



Universiteit
Leiden
The Netherlands

Spatio-temporal framework for integrative analysis of zebrafish development studies

Belmamoune, M.

Citation

Belmamoune, M. (2009, November 17). *Spatio-temporal framework for integrative analysis of zebrafish development studies*. Retrieved from <https://hdl.handle.net/1887/14433>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14433>

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 7

CONCLUSIONS AND DISCUSSIONS

7.1 General overview

In this thesis we presented a 3D spatio-temporal framework for the zebrafish model system. This framework is intended to assist biologists in their studies of vertebrate development and is based on a number of components. The separate components were presented in the different chapters of this thesis. This chapter summarizes our conclusions and provides a discussion for each chapter separately.

7.2 The Developmental Anatomy Ontology of Zebrafish

In Chapter 2 we presented the Developmental Anatomy Ontology of Zebrafish (DAOZ). DAOZ is being developed to fill our need for an anatomy ontology to be used and adapted by computer-based applications that require anatomical information for data annotation and retrieval. DAOZ provides both a spatial and temporal structured ontology based on the anatomical vocabulary provided by ZFIN (<http://zfin.org>). Furthermore, we extended DAOZ with additional concepts and relations to cover the phenotypic structure of the zebrafish model organism at the most biologically relevant levels of granularity. Through the different relationships the ontology concepts are being structured as a Directed Acyclic Graph (DAG) and provided with clear semantics. DAOZ structure has been expressed in a set of axioms for a formal and consistent description. Furthermore, the complete ontology has been organized in an object-oriented database where we assigned to each ontological concept a unique identifier. Storing concepts in a database enables DAOZ to be readily navigable and understandable by curators for data annotation and by computers for data retrieval and mining. DAOZ concepts are in compliance with these known inside the zebrafish community; this assures integration and data dissemination.

An ontology can always be extended with more granularity for a larger data description and propagation; the DAOZ DAG structure enables such extensions with additional concepts and relations to improve the ontology maintenance and dissemination.

7.3 The 3D Digital Atlas of Zebrafish

In Chapter 3 we presented the 3D digital atlas of zebrafish model organism that has been restructured in an object oriented database system. This atlas contains 3D models at representative stages of zebrafish development. The 3D digital atlas serves as a coordinate framework for data comparison and analysis. Additionally, it could be a particularly useful tool for education. In our spatio-temporal framework, the 3D digital atlas is an indispensable component; to be used as a reference tool for submission of patterns of gene expression and retrieval of data it is has been restructured in a database system using DAOZ concepts as common nomenclature for its data annotation. Each 3D atlas stage model consists of a complex set of 3D volumetric anatomical structures. Anatomical structures – or domains- are annotated with unique identifier of spatio-temporal anatomical concepts from the DAOZ. Through DAOZ concepts anatomical domains of the 3D atlas models and set of images are restructured in a database system using different levels of resolution. Using the three-tier web-application, i.e. 3D ZFAtlasServer users can explore and query 3D models through the internet. This application uses the different levels of data organization to search anatomical domains in a 3D model. Queries are composed at a gross level of granularity based on general concepts to get detailed results at a finer level of abstraction. This query facility offers a readily access to this complex anatomical data for a wide range of users. Users can specify data to view according to their needs following the principle of ‘*you get what you want*’.

7.4 The Gene Expression Management System

Chapter 4 is dedicated to the description of another component in our framework, i.e. the Gene Expression Management System (GEMS). Molecular biology sometimes redefines anatomical borders through patterns of gene expression. The GEMS has been developed to manage, link and mine *in situ* expression patterns of marker genes during the embryonic development of zebrafish. Patterns of gene expression are 3D images; they are resulting from whole mount Fluorescent *In Situ* Hybridization (FISH) experiments, i.e. zebrafish (Welten et al, 2006) and imaging with the Confocal Laser Scanner Microscopy (CLSM) imaging. Where and when certain sets of genes are expressed regulate processes such as responses to cellular differentiation by growth factors. A number of developmental studies have already been realized using microarrays. These studies provide high throughput gene expression information at a gross level of granularity, i.e. in a mixture of cells or in whole organisms. However, the ideal situation to study development is to provide expression information at a finer level, i.e. in population of identical cells. Whole mount *in situ* hybridization is a process that enables gene expression visualization at cellular level in a whole organism. In our case, patterns of gene expression have a spatio-temporal dimension since expression information is visualized within cells (spatial information) of an intact zebrafish embryo (temporal information). The GEMS is the central repository where this spatio-temporal gene expression images are managed into the proper spatial and temporal context so that data can be adequately analyzed and thus contribution to developmental study can be usefully initiated. Most raw data from *in situ* studies are never published. With the GEMS, we facilitate online data submission and annotation of the original images from collaborating laboratories. The online submission to collect this raw data is the key for our system success.

The system allows management of several data types. The experimental protocol is part of the images submission. An experimental protocol could always play an important role for data analysis. Therefore, an image submission starts by presenting the experimental protocol as a form filled by pre-defined values that the users could always change and submit. An experimental protocol can always be modified. Therefore, we managed protocol information using XML format that offers a great flexibility for adaptation and maintenance then using a relational database. Each submitted protocol is stored as XML-laboratory notebook.

In order to support interpretation and comparison of gene expression datasets, gene expression patterns need to be linked with other resources for data mapping and comparison. The mapping of gene expression patterns onto the 3D atlas is realized using the consistent DAOZ concepts. DAOZ concepts also provide data consistency with data inside the zebrafish community. Additionally, Gene Ontology (GO) terminology is used as a vocabulary to annotate gene and gene product of GEMS expression patterns; through the GO terminology, GEMS data integration is extended to other resources such as NCBI and Entrez gene.

The most critical issue in a database system design is information access. We provide several ways to access the information residing in GEMS. The GEMS supports a variety of search query possibilities through different Graphical User Interfaces (GUIs). In its actual release, the GEMS provides data visualization in 2D format. As it is known, it is difficult to visualize in our mind 3D data. Therefore, 3D representation of the expression patterns is needed in order to give molecular information for developmental components in a 3D context. To this end, graphical models are derived. These models in the future will be integrated into the system to offer an additional 3D visualization to end-users. The GEMS, as many other central repositories, is confronted with the task how to improve its data exchange with the rest of the community. To this end, at the moment the Distributed Annotation System (DAS) is under consideration. DAS is being used to exchange

biological annotations between data distributed among different web-sites and a system such as DAS will be very useful to improve our data exchange with other resources.

7.5 The 3D Visual Query System

Chapter 5 was focused on a query system based on visual formulation of search queries. The complexity of spatio-temporal data requires tools that support and facilitate interactive data exploration. We have built a prototype environment for interactive querying and exploration of spatio-temporal data. When formulating visual queries, users start with zebrafish 3D models. Each 3D model represents the graphical representation of the input data. Users navigate through the data to look for domains of interest. The state of the visualization environment forms a visual query. The query output is a set of patterns of gene expression from the GEMS and expression models if available. In this chapter we described the prototype of the 3D Visual Query System (3D-VisQus). As an additional component to our information framework, 3D-VisQus offers a portal and intuitive interface to GEMS database through 3D models of the atlas. This system links the two components with unique identifiers of DAOZ concepts. With the 3D-VisQus we demonstrated two key elements; i.e. (1) we showed a query method based on perception and recognition of the visualized elements facilitating access and exploration to complex anatomical data; (2) by the 3D-VisQus we demonstrated that the consistent annotation and organization of GEMS data and atlas models in well structured databases allows their integration and mapping.

7.6 The GEMS: a mining tool for spatio-temporal patterns

In Chapter 6 we explored data mining strategies and we focused on mining association rules between sets of gene expression patterns in the GEMS database. Furthermore, we focused on user interaction with the rules resulting from the classification in the data. The GEMS, i.e. spatio-temporal framework for patterns of gene expression has been extended

by introducing an additional functionality to construct and mine association rules. Within the graphical user interface of GEMS, researchers are able to run operations to mine association rules between gene expression data. Genes have different temporal expression profiles. To take their temporal characteristics into consideration, we adopted the Progressive Partition Miner (PPM) algorithm to mine association rules over annotated images in the GEMS. Furthermore, we developed java agents to execute the algorithm autonomously with each submission. Through the GEMS user interface users are able to send requests to the system to run the mining algorithm and to generate rules on the fly. The same GEMS framework is used to put the discovered rules by intuition into context with existing biological knowledge. The spatio-temporal images and descriptive annotations of the generated rules represent a first attempt for users to start data analysis. Furthermore, the GEMS framework enhances each expression pattern analysis by providing links to external resources (cf. chapter 4).

We started with the PPM algorithm to introduce and explore mining aspect within the GEMS framework. In our future work, we intend to use other promising approaches such as Frequent Episode Mining in Developmental Analysis (FEDA). This algorithm is more tailored to our developmental data as it is based on analyzing sequences of developmental characters to discover episodes and these are used to determine differences between developmental sequences (Bathoorn et al, 2007).

With this additional functionality to mine for association rules we showed that the GEMS offers a platform for linking and mining 3D spatio-temporal patterns of gene expression of zebrafish model system. We mined for association rules over annotated images. Moreover, mining for associations over features in the images is under considered (Jano et al, 2007).

7.7 General conclusions

In this thesis we presented a spatio-temporal framework to facilitate studies in developmental genetics and biology. This framework is composed of three main components, i.e. the 3D atlas of zebrafish that represents the reference framework for zebrafish data submission and retrieval, the GEMS which is the central repository for 3D spatio-temporal patterns of gene expression for data storage, retrieval and mining and the DAOZ corresponding to the standard semantic framework for data annotation in both the 3D atlas and the GEMS. Through dedicated three-tier web-applications each of these components can be accessed separately for an easy data exploration and analysis. An additional component, i.e. the 3D visual query system has been developed to serve as a portal interface to access our spatio-temporal data as a whole. It facilitates user's tasks to access GEMS data through 3D atlas models by using DAOZ entities. The study presented here is in its proof-of-principle stage that will be followed by the widespread-adoption stage to cover more model systems.

The actual developmental framework should be extended with additional components such as microarray elements. A wealth of information can be obtained from microarrays to understand genetic network through development. Microarray for genes expression profiling when combined with spatial and temporal patterns of gene expression may highlight the processes that take place during embryonic development. Comprehensive microarrays covering large numbers of the predicted expressed transcripts for zebrafish are available. Whole mount *in situ* data stored in the GEMS is typically used to provide gene function information within a biological process. Microarrays are necessary for pathways analysis and understanding the gene expression network within which a particular gene operates. Therefore, with the addition of microarrays, that all of the tools for complex dissections of both cellular and genetic pathways are available to developmental biologists. Given this data availability of arrays and our computing

infrastructure, we should make increasing use of this wealth of data to improve our framework for computational assessments. To conclude, in this thesis we presented our platform for linking and mining spatio-temporal databases of gene expression for zebrafish developmental model organism. This platform is dedicated to study zebrafish development. However, it has been designed in such a way to be scalable to cover other model organisms for an improved environment for developmental studies. Moreover, through this platform data linking and exchange with other model organisms should be facilitated and be more easy to realize. This spatio-temporal framework is an on-going research that opens up cross platform validation and search to improve developmental studies.

