European Commission

Science for Environment Policy

March 2016
Thematic Issue 54

Research for Environmental Policymaking

Subscribe to free weekly News Alert

Source: Boyack, K.W., Klavans, R. (2013). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Science and Technology* 65(4): 670 - 685. DOI 10.1002/asi.22990

Contact:
kboyack@mapofscience.com

Read more about:
Environmental information services

Creating a map of science: a visual representation of global research

A map of science could assist research planning strategies by helping to identify emerging topics. The map — which is based on links to almost 20 million scientific articles that have been published over the past 16 years — clusters and links scientific disciplines by citation-based relationships and serves as a highly detailed and scalable infographic. The authors hope it will be used by research planners to help distinguish — and potentially forecast — the research areas in science which have longevity, and also those which are innovative.

Creating a map of science requires a combination of classification and visualisation. Classification takes place by data mining and then partitioning sets of documents into discipline-based groups, such as chemistry, medical science, earth science, etc., and determining the relationships between these groups. This information is then used as input to a system which can visually simulate the information in an easily digestible way.

The first attempt to make a worldwide map of science using citation-based techniques in the 1970s used 1 310 highly cited references from a single year. The next step was taken in 1999, when a four-level map was created from 130 000 reference papers. In 2006, co-citation models were made from 700 000 papers, and throughout the rest of the decade, several models were created using data sets of millions of citing papers spanning several years. This particular map draws on nearly 20 million scientific documents.

The model is based on a co-citation technique, which is a way of determining similarities and relationships between articles. The thresholds that were used to determine co-citation sources were adjusted to achieve greater scope — from the most emergent to the most persistent topics — and the technique was combined with bibliographic coupling to find relevant articles that co-citation methods alone may have missed. The process was enhanced with an additional step that directly analysed the texts of the articles for relevant keywords.

The visual mapping of the data set was done by the OpenOrd algorithm to draw intricate layouts of hundreds of thousands of groups. Inputs to the algorithm were further tweaked to achieve the best visual communication of the data, which resulted in an integrated multicoloured map with networked strands, which can be viewed as a whole or scaled up to specific levels of detail.

The map can help define the characteristics of certain topics within science. For example, 'superconductivity' and 'elementary particle physics' have low growth rates, long durability and are related strongly to specific disciplines, whereas 'environmental, energy and economic policy' and 'sleep' have high growth rates, are more temporary, and are multidisciplinary. As a general rule, growth rates tend to correlate inversely with stability, and just under one third of papers published each year are not followed up the next year.

The authors note that much could be done to improve and expand the model[1] to make it more globally representative. Documents regarding patent applications have only recently been added, and the authors suggest that the inclusion of scientific and technical databases in other languages, notably in Chinese, might dramatically change the layout of the map. The addition of information about funding sources would also better enable the model to forecast in which direction research is heading.

1. In 2014, the same researchers more than doubled their original map. They compared two maps of science, one simply based on source documents, and the second, which uses secondary source items (such as journal articles from non-indexed sources, conference papers, books and government reports, software, and even newspaper articles) found on Scopus, over a 16-year period. The map has now expanded from 19 to 43 million documents. Boyack, K. W., & Klavans, R. (2014). Including non-source items in a large-scale map of science: What difference does it make? *Journal of Informetrics*, 8, 569–580.