

Penta and Hexa Statements are Suitable for Representing Omics Knowledge on the Semantic Web

Norio Kobayashi and Tetsuro Toyoda

Bioinformatics And Systems Engineering division (BASE), RIKEN, JAPAN
toyoda@base.riken.jp**

1 Abstract

A triple or a quad in RDF is suitable to describe a relationship among a subject, predicate and object. In a knowledge management of omics data, however, a predicate is often used to describe a relationship between a part of subject or object rather than whole of subject or object. For example, a statement that ‘a feature (subject) exists within (predicate) a genomic DNA sequence (object)’ requires the positional information of the feature within the genome. Therefore, an additional URI that specifies the feature’s position within the object should be added to the RDF statement to form a penta statement composed of five URIs representing the statement itself, its subject, predicate, object and object’s region that specifies the position within the object. Similarly, a hexa statement composed of six URIs representing the statement, its subject, subject’s region, predicate, object and object’s region is conceivable as a useful representation to associate a part of a subject with a part of an object. Here we show penta and hexa statements are suitable for representing knowledge among huge volumes of omics data. We also demonstrate an alternative way of representing a hexa statement that is defined as an extension of conventional quad RDF, in which we define the format of the URI of subject or object as a combination of the subject’s or object’s URI and the internal coordinates that specify the parts within the subject or object, respectively. The extended RDF with penta and hexa statements are implemented in the Semantic Web Folders (SWF for short) system which is an information platform to integrate various life science databases and thus providing a collaboration tool among researchers to develop new databases. SWF is available at <http://database.riken.jp>.

** To whom correspondence should be addressed.