ApiNATOMY: Towards Multiscale Views of Human Anatomy

Bernard de Bono[2], Pierre Grenon[2], Michiel Helvenstijn[1], Joost Kok[1], and Natallia Kokash[1],[*]

[1] Leiden Institute of Advanced Computer Science (LIACS), Leiden, The Netherlands
[2] University College London (UCL)

Abstract.

1 Introduction

An ontology consists of a set of terms and their relations that represent a specific domain of knowledge. Ontologies are created and maintained by knowledge domain experts, and are used as computer-readable taxonomies by software tools intended to support knowledge management activities in that domain.

The complexity of some of the ontologies in current use, as well as the complexity of handling semantic metadata that annotate third party resources with ontology terms (e.g. as described in [1]), has generated considerable demand for effective visualization in the design, authoring, navigation and management of (i) ontology-based knowledge and (ii) semantic metadata that make use of ontologies.

In response to the above demand, a number of generic ontology visualization tools have been developed to assist knowledge acquisition, browsing and maintenance of ontologies [2]. Such tools, however, put a considerable and unrealistic demand on the users' familiarity and expertise in both (i) semantic web technologies and (ii) the design principles of ontologies. It is unlikely that a user with expertise in the domain of an ontology also has expertise in the technologies managing ontologies.

The domain of biomedical physiology is a case in point. Physiology experts deal with complex biophysical operations, across multiple spatial and temporal scales, which they represent in terms of the transfer of energy from one form to another and/or from one anatomical location to another. Different kinds of descriptions of these biophysical operations are produced by different disciplines in biomedicine. For instance, (i) a medical doctor may describe the mechanism by which a stone in the ureter causes damage in the kidney; (ii) a pharmacologist may depict the process by which a drug absorbed from gut transits to the hip joints where it reduces inflammation; (iii) a molecular geneticist may trace the anatomical distribution of the expression of particular gene to understand the

---

[*] Corresponding author, email: `nkokash@liacs.nl`

cause of a skeletal malformation; and, (iv) a bioengineer may build a mathematical model to quantify the effect of hormone production by the small intestine on the production of bile by the liver. These descriptions take diverse forms, ranging from images and free text (e.g. for a paper in a journal) to XML documents bearing well-defined data or sets of variables and related equations (e.g. as input into a simulation tool).

Automating the discovery of relationships, in terms of physiological meaning, between the above types of description is a key goal to the physiology community. To that end, this community is investing considerable effort in building ontologies for the annotation and semantic management of resources describing physiology. A number of reference ontologies have been created to represent the various entities required to describe physiology, including gene products [3], chemical entities [4], cells [5] and gross anatomy [6]. Cumulatively, these reference ontologies consist of hundreds of thousands of terms such that the volume of semantic metadata arising from annotation of resources with these terms is considerable. However, conventional technology for the management of ontologies and metadata is not accessible to physiology experts.

The ApiNATOMY effort has emerged to provide intuitive graphical means to manage ontologies and semantic metadata relevant to physiology. In this paper, we present an integrated web-based ApiNATOMY environment for physiology experts to navigate through circuitboard visualizations of body components and their connections across scales. In particular, we present a tool prototype that visualizes (i) schematics of ontology-based knowledge about body parts, cardio-vascular and neural connections, over which (ii) graphical renderings of semantic metadata linking to gene products and mathematical model of processes are overlaid.

In this work, we first overview the ontology, metadata and data resources that we focused on, and outline usage scenarios that motivate our work (Section 2). We then we discuss representation and visualization methods applied to arrangement and display relevant resources (Section 3). In Section 4, we describe a graphical tool prototype that implements these methods. Furthermore, in Section 5, we overview related methodologies, efforts and techniques in the field. Finally, we conclude the paper and discuss future work in Section 6.

2   Use case, resources and early-stage results

In this section, we briefly discuss a core use case for the application ApiNATOMY schematics in support of genomics and drug discovery studies. In so doing, we introduce (i) some of the key ontological and data resources required in this case, as well as (ii) early-stage results of this effort.

The domains of genomics and drug discovery are dependent on physiology knowledge, as both domains take into account the manufacture of gene products in different parts of the body and the regulated long-distance transport of molecules that interact with these products. The location of gene product manufacture (i.e. gene expression data, such as [7]), as well as routes associated with

pharmacokinetic modeling of molecular interactors (drawn from resources such as [8]) may be usefully depicted in the form of a physiology circuitboard.

In ApiNATOMY, a physiology circuitboard schematic consists of a combination of (i) an anatomical treemap and (ii) an overlay of process graphs. In our earlier prototypes (described in [9][10]), templates were applied to constrain the layout of tiles in treemaps of the Foundational Model of Anatomy (FMA) [6] ontology, such that nesting of one tile inside another indicates that the child tile is either a part of a subclass of the parent tile. The Graphical User Interface (GUI) providing the interaction with the circuitboard allows point-and-click navigation of the treemap content. This type of interaction extends to also to involve process graphs - in this paper we report on the graphical projection of routes of (i) blood flow processes linking different regions of the human body (using data generated in [11]), as well as (ii) transport processes along neurons of the central nervous system (i.e. brain and spinal cord) with data obtained via the Neuroscience Information Framework [12].

The ApiNATOMY Graphical User Interface is built from inception as a three dimensional (3D) environment. This feature facilitates interaction not only with 3D renderings of the circuit boards themselves, but also with a wide range of geometry/mesh formats for volumetric models of biological structure across scales. For instance, it is already possible to overlay Wavefront .obj data from BodyParts3D [13] as well as SWC data supported by the neuromorpho.org [14] resource.

In the next sections, we discuss:

- the constraining of treemap layout to generate stable anatomical treemaps,
- the design and overlay of routes of communication for the cardiovascular and neural systems, and
- the querying and 3D depiction of protein architecture schematics for the anatomical overview of gene expression data.

We arrange data in a hierarchical way starting from the large-scale views showing external and internal surfaces and organs arranged to resemble the longitudinal section through the middle of the human body as the top-level taxonomy. Figure 1 shows the idealized radially symmetric body plan, apportioned over cylindrical regions. This homunculus has a central longitudinal axis of rotation located in the idealized gut lumen which runs in the cephalocaudal direction. Each of the organs in the plan is composed of multiple tissues and sub-organs, the structural information about them is obtained from the FMA ontology [6]. As the visualization of the full ontology may obscure the details a user is interested to see, it is essential for the visualization tool to support data filtering across multiple levels and contextual zooming into selected areas. The user should be able to create a custom view with the internal structure of the selected body parts placed in a way that simplifies the analysis of these data.

(a) Schematic body plan   (b) Visualizing medical ontologies using treemaps
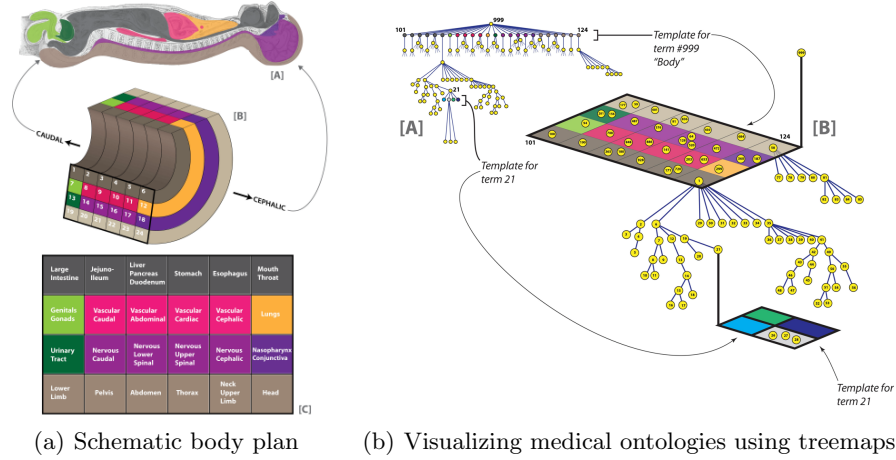
Fig. 1. Longitudinal section through the middle of the male human body

3   Visualizing taxonomies and connectivity data

Treemaps [15] are an effective technique to visualize hierarchical data by using nested shapes in a space-filling layout. Each shape represents a geometric region, which can be subdivided recursively into smaller regions. The standard shape is the rectangle. Nodes in a treemap, also called *tiles*, represent individual data items in a dataset. Node size, color and text label can be used to represent attributes of the data item. One-layered treemaps can display data attributes but are not very good at emphasizing the place of an item in the overall hierarchical structure. To compensate for that, a small margin with structural labels is typically used. In treemaps displaying hierarchical structures, it is possible to navigate among different layers and zoom into selected tiles [16].

To create a treemap, one must define a tiling algorithm - a way to divide a tile into sub-tiles of specified areas. Tiling algorithms used for typical applications of treemaps such as e.g., visualization of folders in files in the computer file system with their respected sizes, do not associate tile positions with any characteristic of the data. This is not the case in our scenario: while a user navigates among different layers, filters data and zooms into selected areas, the tool should keep the tiles associated with the data in the same relative positions to each other. Otherwise, the user's perception of the displayed information will be quickly affected. Moreover, our tiling algorithm should allow the user to enforce constraints on tile positions to make the treemap views structurally resemble body regions. Hence, we developed a stable and customizable tiling algorithm that arranges tiles corresponding to data items according to a given template [10].

For a set of n data items with no positional constraints, a default template is created that consists of $\lfloor\sqrt{n}\rfloor$ rows and $\lceil\sqrt{n}\rceil$ columns in each row but the last one (which may contain less columns). If the positional data is available (e.g., FMA

ontology adjacent-to relation) or a user wants to rearrange the data manually, a custom template is associated with the parent node of the dataset items. The template is a hierarchical structure $\{splitType, \{\}, ..., \{\}\}$ where $splitType \in \{slice, dice\}$ defines a way to split the rectangle into sub-rectangles: vertically or horizontally. By recursively splitting the available area into sub-rectangles, one can define complex layouts that enforce two dimensional constraints in the form "x is left/right of y" or "x is above/below of y" where x and y are individual data items or groups of data items that in their turn can be allocated as needed using the same technique.

3.1   Connectivity data

We use treemap-based body plans as background to overlay the schematic representation of biological systems such as circulatory, respiratory, or nervous systems. Body systems are essentially graphs or hypergraphs with nodes corresponding to body parts (treemap tiles), entities that belong to body parts (proteins, cells, etc.) or auxiliary nodes that are not displayed on the treemap but still carry important biomedical information, while their edges or paths represent organ system compounds such as blood vessels or nervous connections that pass through the certain body parts or its sub-parts.

Body systems are intrinsically complex and require efficient data visualization techniques that would help us to avoid clutters induced by the large amount of graph edges and their crossings and allow users to overview large parts of the body systems as well as to trace individual connections and analyze their structure. Edge bundling techniques [17],[18],[19] have been proposed to improve perception of connectivity data in large dense graphs. Such techniques generally rely on edge rerouting strategies that are either solely target at improving visual perception of edges constrained by positions of nodes or exploit the relations among connectivity data as guidelines for more natural allocation of graph edges and nodes. Our application requires the mixture of these techniques

As example, we consider the schematic visualization of blood vessels in human body. The initial dataset is a graph extracted from the FMA database which consists of approximately 11,300 edges and over 10,000 distinct nodes. An edge represents an unbranched segment of a blood vessel. Nodes represent junctions between blood vessel segments. Samples of records from the dataset are shown in Table 1. The first column in the dataset is a unique vascular segment identifier (ID). The second column bears the vessel type, i.e. any one of four numbers that give an indication of the biological type of segment: 1 - for arterial segment, 2 - microcirculation (MC), there are three types of MC edges: arteriole, capillary and venule, 3 - venous, and 4 - cardiac chamber. The third column bears FMA IDs. For non-MC edges (i.e. vessels of type 1, 3 or 4) the number in this column is the FMA ID of the blood vessel of which that segment is part (e.g. there are over 50 segments/edges that form part of the trunk of the aorta, i.e., FMA ID 50010931). For MC edges (type 2), the FMA ID is that of the body region in which the MC is embedded. The fourth and fifth columns in the dataset bear

the unique node identifiers in an edge pair. The sixth column is a free-text label describing that segment (i.e. edge).

Table 1. Vascular connectivity data from the FMA ontology

Segment	Type	FMA	Node 1	Node 2	Description
121a	2	62528	62528_2	62528_4	Arterioles in Microcirculation segment of Wall of left inferior lobar bronchus
121c	2	62528	62528_4	62528_5	Capillaries in Microcirculation segment of Wall of left inferior lobar bronchus
121v	2	62528	62528_3	62528_5	Venules in Microcirculation segment of Wall of left inferior lobar bronchus
...	...	...	...	...	...
8499	1	69333	8498_0	62528_2	Arterial Segment 8499 of Trunk of left second bronchial artery from origin of supplying terminal segment to the arteriolar side of the Wall of left inferior lobar bronchus MC
9547	3	66699	9546_0	62528_3	Venous Segment 9547 of Trunk of left bronchial vein from origin of supplying terminal segment to the venular side of the Wall of left inferior lobar bronchus MC

When it comes to describing routes of blood passage through the heart, one has to distinguish between cardiac chambers, (i.e. main lumen of left ventricle, left atrium, right ventricle or right atrium) and the MC of the wall of heart (e.g. wall of left ventricle (FMA ID 7101), left atrium (7097), right ventricle (7098) or right atrium (7096)). There are two ways by which blood passes through the heart: (i) though the chambers and (ii) through its walls. Consequently, we exclude cardiac chamber segments from consideration as they are redundant for the representation of vascular connectivity data in our application. An MC is represented by three edges connected in series: one edge represents tissue arterioles, a second edge stands for the bed of capillaries, while a third denotes the venules. In one MC, therefore: a) the end node of the arteriolar edge and the start node of the capillary edge are equivalent, and b) the end node of the capillary edge and the end node of the venular edge are equivalent.

In the above example, the anatomical entity in which the MC is embedded is 62528 - the topology of MC segment connectivity is as follows:

$$62528\_2 - [121a] \rightarrow 62528\_4 - [121c] \rightarrow 62528\_5 \leftarrow [121v] - 62528\_3.$$

MC segment 121a is supplied with blood by the arterial segment 8499 while MC segment 121v is drained of blood by the venous segment 9547.

The accurate visualization of the cardiovascular system in a comprehensible way requires complex pre-processing (about 12 rules were identified to extract the data of interest from the presented dataset by a biomedical expert in our team). In this paper, for the illustration purpose we show only paths connecting

MCs of the walls of the heart to MCs belonging to the sub-organs of the organs in our 24 upper level tile body plan. To obtain this view, we looked for the shortest paths (due to the way the data is represented in the initial data set, loops are possible) from the MCs of the heart walls to the final FMA tiles. For example, the path from the left ventricle to the wall of left inferior lobar bronchus MC is as follows: $7101 \rightarrow 2406 \rightarrow 2407 \rightarrow 2573 \rightarrow 2657 \rightarrow 3780 \rightarrow 4288 \rightarrow 4707 \rightarrow 4709 \rightarrow 4711 \rightarrow 4713 \rightarrow 4714 \rightarrow 8492 \rightarrow 8494 \rightarrow 8496 \rightarrow 8498 \rightarrow 8499 \rightarrow 62528$, while the path from this organ to the right atrium is $7096 \leftarrow 771 \leftarrow 1333 \leftarrow 1334 \leftarrow 1731 \leftarrow 1993 \leftarrow 1994 \leftarrow 1996 \leftarrow 1998 \leftarrow 1999 \leftarrow 1983 \leftarrow 9534 \leftarrow 9536 \leftarrow 9538 \leftarrow 9540 \leftarrow 9542 \leftarrow 9544 \leftarrow 9546 \leftarrow 9547 \leftarrow 62528$.
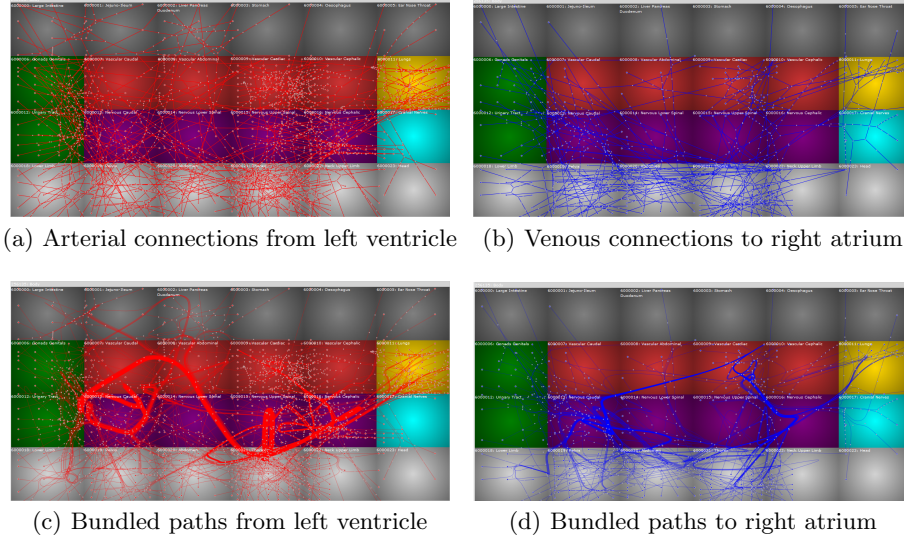
The first and the last IDs in this path correspond to the tiles in the treemap view, while the intermediate IDs will be represented using auxiliary nodes with undefined coordinates. One of the issues we encountered is the need to determine optimal positions for these nodes. Since several paths as above can have common sub-paths, the intermediate nodes should not deviate too much from the way from the heart MC to each of the end tiles sharing such sub-paths. This motivates our application of the sticky force-directed graph visualization method [20][21] that given a sub-set of nodes with fixed coordinates arrange other nodes to optimize forces on the graph edges. As our only objective is to allocate intermediate nodes as close to the paths they belong to (which most naturally can be represented by straight lines) as possible, we set gravity and charge to 0 and the desired link length to a minimal possible length. The extracted connectivity graphs for two heart walls: left ventricle (7101) and right atrium (7096), are shown in Figures 2(a) and 2(b).

Once the coordinates for all connectivity graph nodes are determined, we apply hierarchical edge bundling method that uses the path structure to bundle paths with common sub-paths in the graph obtained on the first stage. The final set of paths from left ventricle and right atrium are shown in Figures 2(c) and 2(d), respectively. In these images, both left ventricle and right atrium are descendants of tile 6000009: Vascular Cardiac. By setting bundling algorithm tension parameter close to 1 (0.96-0.99), we obtain a view that clearly shows that individual paths extracted from our dataset form "highways" which correspond to major arteries and veins in the human body.

After one-time pre-processing to import data from available external sources, we store connectivity data in a generic JSON format. A user can interact with and edit these data using functionality of our application. One of the prime goals of our tools is to simplify the access and maintenance of biomedical taxonomies, vascular system among others. A user should be able to choose a way to represent connectivity data. There is ongoing work on the implementation of the orthogonal connector visualization algorithm that would place links into margins between tiles so that they will not obstruct the interaction with the tiles.

4   Tool prototype

– Architecture
– Description: views, interaction
– ...

(a) Arterial connections from left ventricle (b) Venous connections to right atrium

(c) Bundled paths from left ventricle (d) Bundled paths to right atrium

Fig. 2. Cardiovascular system

5   Related work

The need for the multi-scale visualization and analysis of human body systems is well recognized by biomedical communities. For example, 3D Multiscale Physiological Human initiative deals with the mixtures of physiological knowledge and computational approaches to help scientists in biomedicine to improve diagnostics and treatments of various disorders [22],[23]. Numerous taxonomies and databases have been created and are widely used by researchers in the biomedical field [24]. While various generic diagrams and visualization techniques can be used to display biomedical ontologies [2], to the best of our knowledge, our tool is the first systematic approach to integrate such knowledge in one extensible and configurable framework.

Among the most effective taxonomy visualization techniques are space-filling diagrams, and in particular, treemaps. Burch and Diehl [25] discuss the ways to display multiple hierarchies and conclude that overlaying connectors on top of treemaps is the most visually attractive and easy to follow approach. Among the alternative options they considered are separate, linked and colored tree diagrams, sorted and unsorted matrices and sorted parallel coordinate views. Regarding the way to layout the connectors, two naive methods were considered: straight connections and orthogonal connections. de Bono et al. [9] describes limitations of existing treemapping tools for biomedical data visualization. To overcome these limitations, we introduced a generic method to build custom templates which is applied in our tool to control layout of ApiNATOMY body tissues. Among the advantages of the proposed tremapping method are cus-

tomizable layouts, visualization stability and multi-foci contextual zoom. The detailed comparison of our method with existing treemaping algorithms can be found in [10].

Our application requires multiple taxonomies consisting of thousands of items to be displayed on relatively small screens of handhold devices. We employ the same visualization technique with more advanced treemapping and connector layout algorithms. Due to the large amount of vascular connectivity data, we employ hierarchical edge bundling method [17] to get intuitive and realistic representation of blood flow across a teemap-based plan of human body. In contrast to the scenarios in the aforementioned work, not every node in our vascular connection dataset has a corresponding node in the treemap. Thus, force-directed graph drawing method [26] is added to the scene to find optimal positions of intermediate nodes on the paths that connect the root of the taxonomy (heart) with its leaves (body tissues shown as treemap tiles). The variation of the force-bundling method suitable for our application is known as sticky force-directed placement [20] which allows to fix the positions of certain nodes and allocate other nodes to achieve mechanical equilibrium between forces pulling the free nodes towards fixed positions.

Other potentially useful methods did not provide the desired result. The first approach we tried consists of applying the force-directed edge bundling method [27] to bundle entire paths among the heart chambers and body tissues does not reflect the hierarchical structure of vascular connectivity graph. The second approach, force-based edge bundling over a graph produced by sticky force-directed node allocation algorithm results into unnatural distortion of short edges towards each other. Other edge-bundling methods(e..g, [18][19][28]) operate on graphs with known node positions and thus would produce visualizations on our data that suffer from similar problems.

6   Conclusions and Future Work

References

1. de Bono, B., Hoehndorf, R., Wimalaratne, S., Gkoutos, G., Grenon, P.: The RICORDO approach to semantic interoperability for biomedical data and models: strategy, standards and solutions. BMC Research Notes **4** (2011) 313
2. Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology visualization methods - a survey. ACM Comput. Surv. **39**(4) (2007)
3. Blake, J.A., et al.: Gene ontology annotations and resources. Nucleic Acids Res. **41** (2013) D530–535
4. Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C.: The chebi reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. **41**(D1) (2013) D456–463
5. Bard, J., Rhee, S.Y., Ashburner, M.: An ontology for cell types. Genome Biol. **6**(2) (2005) R21
6. : A reference ontology for biomedical informatics: the foundational model of anatomy. J. Biomed. Inform. **36**(6) (2003) 478–500

 7. EBI: Arrayexpress home — ebi. Web site (2012)
 8. L. Harnisch, I. Matthews, J.C., Karlsson, M.O.: Drug and disease model resources: a consortium to create standards and tools to enhance model-based drug development. CPT Pharmacomet. Syst. Pharmacol. **2** (2013) e34
 9. de Bono, B., Grenon, P., Sammut, S.: ApiNATOMY: A novel toolkit for visualizing multiscale anatomy schematics with phenotype-related information. Hum Mutat **33**(5) (2012) 837–848
10. N. Kokash, de Bono B., K.J.: Template-based treemaps to preserve spatial constraints. In: Proc. IVAPP 2014. (2014)
11. de Bono, B.: Achieving semantic interoperability between physiology models and clinical data. In: Proc. of IEEE Int. Conf. on e-Science Workshops. (2011) 135–142
12. Gardner, D., et al.: The neuroscience information framework: A data and knowledge environment for neuroscience. Neuroinformatics **6**(3) (2008) 149–160
13. Mitsuhashi, N., Fujieda, K., Tamura, T., Kawamoto, S., Takagi, T., Okubo, K.: Bodyparts3d: 3d structure database for anatomical concepts. Nucleic Acids Res. (**37**) D782–D785
14. Ascoli, G.A.: Mobilizing the base of neuroscience data: the case of neuronal morphologies. Nat. Rev. Neurosci. **7**(4) (2006) 318–324
15. Johnson, B., Shneiderman, B.: Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: Proc. of the 2nd Conference on Visualization '91, IEEE (1991) 284–291
16. Blanch, R., Lecolinet, E.: Browsing zoomable treemaps: Structure-aware multiscale navigation techniques. TVCG **13** (2007) 1248–1253
17. Holten, D.: Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. IEEE Transactions on Visualization and Computer Graphics **12**(5) (2006) 741–748
18. Gansner, E.R., Hu, Y., North, S.C., Scheidegger, C.E.: Multilevel agglomerative edge bundling for visualizing large graphs. In Battista, G.D., Fekete, J.D., Qu, H., eds.: Proce of PacificVis, IEEE Computer Society (2011) 187–194
19. Hurter, C., Ersoy, O., Telea, A.: Graph bundling by kernel density estimation. Comp. Graph. Forum **31** (2012) 865–874
20. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force-directed placement. Software Practice and Experience **21**(11) (1991) 1129–1164
21. Bostock, M.: Sticky force layout. Online visualization tool (Accessed 09-14-2014)
22. Magnenat-Thalmann, N., Ratib, O., Choi, H.F., eds.: 3D Multiscale Physiological Human, Springer (2014)
23. Magnenat-Thalmann, N., ed.: Modelling the Physiological Human: Proceedings of the Second 3D Physiological Human Workshop. Volume 5903 of LNCS., Springer (2009)
24. Burger, A., Davidson, D., Baldock, R., eds.: Anatomy Ontologies for Bioinformatics. Volume 6 of Computational Biology., Springer (2008)
25. Burch, M., Diehl, S.: Trees in a treemap: Visualizing multiple hierarchies. In: Proc. VDA 2006. (2006)
26. Battista, G.D., Eades, P., Tamassia, R., Tollis, I.G.: Graph Drawing: Algorithms for the Visualization of Graphs. Prentice Hall (1999)
27. Holten, D., van Wijk, J.J.: Force-directed edge bundling for graph visualization. Comput. Graph. Forum **28**(3) (2009) 983–990
28. Selassie, D., Heller, B., Heer, J.: Divided edge bundling for directional network data. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis) (2011)