Editorial

XMPP Supports the Discovery and Invocation of Asynchronous Web Services in Bioinformatics via Cloud Computing

Robert Alison*

Department of computer science Angola

*Corresponding Author's E-mail: Robert_a@55gmail.com

Abstract

The provision and analysis of data in the life sciences are heavily reliant on the internet. However, in order to be effective, machine-friendly interfaces are necessary due to the expanding amount of data and analysis tools available. In bioinformatics, the most prevalent HTTP-based Web service technologies are Representational State Transfer (REST) services and Simple Object Access Protocol (SOAP). However, these methods have significant drawbacks, such as the inability of services to send status notifications and the lack of discoverability. Although a number of complementary workarounds have been proposed, the resulting ad hoc solutions can be cumbersome and of varying quality With its extensions, XMPP is a powerful protocol for cloud services that has a number of advantages over traditional HTTP-based Web services. These advantages include the following: services are discoverable without the need for an external registry; asynchronous invocation gets rid of the need for ad hoc solutions like polling; and input and output types that are defined in the service make it possible to generate clients on the fly without the need for an external semantics description. XMPP is a highly intriguing option for the next generation of online services in bioinformatics due to its numerous advantages over existing technologies.

INTRODUCTION

The internet is frequently utilized in the life sciences for data access and computational analysis. Web services provide machines with this functionality, whereas HTML-only web pages do not. In bioinformatics, the possibility of having two or more machines interact to solve problems expands computing beyond the scope of manually visiting web pages (Sugawara H, 2003). For instance, many institutions, including those that are members of the International Nucleotide Sequence Database Collaboration, provide services that can be retrieved, submitted, or analysed using bioinformatics data and are machine-readable for the general public. Several tools, including scripting environments like and workflow environments like and Cyrille2, have recently were introduced to aggregate services. The successful experiment in which microarray data genomic sequence information (Oinn T, 2004). And pathway databases were integrated into a workflow to assist in the search

for candidate genes responsible for phenotypic variation exemplifies this aggregation of services' great potential. SOAP and REST are the most widely used technologies that underpin life science Web services. Computer messages like assignment input data, computation results, and database search output are formalized by these technologies (Hoon S, 2003). And SOAP is two widely used protocols in the life sciences; however, the initial version incompatibilities and complexity of SOAP-implemented libraries have made REST an increasingly popular alternative. REST is able to send any type of data around, whereas SOAP and wrap the data in XML envelopes. Although the lack of semantics makes REST less formal and easier to implement and use, writing general clients for it is also challenging. Although it is intended to formalize REST services to address this issue, the Web Application Description Language (WADL) is not yet widely supported. And DAS are prominent projects in bioinformatics that make use of the REST technology to provide data (Fisher P, 2007). The numerous design flaws of

SOAP-based services, which are prevalent in the field, pose a problem in bioinformatics. First, even though this isn't required by the specification, SOAP services typically use HTTP as their communication channel. The original purpose of HTTP was to allow for the querying and retrieval of web pages, not for more complex communication, as is common in bioinformatics. Server timeouts and firewall issues are common obstacles, making the intrinsically synchronous HTTP protocol unsuitable for time-consuming operations like computationally demanding database lookups or calculations. Implementing a ticketing mechanism in the service, in which the client receives a ticket that can be used to repeatedly poll for results, is a common workaround (Wilkinson MD, 2002). Due to the fact that these services implement the ticketing mechanism in a manner that is not standardized, this is not only inefficient but also makes it impossible to create a general client for these services. Second, unlike an XML Schema document, SOAP services do not specify which data types are supported as input and output. This has been addressed by a separate specification, the Web Service Description Language (WSDL), which is similar to WADL (Smedley D, 2009). As a result, two technologies are required instead of just one. Thirdly, and finally, a complementary technology known as Universal Description Discovery and Integration (UDDI) has been developed for the purpose of service discovery in addition to the SOAP specification. UDDI is not used in bioinformatics, according to our knowledge (Vaquero LM, 2009).

Bio-and chemoformatics cloud services

As can be seen in, xws4j was used to create XMPP services for chemoformatics and bioinformatics. Using a proteochemometric model, a service called HIV Pred was created to predict the susceptibility of sequences for seven known HIV protease inhibitors (Kottmann R, 2008). A Resource Description Framework service (RDF Service) has been added to demonstrate that RDF and XML worked well together, making it possible to incorporate RDF into on-going life sciences semantic web efforts. The Chemistry Development Kit can be used to perform a variety of chemoformatics algorithms with the help of two XMPP services: the CDK Service and the Descriptor Service. Small molecules can be given a variety of straightforward molecular properties and their 2D and 3D coordinates thanks to the CDK Service. In particular, larger molecules can take a long time to calculate 3D coordinates, so it is highly recommended to invoke those asynchronously. The Descriptor Service returns molecular descriptors, which are numerical representations of the chemical structure that reflect a specific property; it can, for instance, describe the molecule's shape or the presence of particular functional groups like amine and acid groups. While calculating multiple descriptors for a single molecule can take time, calculating a single descriptor typically takes less time (Lapins M, 2008). The chemoformatics administrations apply the Substance Mark-up Language (CML) to characterize the administrations' semantics. CML offers a desirable level of interoperability with other applications and resources due to its widespread use.

DISCUSSION

With high throughput technologies and large, complex information sources, bioinformatics of today necessitates not only user interfaces but also programmatic interfaces that enable machines to access these services. In bioinformatics, XMPP and the newly developed extension IO Data outperform existing remote service technologies. Bioinformatics services like querying large or multiple databases and computationally intensive calculations like identifying protein domains, families, and functional sites, as well as sequence alignments, take more than a few seconds to complete. The XMPP would directly benefit these services because it eliminates all custom polling mechanisms, which are typical of long-running SOAP-based services (Steinbeck C, 2003). Ad hoc solutions to a problem inherited from HTTP are workarounds such as "ticketing" or "polling." Because XMPP services are asynchronous by nature, they completely remove this obstacle. This is true even if the bioinformatics operation itself takes a long time, but it could also be because of a large job queue or a heavy server load.

CONCLUSION

The Bio-Catalogue and Bio-Moby registries have also addressed this issue. By defining an ontology of scientific data types and annotating services using these data types, Bio-Moby also addressed the lack of strong data typing. The need for services that are simple to use is also reflected in the growing use of REST services in bioinformatics; however, REST does not provide the robust data typing that is required in the life sciences.

REFERENCES

1. Sugawara H, Miyazaki S (2003). Biological SOAP servers and web services provided by the public sequence data bank. Nucleic Acids Res. 31: 3836–3839.

2. Oinn T, Addis M, Ferris J, Marvin D, Senger M et al (2004). Taverna a tool for the composition and enactment of bioinformatics workflows. Bioinformatics. 20: 3045–3054.

3. Hoon S, Ratnapu KK, Chia J, Kumarasamy B, Juguang X et al (2003). Biopipe a flexible framework for protocol-based bioinformatics analysis. Genome Res. 13: 1904–1915.

4. Fisher P, Hedeler C, Wolstencroft K, Hulme H, Noyes H et al (2007). A systematic strategy for large-scale analysis of genotype phenotype correlations: identification of candidate genes involved in African trypanosomiasis. Nucleic Acids Res. 35: 5625–5633.

5. Wilkinson MD, Links M (2002). Bio-MOBY: an open source biological web services proposal. Brief Bioinform.3: 331–341.

6. Smedley D, Haider S, Ballester B, Holland R, London D et al (2009). Bio-Mart-biological queries made easy. BMC Genomics.10: 22.

7. Vaquero LM, Rodero-Merino L, Caceres J, Lindner M (2009). A break in the clouds: towards a cloud definition. SIGCOMM Comput Commun Rev. 39: 50–55.

8. Kottmann R, Gray T, Murphy S, Kagan L, Kravitz S et al (2008). A standard MIGS/MIMS compliant XML Schema: toward the development of the Genomic Contextual Data Markup Language. (GCDML) OMICS. 12: 115–121.

9. Lapins M, Eklund M, Spjuth O, Prusis P, Wikberg JES (2008). Proteochemometric modeling of HIV protease susceptibility. BMC Bioinformatics. 9: 181.

10. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E et al (2003). The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics. J Chem INF Comput Sci. 43: 493–500.