

# Profiling Science and Innovation Policies of Obama Administration by Mining OSTP Web Resources

Zhang Zhixiong, Liu Jianhua, Zou Yimin, Xie Jing, Qian Li, Guo Hongmei

National Science Library, Chinese Academy of Sciences, Beijing, China

Email: zhangzx@mail.las.ac.cn

## Introduction

A project has been carried out in the Chinese Academy of Sciences to support automatically monitoring science, technology and innovation policies based on web resources released from some key institutes (such as some national administrative offices, research councils, funding agencies, and leading research institutes). The Office of Science and Technology Policy (OSTP) is one of the most important institutes for information analysis in CAS since it bears the responsibility to lead interagency efforts to develop and implement science and technology policies for the United States. In this paper, the authors try to bring forth a sound method to mine the web resources released from OSTP Web site to profile the Obama administration science and innovation policy.

## Data sources

By using the automated web information monitoring system developed by the project, the authors gather 582 web resources (including 150 PDF files and 432 HTML files) released from the OSTP Web site, published from Feb 17, 2009 to Apr. 12, 2013.

## Methods

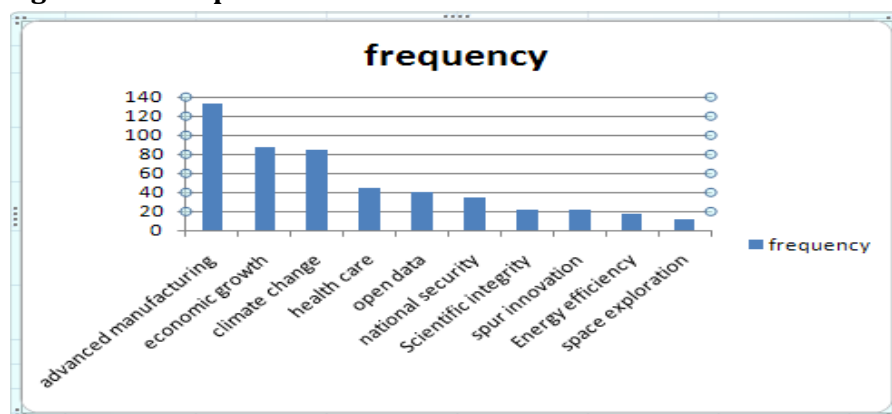
After collecting the data from OSTP Web site, the authors use knowledge extraction tool to automatically analyze the PDF files and web pages to extract key terms and objects (named entities such as important strategies, major projects, significant research reports, R&D spending etc.) from the content and turn the free text into a computable knowledge base, perform topic clustering based on extracted terms and objects, identify key policy activities and key topics within the OSTP, and analyze the hot research activities and hot topics within a period. With the help of other information analysis tools, the authors perform the three kinds of analysis of science and innovation policy taken by the Obama administration. (1) *The key reports analysis* - With the help of the link analysis, the authors identify the most important and influential reports released from the OSTP Web site within the last 4 years, analyze the topics of those reports and perform clustering analysis to identify the relationship between those reports. (2) *The policy focuses analysis* - By using the extracted key terms and objects from the web resources, the authors identify the key topics of the policy in OSTP Web site and depict the hot strategies, major initiatives, important projects in a period. (3) *Inter-relationship analysis* - Based on the gathered web resources, identified key reports, identified key topics, extracted terms and objects, the authors perform a network analysis to reveal the relationship between them, disclose inter-relationship within science and innovation policies of the Obama administration.

## Results

*The Key Reports Analysis* - With the help of Majestic SEO tools, the authors figure out the 15 most influential reports from the OSTP website, which include “Increasing Access to the Results of Federally Funded Research”, “National Space Policy of the United States of America”, “A Policy Framework for The 21st Century Grid : Enabling Our Secure Energy Future” etc. After topic analysis, the authors find the topics of those reports include Open Data Policy, Space Policy, Bio-Economy, STEM Education, Genome, Advanced Manufacturing, Environmental Capital, CyberSecurity, Innovation Strategy, Open Innovation etc. STEM Education is the most concerned topic, since 3 reports from different publish time care about this topic.

*The Policy Focuses Analysis* - By using the extracted key terms and objects from the web resources. The authors find hot topics expressed by extracted term in OSTP Web site include Advanced Manufacturing, Economic Growth etc. As could be seen in figure 1.

**Figure 1: Hot topic in OSTP Web site**



The authors also find the hot topics of each year are changing. For example, “Advanced Manufacturing” becomes a hot topic in 2011 and popular in 2012, while “open data” emerges as hot topic in 2012, and it become the hottest topic in 2013.

By extracting initiatives or projects from the OSTP web resources, the authors figure out the most important initiatives OSTP developed include Smart Grid, Materials Genome Initiative, National Nanotechnology Initiative, Methane Opportunities for Vehicular Energy, US Global Change Research Program, Open Government Initiative, Advanced Manufacturing Initiative etc.

*Inter-relationship Analysis* - The authors perform a network analysis to disclose inter-relationship within OSTP web site. Take “Advanced Manufacturing Initiative” as a example. The author find the foundations related to the initiative include NSF, Information Technology and Innovation Foundation; organizations related to the initiative include Harvard, MIT, Department of Energy, NIST etc; projects related to the initiative include Manufacturing Technology Program, Nanoelectronics Research Initiative, National Nanotechnology Initiative etc.