CrossMark

Topic identification challenge

**Kevin Boyack[1] · Wolfgang Glänzel[2] · Jochen Gläser[3] ·
Frank Havemann[4] · Andrea Scharnhorst[5] · Bart Thijs[2] ·
Nees Jan van Eck[6] · Theresa Velden[3,7] · Ludo Waltmann[6]**

Over the last two years, a group of researchers used a shared dataset in order to compare
their approaches to the identification of thematic structures in a set of 111,616 papers on
astronomy and astrophysics published in 59 journals between 2003 and 2010. The out-
comes of this comparative exercise are published in a special issue of *Scientometrics*

✉ Jochen Gläser
Jochen.Glaser@ztg.tu-berlin.de

Kevin Boyack
kboyack@mapofscience.com

Wolfgang Glänzel
wolfgang.glanzel@kuleuven.be

Frank Havemann
frank.havemann@ib.hu-berlin.de

Andrea Scharnhorst
andrea.scharnhorst@dans.knaw.nl

Nees Jan van Eck
ecknjpvan@cwts.leidenuniv.nl

Theresa Velden
velden@ztg.tu-berlin.de

Ludo Waltmann
waltmanlr@cwts.leidenuniv.nl

[1]   SciTech Strategies, Inc., Albuquerque, NM 87122, USA

[2]   ECOOM and Department of MSI, KU Leuven, Louvain, Belgium

[3]   ZTG, TU Berlin, HBS1, Hardenbergstr. 16-18, 10623 Berlin, Germany

[4]   Institut für Bibliotheks-und Informationswissenschaft, Humboldt-Universität zu Berlin,
      Dorotheenstr. 26, 10099 Berlin, Germany

[5]   DANS-KNAW, Anna van Saksenlaan 51, The Hague, The Netherlands

[6]   Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands

[7]   University of Michigan, School of Information, Ann Arbor, MI 48109, USA

⚛ Springer

(Gläser et al. 2017). Now that *Clarivate Analytics* kindly agreed to make this dataset available to interested researchers in the bibliometrics community, we suggest to extend this comparative approach.

We challenge you to participate in the comparative topic identification exercise.

The challenge is not to develop the best partitioning of the dataset. We believe this to be impossible because there is not one single best solution. Instead, we challenge you to gain as much information as possible about your own approach and the reasons why it produced a particular solution, and how it compares to solutions produced by other approaches. We challenge you to comparatively discuss advantages and disadvantages of approaches to topic identification and thus to contribute to a cumulative body of knowledge on the suitability of data models and algorithms for the identification of topics.

Thanks to *Clarivate Analytics*, we are able to offer access to the dataset with an efficient license agreement.

While participating in the "Web of Science comparative topic identification exercise," you will be provided with access to the Clarivate Analytics "Web of Science comparative topic identification exercise" dataset. You may access and use this dataset from March 1, 2017 through December 31, 2018 only for the exercise above, subject to the "Clarivate Analytics Terms", including the "Web of Science: Custom Data Set Product Terms" in the "Product/Service Terms", available on our Terms of Business site: http://clarivate.com/tob/. By accessing and/or using our data, you are legally bound by and hereby consent to these terms. If you do not agree to these terms, then you may not access or use our data. Any extension or further use of our data beyond December 31, 2018 is strictly prohibited unless you receive prior written permission from Clarivate Analytics.

The dataset can be obtained by sending an email to jason.rollins@thomsonreuters.com.

We will also offer access to a website where solutions can be deposited and downloaded for comparisons. The website, which also offers some tools for a comparative analysis of solutions and individual clusters, is www.topic-challenge.info.

If there are enough participants, we will run sessions on the comparative exercise at the next ISSI conferences and dedicated workshops.

We hope that many of you will take up the challenge and thus contribute to cumulative progress in bibliometrics.

Reference

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data—Different results? Towards a comparative approach to the identification of thematic structures in science. In J. Gläser, A. Scharnhorst & W. Glänzel (Eds.), *Same data—Different results? Towards a comparative approach to the identification of thematic structures in science*. Special Issue of *Scientometrics*. doi:10.1007/s11192-017-2296-z.