

An infrastructure for data mining public chemical & biological information

David J. Wild¹ and Rajarshi Guha¹

¹Indiana University School of Informatics, Bloomington, Indiana, USA.

At Indiana University School of Informatics, we are exploring several projects aimed at increasing the accessibility and usability of the increasing volumes of chemical and biological information that are readily available on the web, in public databases, and in accessible documents. In this presentation, we will describe our progress with and results for each of the projects and we will detail our vision for how these projects might together enable better data mining of public information in drug discovery projects. The projects that will be described include: public information access using aggregate compound information web services; automatic generation of workflows of web services using semantic web techniques; calculation of chemical similarity between web documents; mining of information in chemistry journal articles; and network models to integrate compound, protein and assay information.