Ontologies and Processing Patterns for Microarrays

Mónica Miguélez Rico
University of A Coruña, Spain

José Antonio Seoane Fernández
University of A Coruña, Spain

Julián Dorado de la Calle
University of A Coruña, Spain

INTRODUCTION

The researchers currently have a new tool for dealing with the solution of biomedical problems: the Microarrays. These devices support the study and the acquisition of information related to many genes at the same time by means of a unique experiment, providing multiple potential applications such as mutation detection of microorganism identification.

Some of the problems that exist when working with this type of technologies are the high number of data and the complex technical nomenclature to be dealt with. These facts imply the need of using several standards and ontologies when performing this type of experiments.

BACKGROUND

The microarrays have been a key element in the biotechnological revolution of the last years; however new problems regarding both, data handling and statistics analysis, have arisen due to the vast volume of information and to the structure of the data used.

The main concern lies in the vast amount of data to be stored, processed and analysed. Besides, as the microarrays are a new technique, most of the methods, protocols and standards are still being defined.

The fact of dealing with such amount of unstructured information leads to believe that is quite difficult for the descriptors of the stored concepts or their units to be the same at the different data bases where it is accessed. In order to support the vocabulary unification task, the ontologies (Chandrasekaran, 1999) enable a hierarchical definition of concepts for framing the schemas of the accessed data bases. There are fully established ontologies also quite used as the UMLS medical vocabulary (UMLS, 2006), that has information about symptoms and illnesses, or the GO (Gene Ontology) genomic ontology (Gene Ontology, 2006), regarding information about the function and the expression location of the different human genes.

Once the use of ontologies has been established, they are also quite useful for searching hidden relationships among data. Consultations with SQL-type (Structured Query Language) (Beaulieu, 2005) query languages may be performed in an ontology and translated to query languages owning to each underlying data base. In this way, by the use of the ontology, it could be known that the presence of fever is a symptom and which are the illnesses that present fever as a symptom.

Currently, there are special data formats in medicine science as the DICOM standard (Oosterwijk, 2001) for storage and transfer of the increasing amount of medical images that support new imaging modalities. Nevertheless, the typical biomedical images, as the microarrays or the DNA gels, are not currently considered at DICOM, although their future integration is foreseeable in incoming revisions, as the clinical test based on these techniques might be increasingly used in routine medical practice. At the moment, however, the management of this type of images is quite sensitive.

MAIN FOCUS OF THE CHAPTER

This paper presents a description of the most important standards and ontologies for working with microarrays experiments; it also tackles the integration options of some of these ontologies and standards into an information system for managing microarrays.

The first standardisation initiatives appeared in 1998. They were more or less isolated initiatives where three standardisation areas could be distinguished: hardware, fixed material and procedures for analysis and storage of studies information. Several organisations as the MGED Normalization Working Group (MGED Data, 2006) were created for the standardisation of the information. The MGED (Microarray Gene Expression Data) Society is an international organisation devoted to the standardisation and to the exchange of information related to microarrays experiments. Other organisations to be mentioned are the OMG (Object Management Group) (OMG, 2006) or the UCL/HGNC (Human Gene Nomenclature) (HGNC, 2006).

As far as terminologies, vocabularies, nomenclatures and ontologies is concerned, it should be highlighted the MGED Ontology (MGED OWG, 2006), which describes the experiments and the gene expression data, or the GO (Gen Ontology Consortium) (Gene Ontology, 2006), which provides controlled vocabularies for describing the molecular function, the biological process and the cellular components of the gene products. Also the UCL/HGNC (Human Gene Nomenclature) (HGNC, 2006), the TaO (TAMBIS Ontology) (TaO, 2006), the RiboWeb (RiboWeb, 2001) or the EcoCyc (EcoCyc, 2005) should be mentioned.

Regarding the data exchange standards in the microarrays field, the MicroArray and Gene Expression Markup Language (MAGE-ML) (MAGE-ML, 2006) is language designed for describing and communicating information among microarrays experiments.

Other data exchange standards are the Bioinformatics Sequence Markup Language (BSML) (BSML, 2006), the Gene Expression Markup Language (GeneXML) (NCGR, 2006) or the Genome Annotation Markup Elements (GAME) (Bioxml, 2006).

The MGED Group is the standardisation organisation that presents the wider scope regarding the microarrays field and presented in November 2000 the standard **MIAME (Minimun Information About a Microarray Experiment)** (MIAME, 2006). This acronym describe the minimal information regarding microarrays that, either should be stored into a data base (from now, DD.BB) used as a public repository, or that should be stored for enabling the non ambiguous interpretation of the experiments results and for repeating such experiments.

After defining the information that is going to be stored (MIAME), there should be a model of objects (UML) for describing, not only how the data of these experiments should be expressed, but also the mecha-

Figure 1. MGED ontology

3 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/ontologies-processing-patterns-microarrays/10403

Related Content

An Optimal Way of VM Placement Strategy in Cloud Computing Platform Using ABCS Algorithm
Pushpa R.and M. Siddappa (2021). *International Journal of Ambient Computing and Intelligence (pp. 16-38).*
www.irma-international.org/article/an-optimal-way-of-vm-placement-strategy-in-cloud-computing-platform-using-abcs-algorithm/279584

An Ambient Intelligence Based Multi-Agent System for Alzheimer Health Care
Dante I. Tapiaand Juan M. Corchado (2009). *International Journal of Ambient Computing and Intelligence (pp. 15-26).*
www.irma-international.org/article/ambient-intelligence-based-multi-agent/1369

Explainable AI (XAI) for Cybersecurity Decision-Making in Industrial Automation
Pawan Kumar Goel, Chin-Shiuh Shiehand Mong-Fong Horng (2025). *AI-Enhanced Cybersecurity for Industrial Automation (pp. 279-294).*
www.irma-international.org/chapter/explainable-ai-xai-for-cybersecurity-decision-making-in-industrial-automation/379630

Management and Optimization Methods of Music Audio-Visual Archives Resources Based on Big Data
Hongyu Liuand Chenxi Lu (2023). *International Journal of Ambient Computing and Intelligence (pp. 1-15).*
www.irma-international.org/article/management-and-optimization-methods-of-music-audio-visual-archives-resources-based-on-big-data/332866

Transforming Mental Health Care Through AI Technology
Rathi Priya Bollu Yashwantand Eng Hwa Wong (2027). *Encyclopedia of Modern Artificial Intelligence (pp. 1-30).*
www.irma-international.org/chapter/transforming-mental-health-care-through-ai-technology/407379