Integrated Visualization and Analysis of a Multi-scale Biomedical Knowledge Space

Asan Agibetov¹, Ricardo Manuel Millán Vaquero², Karl-Ingo Friese², Giuseppe Patanè¹, Michela Spagnuolo¹ and Franz-Erich Wolter²

¹Istituto di Matematica Applicata e Tecnologie Informatiche, Consiglio Nazionale delle Ricerche Genova, Italy
²Welfenlab, Division of Computer Graphics, Leibniz University of Hannover, Germany

Abstract
The study and analysis of relationships in a complex and multi-scale data set is a challenge of information and scientific visualization. This work proposes an integrated visualization to capture all the important aspects of multi-scale data into the same view by leveraging the multi-scale biomedical knowledge encoded into an underlying ontology. Ontology supports visualization by providing semantic means to identify relevant items that must be presented to the user. The study and analysis of relationships across the scales are presented as results of queries to the multi-scale biomedical knowledge space. We demonstrate the prototype of the graphical interface of an integrated visualization framework and the knowledge formalization support in an example scenario related to the musculoskeletal diseases.

Categories and Subject Descriptors (according to ACM CCS): H.5.0 [Information Systems]: Information Interfaces and Presentation—General; J.3 [Computer Applications]: Life and Medical Sciences—Health; I.2.4 [Computing Methodologies]: Artificial Intelligence—Knowledge Representation Formalisms and Methods

1 Introduction
This paper investigates visualization methods for the biomedical domain that studies musculoskeletal articulation of the human body and related diseases, focusing on the anatomical district of the human knee. In particular, we are interested in studying pathologies that may have different disease features at different scales (e.g. cellular, molecular, tissue, anatomy and behavior). Consider a pathology that was evidenced as the result of a gait pattern study (behavior scale), which might have been caused by the disruption of the macromolecules content during the cellular behavior change (cellular scale). For a complete understanding of this disease, information sources from all different scales and their relationships have to be considered.

In the following, we assume that the information sources that come from different scales, their relationships and data representing or accompanying them, reside in a multi-scale biomedical knowledge space, and for simplicity we are going to refer to it as $M$. In $M$, different specialists, such as tissue or biomechanical engineers, contribute with their data and expertise. Sharing the information contained in $M$ is required but it is not an easy task. On the one side, knowledge formalization may be used to specify explicitly a shared conceptualization, for instance using ontologies [SBF98]. Ontologies are a means to identify relevant items in a given domain and formally define what are the properties or attributes necessary to document them for an effective sharing. On the other side, ontologies alone are not enough. To give a cognitively rich and interactive exploration of $M$, smart visualization means are needed to integrate visualization at the conceptual level with visualization of patient data. Current ontology visualization tools are not sufficient to this aim, hindering the use of ontologies, and formalized semantic description in general, in the biomedical domain.

In this work we propose an integrated visualization environment to provide effective means to visually navigate the knowledge space of a complex domain such as musculoskeletal diseases. In this context, ontology serves as knowledge formalization of $M$ and it also registers the suitability of visualization techniques of the patient-specific data. Queries to the ontology determine semantically relevant data and information to be visualized together in a single frame of view, alleviating the search across all the patient specific-data and allowing a local and global understanding.
2 Previous work
Multi-scale biomedical visualization aims at the integration of biomedical data distributed at different spatio-temporal scales and their simultaneous presentation with detailed and global information [MMC12]. Except in the domains of genomics and proteomics, it is mainly covered by Scientific Visualization (SciVis), focusing primarily on physical data and providing realistic representations. Advances in multi-scale biomedical visualization have been made during the last decade, and the level of integration of multidisciplinary research has been increasing [OGG10]. Recent projects [HCB10] enable collaborative investigation of the human body as a single complex system and demonstrate how multi-scale approaches can solve specific multidisciplinary biomedical challenges. However, improvements are slow due to the difficulties in the development of scientist-centric visualizations that provide understanding [Joh04]. Fully integrated visualization across all the range of data types has not been achieved yet [GOB10].

An ontology is a knowledge representation, whose role is to define the concepts, relationships, and other distinctions that are relevant for modeling a domain [Gru93]. Applications of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97] that are relevant for modeling a domain [Gru93]. Applications of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97], to extensions of ontologies in medicine range from definitions and classifications of common medical terms [SCC97].

The traditional distinction of visualization techniques into SciVis and InfoVis delimitates their uses [Rhy03]. The need of overcoming this differentiation led to the trend of proposing new visualization classifications [Hag11] and the convergence of visualization techniques developed in parallel [Hau06]. An example in this direction is the visualization of the anatomical hierarchy integrated with volumetric data [BVG10]. Indeed, the techniques of these subfields are complementary and can be smoothly integrated in order to represent the features contained in $\mathcal{M}$.

3 Representing the knowledge formalization of $\mathcal{M}$
In order to have a complete understanding of a pathology with different disease features on different scales, we need to organize heterogeneous and multi-scale information sources. We also have to take into account different relationships between these information sources, as well as data that represent them. To this end, we assume to have an ontology that encodes such a multi-scale biomedical knowledge, likely a complex ontology, and we want to find an effective means to visualize relevant items to be presented to a specific user within the domain $\mathcal{M}$.

In our approach, relevant items are identified as results of queries to the ontology, exploiting a graph representation of the ontology. More precisely, we represent ontology as a labelled directed graph $G = (V, E)$, where nodes $V = \{C, I\} = \{\text{concepts (C)} \text{ and instances (I)} \text{ of concepts, and edges E = \{R, isa\}, where R are relations between instances and "is a" is a relation between instances and concepts. G is labeled with I : V \mapsto L, that maps nodes to the corresponding labels (L, labels of concepts, individuals and relations). For example, cellular change, loss of biomechanical function, MRI evidence and alteration in gait pattern are instances of a degradation process feature, represented as a graph $G$ as depicted on Figure 1. The general structure of the ontology focuses on multi-scale degradation process features $\langle DPF \in C \rangle$ that may cause one another ($cause, caused ^{-1} \in R$), are evidenced ($evidenced \in R$) by sources of evidence ($SOE \in C$) which in turn are measured ($measured \in R$) by different techniques. This essentially formalizes propagation of degradation process features in hierarchical pathologies. The ontology supports visualization, by formalizing relationships between multi-scale data, patients, techniques, user profiles and relevant visualization items (e.g. spatio-temporal scale, visualization suitability). Succinctly the multi-scale biomedical knowledge space is represented by $G$ on Figure 2.

4 Methodology of the integrated visualization
The proposed visualization combines a multi-scale approach and InfoVis and SciVis techniques in order to merge and present visually relevant features of $\mathcal{M}$ into the same view.

User profile We classify data of $\mathcal{M}$ according to their scale of data and visualization suitability (Table 1). Micro-scale, medium-scale and macro-scale classify data acquired or derived from different spatio-temporal ranges. Abstract scale encompasses all the nonspatio-temporal knowledge that cannot be directly implied from one of the previous scales, e.g. anatomical structure or relations between evidences in a multiscalar pathology. The main scale of interest of the user is saved in the user profile. Visualization suitability is the ap-

![Figure 1: Degradation process feature (DPF).](image-url)
propriateness of a given data set (sub-space) of \( \mathcal{M} \) to be represented by using a concrete SciVis (e.g., volume rendering for MRI) or InfoVis technique (e.g., node-link diagram for hierarchical data or bar chart for statistical data). The aforementioned parameters, which are encoded in the ontology, allow the positioning of the multi-scale data on the visualization framework in a suitable way according to their properties and user interests.

**Integrated visualization framework** The visualization scene (Figure 3) consists of three layers: focus, context and background. Data sets from each spatio-temporal scale (i.e., micro, medium and macro) are positioned on one of these layers, which are mainly distinguished by their z-order. Focus constitutes the main scale and data sets placed there are visualized on the front layer. Context is placed behind and its data sets are spatially aligned with the previous data to provide context to the data on the focus layer. Background is the last layer, least seen and less important. Its data sets are not directly related to the focus layer data, but they complete the general view across all the spatio-temporal ranges (e.g., macro-scale data set in the Figure 3). Data are mainly included in nodes, which allow a consistent representation. The actual visualization of each node depends on the visual-ization suitability (e.g., graphs, images or 3D content).

The positioning of the nodes depends on the visualization suitability of the data and the current main scale of interest. Given the main scale, its data representation will be positioned on the focus layer, represented by a main node. This node contains the most relevant SciVis data for the scale. InfoVis data and other specific SciVis data sets are visualized with subnodes. The integration of the adjacent scales depends on their spatio-temporal proximity. The most adjacent scale is aligned in the view on the context layer, and the last one is presented on the background layer using the call-out technique. The use of this augmented representation alleviates the differences in the order of magnitude of data, and allows the direct extraction of information from the context. The abstract scale is represented by perceptual cues in the relations between nodes (labels, arrows, colored lines), enriching understanding, e.g., spatial origin of sources, anatomical structures, hierarchy of evidences in a pathology.

Figure 2: Graph representation \( G \) of \( \mathcal{M} \)'s ontology.

<table>
<thead>
<tr>
<th>Visualization/suitability</th>
<th>Scale of Data</th>
<th>SciVis technique</th>
<th>InfoVis technique</th>
<th>User profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro-scale</td>
<td>Molecular and histological images, micro-CT</td>
<td>Properties extracted at molecular, cellular and tissue level</td>
<td>Tissue engineer</td>
<td></td>
</tr>
<tr>
<td>Medium-scale</td>
<td>PET-MRI, CT and segmentations</td>
<td></td>
<td>Radiologist</td>
<td></td>
</tr>
<tr>
<td>Macro-scale</td>
<td>Gait pattern animation, original video sequence</td>
<td>Gait pattern graphics</td>
<td>Biomechanical engineer</td>
<td></td>
</tr>
<tr>
<td>Abstract scale</td>
<td>-</td>
<td>Anatomical structure, relation among evidences, other derived knowledge</td>
<td>Computer scientist, Generic</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Data structure and user profiles of \( \mathcal{M} \).

Figure 3: Proposed visualization framework, with micro-scale as main scale.

5 Example scenario

Musculoskeletal diseases are an example of hierarchical pathologies of multi-scale nature. Figure 4 represents the articular cartilage degradation during osteoarthritis. Degradation features on the cellular scale propagate upwards through molecular, macromolecular, and tissue scales, causing finally the alteration in gait pattern [Gol12, ADC*08]. For a complete understanding of the process, data acquired through different techniques across different scales and their relationships have to be considered. In this scenario, the multi-scale biomedical knowledge space \( \mathcal{M} \) basically consists of the features of the degradation process, their relationships and the aforementioned information sources as evidences of the features. Ontology querying languages, such as SPARQL (SPARQL Protocol and RDF Query Language) [SP08], use graph pattern matching techniques to evaluate answers. Evaluation of queries for graph-based structures is completely out of the scope of this paper, we refer to [PAG09] for details. We only present here schematically intuition for graph pattern matching techniques and how the results of these queries may be used to identify semantically relevant items to support visualization. For example the results to the following query, expressed in English as “Given a technique, what is the DPF evidenced by its sources of evidence?” may be obtained by evaluating graph pattern match from query graph \( Q \) to the Knowledge Base graph \( KB \) (see Figure 5). The answer set of mappings from
Figure 4: Degradation of articular cartilage during osteoarthritis. Each DPF is evidenced by sources of evidence and those are measured by certain techniques. Note: this example is a particular instance of M’s ontology.

Variables in Q to values in KB (depicted in different colors on Figure 5) is summarized in the following table:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Source of Evidence</th>
<th>Spatiotemporal Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live/Death count</td>
<td>Cell viability</td>
<td>Micro</td>
</tr>
<tr>
<td>PCR</td>
<td>Cellular change</td>
<td>Micro</td>
</tr>
</tbody>
</table>

Queries to the ontology of M can also retrieve all the relevant information of a given patient (“Which are all the Data which concern a specific patient?”), and the techniques used (“Which is the technique that obtains specific Data?”).

Visualization The support of the ontology allows for obtaining all necessary visualization parameters for positioning the different spatio-temporal data sets on the scene and their proper representation, as visualization suitability and scale of data (“In which Spatio-temporal scale a given DPF is placed?”, “Which Visualization technique visualizes the given Data?”). Accordingly, the data sets are represented and positioned, as described in Section 4. Another important fact is that ontology can also be queried to get information from the abstract scale, e.g. the relations between the sources of evidences which proof the different cartilage degradation process features (Figure 6).

6 Conclusion and future work

The next step for the proposed system is to perform the interactive exploration of multi-scale data sets [MVRFW14]. This will take benefit from the visualization criteria adopted, such as scene layers or the consistent representation of data sets in nodes. Specifically, the system aims to provide two different view modes. In the 2D focus view mode, the user viewport is perpendicular to the layers, and the user can manipulate the nodes, enlarge or hide them and change the view of the data set contained in the node. In the 3D overview mode, the three layers are presented in 3D perspective, which allows the multi-scale navigation by sorting the layers. Another step is to better support collaborative diagnosis in which different medical specialists work together in the same environment while preserving their habitual way of working. They augment M with their findings; discuss their conclusions and continue the knowledge discovery in M.

Acknowledgements

This work was supported from the EU Marie Curie ITN MultiScaleHuman (FP7-PEOPLE-2011-ITN, Grant agreement no.: 289987). The authors would like to thank Marta Ondresik (3B’s Research Group) for providing the description of the example scenario (Figure 4), and all the partners for providing biomedical data sets. The two first authors Asan Agibetov and Ricardo Millán have contributed equally to the writing of the paper.
References


[BDBS04] Baecker E. H., Dang N., Babaria K., Shneiderman B.: Visualization and analysis of microarray and gene ontology data with treemaps. BMC Bioinformatics 5, 1 (June 2004), 84. PMID: 15222902. 2


© The Eurographics Association 2014.