

A semi-automatic web based tool for the selection of research projects reviewers

Valeria PUPELLA^a, Maria Eugenia MONTEVERDE^a, Claudio LOMBARDO^b,
Filippo BELARDELLI^c and Mauro GIACOMINI^a

^a*Department of Informatics, Bioengineering, Robotics and System Engineering,
University of Genoa, Via Opera Pia 13, 16145 Genoa, Italy*

^b*Alleanza Contro il Cancro, Via Giorgio Ribotta, 5, 00144, Rome, Italy*

^c*Istituto Superiore di Sanità, Viale Regina Elena 29, 00161, Rome, Italy*

Abstract. The correct evaluation of research proposals continues today to be problematic, and in many cases, grants and fellowships are subjected to this type of assessment. A web based semi-automatic tool to help in the selection of reviewers was developed. The core of the proposed system is the matching of the MeSH Descriptors of the publications submitted by the reviewers (for their accreditation) and the Descriptor linked to the research keywords, which were selected. Moreover, a citation related index was further calculated and adopted in order to discard not suitable reviewers. This tool was used as a support in a web site for the evaluation of candidates applying for a fellowship in the oncology field.

Keywords. Reviewer selection, MeSH, Citation based index, research in oncology

Introduction

Today, new postdoctoral scientists are poorly integrated into the world of research, therefore a trend inversion was deemed crucial in order to promote their research projects at an international level. Hence, a new mobility program, named Training through Research Application Italian iNitiative (TRAIN), was conceived as a part of a strategy carried out by the International Program of the Italian Cancer Network "Alleanza Contro il Cancro" to promote and develop translational mobility in the oncology field at different levels of scientists' research careers [1],[2]. This program is financed by the European Commission Seventh Framework Program (FP7) thanks to a "Cofinancing of regional, national and international programs" (COFUND) initiative which provides fellowships which depend on admission criteria verified, both manually and automatically, by the "Istituto Nazionale per la Ricerca sul Cancro" (INRC). The project has as its main aim the call for proposals for young scientists in oncology and related fields, who should submit specific research projects to be evaluated by independent reviewers.

As many editors of scientific journals well know, the identification of appropriate reviewers for manuscripts and research projects is extremely challenging [3–5].

Consequently, a Web tool to obtain the evaluation and association of reviewers with the relative proposals was developed. One of the most important parts of this tool was an algorithm that partially automated the reviewer selection. Such tool (based on the controlled vocabulary Medical Subject Headings (MeSH) and on a citation

correlated index) attempts to measure both the productivity and impact of a scientist's publications and their correlation with the research fields relating to the presented proposals.

1. Materials and Methods

1.1. Materials

TRAIN foresees three different mobility schemes for scientists, either outgoing or ingoing and the re-integration into the oncology field. In order to receive a fellowship, the applicant must complete five forms and provide a PubMed ID (PMID) list of a maximum of the 20 most relevant publications related to the selected research fields. As a matter of fact, there are 20 different research areas described by as many keywords. Reviewers must carry out this selection as well, however a maximum of five PMID for each keyword can be inserted to provide proof of their expertise. The E-utility Web Service provided by the National Center for Biotechnology Information (NCBI) [6] was used to retrieve significant data about each publication, inclusive of MeSH descriptors and qualifiers. In particular, `eUtilsServiceSoapClient`, `eSummaryRequest` and `SummaryResult` classes were used in order to call the service, request the publication details and retrieve these details. MeSH, as a metadata system, is used to index scientific literature in biomedical fields [7]. It is available for download in various formats, such as XML, but it does not supply any Web Service utility, thus a database has been developed in order to retrieve data from the XML file and for an improved integration with the system created. To achieve a better evaluation of the reviewers' expertise and to present applicants in a more complete form, data acquired with the PubMed E-Utility has been enhanced with their bibliographic information provided by the Web Of Knowledge (WOK), such as the number of citations for each publication, which was manually retrieved. This is highly significant, in ranking scientific literature: indeed, it objectively presents how fundamental a publication is considered.

After taking into account all the requirements, an Entity-Relation diagram (E-R diagram) was designed for the future database development. Particular attention was given to the Publication entity, which merges all the information from the different sources listed above. Thus, the Publication entity contains the PMID, title, number of authors, abstract, and it is linked to the Descriptor and the Qualifier entities, characterised by their MeSH identifiers. This E-R diagram was later implemented, as a database, in SQL Server 2008. A feasibility study was carried out together with the other parties involved and a web-based application was chosen as the interface for the future users of this program. Thanks to the wide diffusion of Internet [8], [9], even on tablets and smartphones, websites are easily reached by users and they can be accessed with different devices. Moreover, since users would probably not be acquainted with the IT, the interface was designed to be appealing and user-friendly, to encourage its utilization, and to supply an intuitive tool for the completion of the forms. Visual Studio 2010 was adopted to implement the platform.

Hence, the evaluating tool is given sufficient data for a complete analysis on both the pertinence of the reviewer's publications to the applicant field and the impact factor of their scientific work so far.

1.2. Methods

Throughout this section, the authors will extensively delineate the setting up of the tool for the management of the publications, especially concerning their relevance and aptness to the selected research field.

After the creation of the web pages dedicated to the application for both reviewers and candidates, a function called "Find_Publication_Details" was developed, in order to retrieve the necessary publication details. This function calls the E-Utility Web Service to obtain a publication summary which encloses numerous items characterised by different types. These items also contain the MeSH Descriptors and Qualifiers of each publication. Since the provided keywords were not defined by MeSH Descriptors and Qualifiers, a tool already developed by the MEDINFO laboratory (10) was utilized in order to achieve these associations and save them into the database, even though not every keyword could be directly linked. Each keyword is saved into a table named "Keywords" together with its unique identifier and, if present, its DescriptorUI. Since reviewers must choose one or more keywords to describe their works and publications, two tables were created to depict this association. The first table is the "Keyword_Exp", where the expert's unique identifier is associated to the selected keywords, while the "Keyword_Pub" table describes the link between the publication and the related keywords. These tables are necessary for the ensuing evaluation steps, since there is a coherence control for each keyword, expert and publication involved. This check is performed by a process involving different automatic and manual steps. If the DescriptorUI for the keyword is present, then the DescriptorUI for the publication is directly compared with the former. Conversely, if the DescriptorUI was not identified for that specific keyword, a web interface, where the keyword itself and the MeSH Descriptors for the publication are shown, was provided. Thanks to this interface, TRAIN administrators could easily check if the selected keyword corresponded to the publication's MeSH Descriptors and therefore accept or reject the reviewer. In both cases, potential reviewers who did not provide relevant publications for the selected research areas were discarded.

Another manual stage was represented by the collection of the number of citations for each publication of the reviewers into the database. Such number was retrieved by the TRAIN administrators from the Web of Knowledge website and subsequently, inserted into a dedicated form of the TRAIN website. The saved numbers were then utilized to calculate a citation related index inspired by the *h*-index, which, in fact, quantifies the scientific impact of a scientist's work [11]. This index was automatically calculated, thanks to the above mentioned number of citations. It represents the number of papers (*n*) which have been cited at least a number (*n*) of times. Therefore, the resulting index could range from 0 to 5, since the reviewers can upload a maximum of 5 papers.

Through another web interface, it was possible for TRAIN administrators to finally accept or deny a candidacy for reviewers. As a matter of fact, the results of the indexes calculation could be visualized and accordingly the reviewer's suitability for the selected keyword could be decided. Moreover, after the acceptance of the reviewers, the system classified them in descending matching order, based on the results of the previous steps. Consequently, the administrators made a manual association of fellowship candidates and reviewers based on referees' rankings, in order to have the most accurate assignment possible.

2. Results

During the first phase of the TRAIN project, an association between research areas provided by the administrators and the MeSH DescriptorUIs was attained for 80% of the keywords. As a matter of fact, for keywords, such as “Medical Oncology”, a direct correspondence, D008495, was found, while for others like “Imaging” a synonym, “Diagnostic Imaging” (D003952), was selected. For the remaining 20%, e.g. “Management, Intellectual Property, Exploitation, Technology Transfer”, a manual approach was used to determine the reviewer’s expertise. The complete set of the 20 research areas can be found at [12].

TRAIN announced four calls for applicants, opening a total number of 26 fellowships consisting of:

- 12 outgoing fellowships, for Italian post-doctoral scientists wishing to improve their research abroad;
- 4 incoming fellowships, for foreign post-doctoral scientists wishing to complete their research studies in Italy;
- 10 re-integration fellowships, addressed to Italian post-doctoral scientists wishing to return to Italy, after research studies in a foreign country.

In order to evaluate the proposals submitted to the website, TRAIN announced an open call for experts who wished to participate. The TRAIN administrators also sent some solicitations to experts indicated by the “Istituto Superiore di Sanità” and the “Organization of the European Cancer Institutes”. Overall, 435 experts took part.

Molecular and Cell Biology presents the highest number of experts registered, 52, while Palliative Care and Rehabilitation and Pharmacogenomics the lowest, 10. The mean value of candidates is 21.75, while the standard deviation is 11.73.

The accepted experts were 205. Among the solicited reviewers, 45 did not upload any publications, as a demonstration of their non-willingness to participate to the project. Since nine research areas had more than 20 experts registered, the aforementioned match algorithm between the keywords and the MeSH Descriptors was carried out in order to level the different categories. For the other fields, it was not necessary. Thanks to this selection, 162 individuals were discarded. 23 out of the remaining 228 did not have a sufficient citation related index. The established threshold was, in fact, 2.

3. Discussion

The evaluation tool was developed in order to obtain a higher objectiveness in proposal reviews. As a matter of fact, thanks to this tool, the choice of reviewers was based on their publications and field of expertise and not only on their candidacy. However, the website administrators could use such a ranking solely as a suggestion and could also choose autonomously the reviewers for each proposal. During the TRAIN project, 340 reviews were carried out; therefore, each expert evaluated at least one proposal. All types of users considered the tool as both valuable and efficient. As a matter of fact, it saved administrators’ time for the associations, and a fewer number of reviewers’ rejections compared to similar tools emerged.

Acknowledgements

The authors would like to thank Moyra Watson for the English revision; Giorgia Pesce, Margherita Marincola, Sergio Bottero, Anna Bogliolo, Germana Gianquinto, Nadia Nasso, Franca Moretti and Angela Guderzo for the administration of the website; Roberta Gazzarata, Susanna Pivetti, Paola Brusini and Daniele Rocca for the implementation; Mikael Raimondo for the graphic interface.

References

- [1] Lombardo C, Albanese D, Belardelli F, d'Alessandro F, Giacomini M, Rondanina T, et al. Training and mobility: a priority for the Organisation of the European Cancer Institutes. How a national mobility initiative could enhance EU cooperation in cancer research contributing to the development of an European Research Area: the example of The I. Tumori [Internet]. [cited 2014 Jan 9];94(2):147–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18564599>
- [2] Lombardo C, Bottero S, D'Alessandro F, Giacomini M, Guderzo A, Moretti F, et al. TRAIN: Training through Research Application Italian iNitiative. Int. J. Biol. Markers [Internet]. [cited 2014 Jan 9];26(2):136–8. Available from: <https://www.ncbi.nlm.nih.gov/m/pubmed/21623583/?i=3&from=/18564599/related>
- [3] Mimno D, McCallum A. Expertise modeling for matching papers with reviewers. Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '07 [Internet]. New York, New York, USA: ACM Press; 2007 [cited 2014 Jan 23]. p. 500. Available from: <http://dl.acm.org/citation.cfm?id=1281192.1281247>
- [4] Dumais ST, Nielsen J. Automating the assignment of submitted manuscripts to reviewers. Proc. 15th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '92 [Internet]. New York, New York, USA: ACM Press; 1992 [cited 2014 Jan 23]. p. 233–44. Available from: <http://dl.acm.org/citation.cfm?id=133160.133205>
- [5] Hartvigsen D, Wei JC, Czuchlewski R. The Conference Paper-Reviewer Assignment Problem. Decis. Sci. [Internet]. 1999 Jun [cited 2014 Jan 23];30(3):865–76. Available from: <http://doi.wiley.com/10.1111/j.1540-5915.1999.tb00910.x>
- [6] Sayers E, Miller V. Overview of the E-utility Web Service (SOAP) [Internet]. National Center for Biotechnology Information (US); 2012 [cited 2014 Jan 9]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK43082/>
- [7] Fact Sheet Medical Subject Headings (MeSH®) [Internet]. U.S. National Library of Medicine; [cited 2014 Jan 9]. Available from: <http://www.nlm.nih.gov/pubs/factsheets/mesh.html>
- [8] Vecchio DT, Gazzarata R, Alloisio S, Giacomini M. From the research to the users Prevention & IT. Int. J. Evid. Based Healthc. 2009;7(3):226.
- [9] Giacomini M, Pastorino L, Soumetz FC, Mielczarski JA, Mielczarski E, Rangelow I, et al. Data Modeling for Tools and Technologies for the Analysis and Synthesis of NANOstructures (TASNANO) Project. J. Inf. Technol. Res. [Internet]. IGI Global; 2009 [cited 2013 Nov 25];2(3):49–70. Available from: <http://www.igi-global.com/article/data-modeling-tools-technologies-analysis/4142>
- [10] Gazzarata R, Pivetti S, Giacomini M. A mesh-based health care prevention information retrieval system. NETTAB 2012 “Integrated Bio-Search.” 2012. p. 18–20.
- [11] Hirsch JE. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2005 Nov 15 [cited 2014 Jan 9];102(46):16569–72. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1283832&tool=pmcentrez&rendertype=abstract>
- [12] Train Research Areas [Internet]. [cited 2014 Jan 31]. Available from: <http://trainoncology.eu/ResearchAreasSUBAbouttrain.aspx>