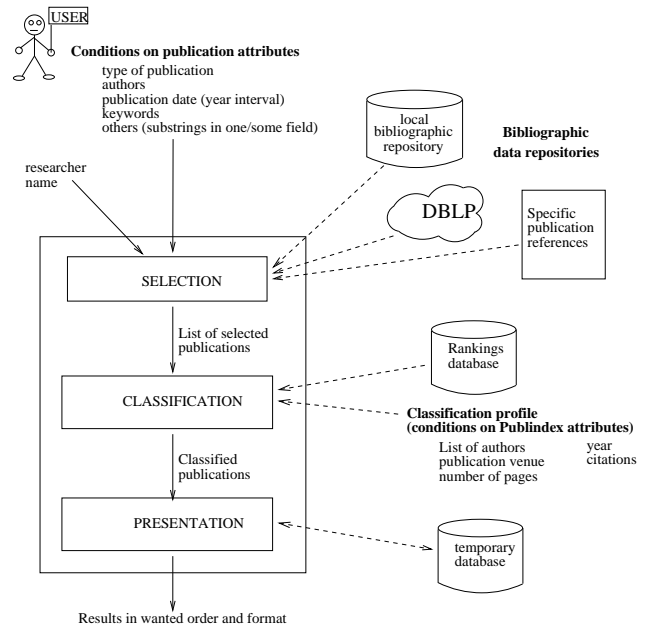


Figure 1: Publindex architecture

2.1 Selection of Publications

The goal of this first step is to select and retrieve, from a specific bibliographic data source, the publications that fulfil the conditions specified by the user on usual search fields (authors, publication type, data of publication, etc.). Regarding bibliographic data sources, we can distinguish two main types: 1) *Publication reference repositories*, from which we can obtain the list of publications of a given researcher (e.g., Web of Science, Scopus, DBLP, Google Scholar, Microsoft Academic, and publication search engines like Publish or Perish, or ResearchGate); and 2) *Journal/conferences rankings*, from which we can obtain different measures to evaluate the relevance of publication venues (e.g., the Journal Citation Reports (JCR) or SJR for journals, GII-GRIN-SCIE (GGS) Conference Rating for conferences, and others like the Computer Research & Education (CORE) rankings for both kinds of publication venues).

Most of these repositories are available on-line, for free or pay-to-access; however, instead of accessing them directly, we advocate storing and unifying such data into a local database¹: 1) to avoid

¹The access to Publindex is granted according to corresponding index copyrights.

This work was supported by the projects TIN2016-78011-C4-3-R (AEI/ FEDER, UE) and DGA/FEDER 2020-2022.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
 JCDL '20, August 1–5, 2020, Virtual Event, China
 © 2020 Copyright held by the owner/author(s).
 ACM ISBN 978-1-4503-7585-6/20/08.
<https://doi.org/10.1145/3383583.3398630>

network communication delays, and 2) to avoid problems when online repositories change or are not available.

2.2 Classification of Publications

Publindex allows users to create publication classes by defining the necessary conditions that a publication must fulfil to belong to each class. For that task, we have developed an interface that guides users 1) to select the parameters and their specific attributes to define conditions on publications, and 2) to define different class profiles by defining boolean expressions on top of the previously defined conditions. See a sample classification profile in the following:

| Conditions | | | | |
|------------|-------------------|-------------------|-------|-----------|
| # | Parameter | Attribute | Index | Condition |
| 1 | authors | number of authors | | <= 5 |
| 2 | publication venue | quartile | JCR | <= Q2 |
| 3 | publication venue | quartile | JCR | >= Q3 |
| 4 | publication venue | class | GGG | = Class 1 |
| 5 | publication venue | class | GGG | = Class 2 |

| Classes | |
|---------------|----------------|
| Class name | Expression |
| Very relevant | 1 AND (2 OR 4) |
| Relevant | 1 AND (3 OR 5) |

To define the conditions, Publindex extracts several parameters (which might have one or more attributes) from the selected publication references: 1) *authors* (researcher position, number of authors, number of relevant authors according to a predefined list), 2) *year of publication*, 3) *number of pages*, 4) *publication venue* (whether it is a relevant venue and/or a relevant publisher according to predefined lists, value/class, top%, quartile, and tertile for each ranking, considering a) the year where the work was published, b) the most recent release of that ranking, or c) the best historical ranking); and 5) *Citations* (number of citations, number of citation from relevant venues or from works in relevant rankings, according to predefined lists).

2.3 Presentation of results

After the publications of a given researcher are selected and classified, the result is presented in the way that the user predefined in the search form: 1) The user can select among different visualization profiles that present each type of publication following a customized syntactical pattern; 2) regarding classification, the user has the option of specifying whether he/she wants to list all the selected publications or just those classified into at least one class; and 3) the user can choose between a) sorting the publications by selecting up to three ordering levels based on publication attributes, or b) sorting the publications according to a given formula that combines and evaluates the specified publication attributes and its given weights to obtain a ranking of the results.

As example, in Figure 2 we can see a partial result of a query, including the matching level color chart, the classification color chart, and a summary of the classification of publications, followed by the (incomplete due to lack of space) list of selected publication references together with their Publindex information about rankings of their publication venues.

3 RELATED WORK

While the evaluation of scholars is frequent in the Academia, there are not so many approaches that offer an automatic evaluation of

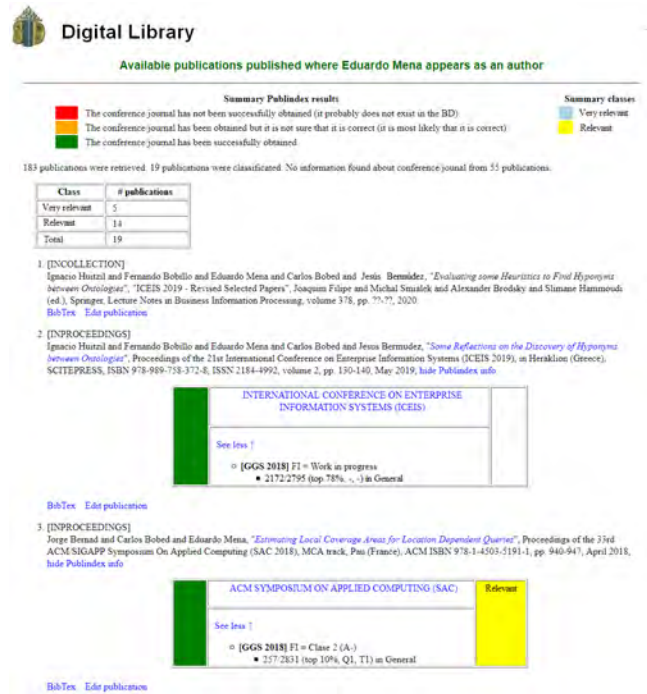


Figure 2: Sample presentation of results

the research publications of candidates, and they are mainly private and closed solutions based on bibliometrics calculated in terms of citations only; no choice to customize our own evaluation criteria.

We could adopt different approaches to improve the disambiguation of author names [1, 3] and publication venues [2] as our goal is the global system rather than the specific submodules used.

4 CONCLUSIONS

To the best of our knowledge, Publindex² is the first attempt to bring together all the main public quality indexes to automatically assess the quality of the publications of a researcher, providing a flexible mechanism to define the criteria that must be used to evaluate publications, which saves time for evaluators as well as improves transparency in the evaluation processes.

As future work, the data obtained, correlated, and returned by Publindex will be exported in different formats including Open Linked Data.

REFERENCES

- Thiago A. Godoi et al. 2013. A relevance feedback approach for the author name disambiguation problem. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2013, Indianapolis, IN, USA, July 22 - 26, 2013*. ACM.
- Denilson Alves et al. 2014. Disambiguating publication venue titles using association rules. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2014, London, United Kingdom, September 8-12, 2014*. IEEE Computer Society.
- Marie Katsurai and Ikki Ohmukai. 2019. Author Matching Across Different Academic Databases: Aggregating Simple Feature-Based Rankings. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*. IEEE.

²The prototype, available at <http://sid.cps.unizar.es/BiD/publindex.html>, was developed as a Web service, exposing its functionalities so it can be integrated into different publication search engines.