

Profiling Science & Innovation Policy by Object-based Computing

Zhang Zhixiong, Liu Jianhua, Xie Jing, Zou Yimin
National Science Library, Chinese Academy of Science, China
{zhangzhx, liujh, xiej, zouyimin}@mail.las.ac.cn

Abstract:

Myriad named entities such as science strategies and policies, key initiatives & research programs, key research institutes researchers and scientists are embedded in many web pages from various science & innovation institutes. The authors refer to these named entities about science & innovation activities as research objects. Usually, these research objects carry the core information of the web pages and are valuable for automatically extracting intelligence from web pages. Hence, It is one of the most important questions that how to dig these knowledge units from these resources and how to use the knowledge to support deep intelligence analysis. In this paper, the authors bring forth a method using object-based computing for profiling science & innovation policies of some key national scientific administrative offices, research councils, funding agencies, and leading research institutes.

After collecting related web pages from the targeted institutes, the authors firstly extract the research objects and their relationships from these web pages. Different from other schemes, the authors add the temporal features of objects when modeling these extracted objects because of the dynamic feature of web resources. Take "July 13, 2010, White House Announces National HIV/AIDS Strategy" for example, the authors transfer the research objects in this sentence into following time-stamped models: (object type, object value, time stamp) → Object A: (Organization, White House, July 13, 2010), Object B: (Strategy, National HIV/AIDS Strategy, July 13, 2010); (Object A, Object B, Relationship Type, Time Stamp) → (White House, National HIV/AIDS Strategy, Announces, July 13, 2010). All the extracted objects are preserved in relation Database. These temporal objects and their relationships representation model can be extended to succedent content analysis tasks involving a temporal dimension, such as novelty object detection and burst topic detection.

As it mentioned above, research objects usually carry the core information of the web pages about science & innovation policy. For example, co- occurrence of three kinds of research objects, which contain Present of America, science roadmap and America, in the title of certain web page, may lead to the high intelligence value of this page. So, based on the structured extracted and sensitive research objects, the authors present a new model to judge the intelligence value of the collected web pages about science & innovation policies for supporting Chinese scientific policy-makers (figure 1). Besides, the authors identify new important science & innovation policies, and classify the policies into more detailed categories such as formal scientific declaration, strategic plan, R&D budget, organization restructure etc.

However, Research profiling requires more fine-grained analysis such as important institutes, important persons and hot topics in current science & innovation. Hence, the

authors use those objects with temporal information again to identify important topics, important research activities, important persons within the targeted institutes, monitor the novelty activities or plans, outline the hot topics in a period of time, depict current active activities in institutes, and cluster the related policies for the institutes.

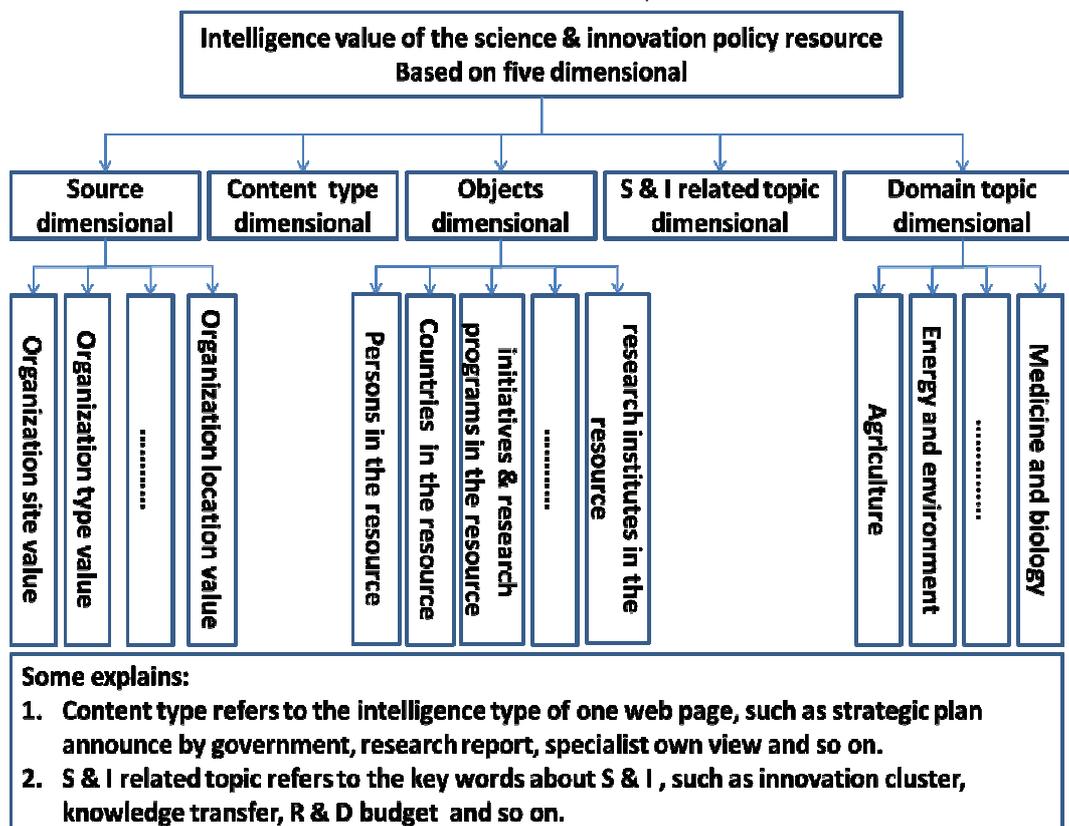


Figure 1 computing the value of pages based on objects

As an application, the authors implement a research profiling system which monitors websites of 86 science & innovation authority organizations, such as Office of Science and Technology Policy, Research Councils UK and Department for Business, Innovation and Skills of UK, to show the effectiveness of profiling science & innovation policy in those institutes by object-based computing. All these organizations are chosen by science & innovation policy intelligence experts, In this system, it pushes the newest important science & innovation policy web resources to target intelligence experts every day , implements automatically classification of these resources, monitors the important topics and objects in recent one month, shows developing tread of the certain topics and objects within target organization and so on. According to the testing token by some intelligence experts, the method presents a good performance.

Keywords: Research Profiling; Science & Innovation Policy; Knowledge Extraction; Research Objects; Object-based Computing.