

The Web as a Laboratory

Bebo White
SLAC National Accelerator Laboratory
2575 Sand Hill Road, MailStop 97
Menlo Park, California 94025-7015
USA
+1 415 680-3356
bebo@slac.stanford.edu

ABSTRACT

Insights from Web Science and Big Data Analysis have led many researchers to the conclusion that the Web not only represents an almost unlimited data store but also a remarkable multi-disciplinary laboratory environment. A new challenge is how to best leverage the potential of this experimental space. What are the procedures for defining, implementing and evaluating “Web-scale” experiments? What are acceptable measures of robustness and repeatability? What are the opportunities for experimental collaboration? What disciplines are likely to benefit from this new research model? The Web Laboratory model provides an exciting new and fertile model for future research.

Categories and Subject Descriptors

H.5.3 [Group and Organization Interfaces]: Web-based interaction.

Keywords

Web Science, Big Data, Analytics, Crowdsourcing, Security

1. INTRODUCTION

This keynote address describes the potential of the World Wide Web as a medium for scientific (including the social sciences) experiments. Crowdsourcing techniques and application programming interfaces (APIs) to established social networks provide researchers the capability to build global population sets for experimentation. Unique challenges in such experiments lie in the areas of experimental design, results analysis, security, and human-computer interaction. In addition, there are ethical considerations and elements of economics and psychology.

The Web offers numerous advantages over the traditional laboratory for the execution of experiments. First, the Web allows researchers to conduct experiments on scales differing by orders of magnitude. However, conducting experiments on the Web also involves limited control that could potentially impact result robustness and repeatability.

Examples of “Web-scale” experimentation to be described include crowdsourcing data collection using surveys and contextual mining of massive datasets. Experiment venues include Twitter, Amazon Mechanical Turk, Ebay, Facebook, and others.