Chapter 2

Modelling Molecular Biological Information: Ontologies and Ontological Design Patterns

Jacqueline Renée Reich
University of Manchester, United Kingdom

The amount of available information in molecular biology is vast due to genome sequencing and gene expression chips. Nowadays, the challenge is to represent and manage the static and dynamic properties of DNA sequence data or annotated information to decipher the structural, functional and evolutionary clues encoded in biological sequences. Therefore, molecular biologists build and use ontologies to represent parts of the molecular biological terminology and to provide a model of biological concepts. Ontological Design Patterns (ODPs) provide a technique to increase the flexibility and reusability of these biological ontologies. There are many useful features of ODPs: 1) they describe simple, flexible and reusable solutions to specific design problems, 2) they can be defined and applied within informal or formal ontologies, 3) they make design approaches transferable, clarify the architecture, and improve the documentation of ontology-based knowledge systems, and 4) they form a framework to deal with different bioinformatics tasks, such as accessing and managing heterogeneous molecular biological databases, analysing scientific texts, or annotating sequence databases. Most of the ODPs are informally described in (Reich, 1999). All ODPs with code examples are available from the author.

INTRODUCTION

Molecular biology encompasses many disciplines each with their own terminology and nomenclature. It has a vast range of information sources and uses various analysis tools with specific semantics. Genome sequencing and gene expression chips are producing a large amount of information, causing the challenge to represent and manage the static and dynamic properties of DNA sequence data or annotated information. Nowadays, the structural, functional and evolutionary clues encoded in biological sequences are still to be deciphered.

Parts of the content of molecular biological knowledge have to be specified and formalised to become machine-readable and be represented on the computer. To organise the molecular biological vocabulary and to provide a model of biological concepts and semantic relationships, molecular biologists build and use hierarchical ontologies (Altman et al. 1998). Ontologies may constitute knowledge bases and semantic frameworks to improve 1) the specification of existing or new database schemata, 2) the matching of data within heterogeneous databases, and 3) the sharing and reuse of knowledge. The matching of heterogeneous data presupposes the collaboration of software components in which, for instance, queries and assertions can be exchanged. Assembling reusable components of precisely specified knowledge can prevent the rebuilding of future knowledge segments from scratch.

To increase the adaptability of an ontology to future demands and to avoid time consuming and expensive redesign, abstractions that emerge during design must be incorporated into the ontology (Guarino, 1997). This situation is comparable to software design, where the application of object-oriented design patterns is emphasised (Coad, 1992; Gamma et al., 1994). At this high level of abstraction

\[\begin{array}{|c|c|}
\hline
\text{TerminologicalHierarchy ODP} & \text{to compose concepts and instances into part-whole hierarchies} \\
\text{UpdateDependencies ODP} & \text{to notify interdependent concepts and instances if one concept is changed} \\
\text{DefinitionEncapsulation ODP} & \text{to define a family of definitions, and encapsulate each one} \\
\text{ExpressionComposer ODP} & \text{same construction process for different concepts} \\
\text{ChainOfExpressions ODP} & \text{multiple concepts can fulfill a context} \\
\text{InteractionHider ODP} & \text{concept encapsulates interaction of concepts and instances} \\
\text{TerminologyOrganizer ODP} & \text{interface creates and organises concepts or instances, realised by sub-concepts} \\
\text{Mask-ODP} & \text{to represent complete sub-concepts as instances} \\
\text{UnspecificTerm-ODP} & \text{to manage unspecific instances at fine granularities} \\
\text{AddDynamicInfo-ODP} & \text{to add information and meaning dynamically} \\
\hline
\end{array}\]
Related Content

Mining Users' Intents from Logs
[www.irma-international.org/article/mining-users-intents-from-logs/126306/](www.irma-international.org/article/mining-users-intents-from-logs/126306/)

VIPER: Evaluation of an Integrated Group VoiceIP Software Application for Teaching and Learning in Higher Education
[www.irma-international.org/chapter/viper-evaluation-integrated-group-voiceip/29452/](www.irma-international.org/chapter/viper-evaluation-integrated-group-voiceip/29452/)

Automated Ranking of Relaxing Query Results Based on XML Structure and Content Preferences
[www.irma-international.org/article/automated-ranking-relaxing-query-results/55060/](www.irma-international.org/article/automated-ranking-relaxing-query-results/55060/)

Context Inference Engine (CiE): Classifying Activity of Context using Minkowski Distance and Standard Deviation-Based Ranks
[www.irma-international.org/chapter/context-inference-engine-cie/108811/](www.irma-international.org/chapter/context-inference-engine-cie/108811/)

Composition of Optimal Service Workflows with Quality-of-Service Enabled Multi-Criteria Uniform Cost Search Algorithm