Methods and applications for visualization of SNOMED CT concept sets

A.R. Højen[1]; E. Sundvall[2,3]; K.R. Gøeg[1]

[1]Department of Health Science and Technology, Medical Informatics, Aalborg University, Denmark;
[2]Department of Biomedical Engineering, Linköping University, Sweden;
[3]County Council of Östergötland, Sweden

Summary

Inconsistent use of SNOMED CT concepts may reduce comparability of information in health information systems. Terminology implementation should be approached by common strategies for navigating and selecting proper concepts. This study aims to explore ways of illustrating common pathways and ancestors of particular sets of concepts, to support consistent use of SNOMED CT and also assess potential applications for such visualizations.

The open source prototype presented is an interactive web-based re-implementation of the terminology visualization tool TermViz that provides an overview of concepts and their hierarchical relations. It provides terminological features such as interactively rearranging graphs, fetching more concept nodes, highlighting least common parents and shared pathways in merged graphs etc.

Four teams of three to four people used the prototype to complete a terminology mapping task and then, in focus group interviews, discussed the user experience and potential future tool usage. Potential purposes discussed included SNOMED CT search and training, consistent selection of concepts and content management.

The evaluation indicated that the tool may be useful in many contexts especially if integrated with existing systems, and that the graph layout needs further tuning and development.

Correspondence to:

Anne Randorff Højen
Aalborg University
Fredrik Bajers Vej 7
9220 Aalborg Øst
Denmark
E-mail: arra@hst.aau.dk

© Schattauer 2014

A.R. Højen, E. Sundvall, K.R. Gøeg: Methods and applications for visualization of SNOMED CT concept sets.

1. Introduction

1.1 The challenge of consistent SNOMED CT implementation

Implementation of clinical terminology should support health information interoperability [1, 2, 3, 4] and data comparability [5]. Data comparability is necessary for enabling retrieval and aggregation based on specific predicate queries. Therefore, comparability of data is important both at the point of care and for secondary purposes [5, 6, 7]. To enable data comparability, a controlled clinical terminology like SNOMED Clinical Terms (SNOMED CT) can be used. SNOMED CT is the world's most comprehensive clinical terminology and is developed to support reuse of meaningful health information. SNOMED CT is complex due to its comprehensiveness and underlying reference model [8, 9] and therefore knowledge on SNOMED CT is required when approaching implementations that should support data comparability [10].

A typical approach for SNOMED CT implementations is to select concepts from SNOMED CT to represent legacy codes or interface terms applied in a *Clinical Information System* (CIS) [10, 11, 12]. In this case, SNOMED CT implementation is about selecting sets of concepts to be applied in a specific context, which can span from small groups of concepts to be included in a pick-list or larger sets of concepts for example concepts which cover all procedures and findings related to a physical examination. IHTSDO[1] provides a formal way of representing these sets of concepts through the use of *Reference Sets*. Hansen et al. [13] state the importance of Reference sets for building usable CISs. However, inconsistent concept selection is one of the challenges when implementing SNOMED CT to improve comparability of information [8, 14], and even within defined sets of concepts, coherency between the selected concepts is essential for supporting data retrieval and reuse [11, 15, 16].

Studies, [14, 15], conclude that terminology implementation should be approached by common strategies for navigating and selecting appropriate concepts. Even though IHTSDO suggests that each SNOMED CT hierarchy is used for a certain purpose, studies indicate that these purposes are not clear and specific enough to support consistent mapping. "The problem is not merely redundant concepts, but that two hierarchies can be used interchangeably" [16].

In an earlier study [15] we proposed a set of mapping guidelines to support concept selection in SNOMED CT. The primary focus in the guidelines is to ensure consistency in the concept selection by using concepts from same sub-hierarchies when representing similar information types. However, from an implementation viewpoint, conforming to the guideline is a major challenge. Having worked on a SNOMED CT project with an EHR implementation team in Northern Jutland in Denmark, we experienced the need for getting an overview of existing terminology. This was a prerequisite for choosing appropriate terms for new applications. In addition to this, we experienced that it is important to know that the concepts chosen for a template for example the *'rhythm of periphery pulses'*, *'heart rate'*, *'blood pressure findings'* and *'cyanosis'* have the common parent *'cardiovascular findings'* as knowing this supports meaningful clinical retrieval.

1.2 Terminology visualization to support consistency

Getting the needed overview of the existing terminology could be accomplished by visualization based on a whole set of concepts used in a certain template or a number of templates. Yet, the existing terminology-support in the CIS typically only provides alphabetically ordered lists of local codes and terms used. Commonly available SNOMED CT browsers with graphical features usually show hierarchical structure built up from a single concept at a time [17, 18]. For example the CliniClue browser supports a static graphic view of a single concept and its supertype ancestors [19]. The Snoflake browser [20] supports visualization of a single concept and its direct parent and children concepts, allowing for dynamic navigation through the hierarchies. Illustrations of the visualizations provided in CliniClue Xplore and the Snoflake browser is shown in ▶ Figure 1.

---

1   International Health Terminology Standards Development Organisation (www.ihtsdo.org)

Visualization has also been a matter of attention in earlier terminology studies. In [21] the ontology management system, Protégé is presented. Here, the use of visualization to provide an overview is acknowledged, and it is stated that different paradigms for visualization could be desirable dependent on use. Also, [17, 18] indicate a possibility for improved technologies for displaying the terminology structure and especially the subsumption hierarchy.

1.3 Visualization of complex terminology using TermViz

Visualizing complex terminologies such as SNOMED CT is challenging [22]. This is partly due to the poly-hierarchical structure, where nodes (concepts) can have multiple parents and several other relationship types. This makes it difficult to build a graph originating from more than one node at the time. Graphs very soon become crowded and entangled when the number of nodes increases [22].

It is difficult to dynamically highlight current points of interest and reduce clutter from other parts of the graph without losing context. Therefore, experience was leveraged from a former SNOMED CT visualization project called TermViz. The first TermViz implementation was a Java application that applied information visualization methods such as "focus+context" and self-organizing force-based layouts. This tool allows to focus on several nodes in the terminology simultaneously. Moreover, it uses interactive animated graph navigation for further exploration, combined with semantic zooming [22] reducing information (e.g. labels) from nodes with a lower degree of interest. In this study, the original TermViz design principles were the starting point for visualizing sets of concepts.

1.4 Objective

The aim with this study is to explore applications for sets of concept visualization, to support terminology implementation and management. An alternative tool for terminology visualization is developed, allowing sets of concepts to be the basis for a single graph. Moreover a qualitative study evaluated the potential usages of this type of visualization.

2. Methods

In this study a tool is developed, where sets of concepts are the basis for a single graph, since no existing tools support this visualization according to the above analysis. The aim of the evaluation is to clarify the usage of the visualization.

2.1 Terminological features

The tool should provide a graphic overview of a set of concepts and their interrelationships using the structure of SNOMED CT, in contrast to the original Termviz that only allows visualization of a single concept and its supertype ancestors. The terminological features implemented in the prototype are.
- Merged graphs
- Least common parents (LCP)
- Common and non-common parents
- Concept children

From this point, the set of concepts that is to be visualized is called an *Interest Set*.

2.1.1 Merged graphs

The graphs for each of the Interest Set members and parents/ancestors should be merged into a single unified graph to provide a graphic overview of an Interest Set and its hierarchical context. An example of this is shown in ▶ Figure 2.

►Figure 2 illustrates how the hierarchical view of the two concepts 'Blood vessel structure' and 'Cardiovascular structure of trunk' have been manually merged, based on the graphic views generated using the graphic view in the CliniClue browser [19], illustrated in ►Figure 3.

In this study, visualizing | is a | relationships was chosen as they are core for supporting hierarchical comparison of SNOMED CT. It would be possible to add other relationship types to the visualization, as the defined | finding site | for clinical findings and | procedure site | for procedures, but it then also becomes more entangled.

2.1.2 Least common parents

The LCP is the most specific common taxonomical ancestor to two concepts [23], ►Figure 2. In terms of comparability, LCPs are of importance as they can be used for information retrieval purposes, for example when designing a query or report. If a query is to be defined that automatically retrieve all concepts in the Interest Set, the LCP can be considered as the entry point for the query predicate, for example *"Select all subtype ancestors of LCP which are part of the Interest Set"*

For example, if the two concepts of interest in ►Figure 2 should be returned as the result of the above query example, the LCP (| structure of cardiovascular system |) is the nearest common supertype of the two concepts which can be used. Therefore, LCPs are emphasized in the visualization (see ►Figure 6 B). LCPs for three or more concepts are also important and will be emphasized as well.

2.1.3 Common parents and non-common parents

Non-common parents are parents of one concept in the Interest Set but not for any other interest-set concepts. It can be useful to reduce information (e.g. labels) of non-common parents from the visualization to reduce the visual complexity of the graph and focus on comparable content.

2.1.4 Concept children

In order to explore the terminology (e.g. to find alternative sibling concepts), it should be possible to view all the children of any concept. Concept children or siblings are obvious candidates for usage in other parts of the EHR-system to specify related inputs. Also, the children can be used to clarify the semantic interpretation of a given concept.

2.2 Visualization tool

As explained in the introduction, graph visualization becomes challenging as SNOMED CT have multiple parents (| is a |) and several other relationship types. Graphs very soon become crowded and entangled when the number of nodes increases, which often is the case in this study, where we want to merge the graphs from multiple source concepts. Interactively highlighting current points of interest and reducing clutter from other parts of the graph without losing context is a challenging task. [18]

Many of the original TermViz functions [18] have now been re-implemented as open source in HTML5 and JavaScript in order to make the new "TermViz.js" embeddable as a part of any web based terminology service. The prototype in this paper aims to explore ways of illustrating common pathways and ancestors of particular sets of nodes, however, TermViz.js can be customized, extended and used for many other purposes.

2.2.1 Overall system design

The TermViz.js implementation is based on the versatile D3.js [24, 25] JavaScript framework that contains support for many types of graphs including force-based ones. This is possible because modern standards compliant web browsers now have the capacity to efficiently handle graphs drawn in Scalable Vector Graphics (SVG) [26] backed by physics-based simulations implemented in JavaScript.

The overall system architecture is shown in ►Figure 4. The prototype was implemented as a web application and SNOMED CT is stored in a MySQL database. The MySQL database stores the full version of the international release of SNOMED CT and is used to retrieve the supertype ancestors of the Interest Set, and to retrieve the subtypes. JSON (JavaScript Object Notation) is used to repre-

sent the server response (as the graph nodes and links) in a formal and structured way which is easy to interpret on the client side. Furthermore, TermViz.js uses this format as input for the graph generation.

CSS is used to style the graph and the general appearance of the prototype. CSS-friendly class-attributes are assigned to different graph elements by TermViz.js. This makes it possible to adjust the appearance of the graph without learning the JavaScript implementation details.

2.2.2 Control Panel

►Figure 5 illustrates the control panel of the prototype, which takes a concept-Id for a specific concept of interest as input. The input is entered in the text field (entry 1 in ►Figure 5). The resulting concept is shown in the area for 'candidate concepts'(entry 2 in ►Figure 5) by its preferred term. From this area the concepts to be displayed in the graph can be dragged into the Interest Set area (entry 3 in ►Figure 5). Moreover, they can be removed from the graph by dragging the concepts back to the candidate area. At the top of the screen, the buttons for displaying the graph and enabling visualizing the terminological features are available (entry 4 in ►Figure 5).

An example of the resulting graph and the implementation of the terminological features are shown in ►Figure 6. In future end-user applications these sets will likely come from other parts of the application (e.g. search results or CIS template mappings) rather than being manually entered.

2.2.3 Graph visualization

For each Interest Concept the supertypes are determined (and fetched from the database), as they each represent a node in the graph. Duplicate occurrences of supertypes are only included once, and they are also categorized as a non-common parent, common parent or LCP.

The Interest Set illustrated in ►Figure 6 consists of three concepts; 'Cardiovascular structure of trunk', 'Organ System' and 'Blood vessel structure'. ►Figure 6A shows the graph as it is displayed and the Interest Concept labels are highlighted in red. The illustration of the LCPs along with the Interest Concepts, as shown in ►Figure 6 B provides an overview of what concepts could potentially be used as entry points when retrieving specific groups of Interest Concepts (e.g. when querying related SNOMED CT coded data in a CIS). For example, a query for all Interest Concepts that are subtype of the LCP 'Anatomical structure' will include all three Interest Concepts in the. In SNOMED CT the subtype relationships is provided by the | is a | relationships [27]. Retrieving subtypes of the LCP 'Cardiovascular System' will include the two concepts 'Blood vessel structure' and 'Cardiovascular structure of trunk'. The result of toning down the visual appearance of non-common concepts and links is shown in ►Figure 6 C. This produces a graphical way to emphasize common structures of the Interest Set.

Large Interest Sets result in large graphs with many nodes, naturally making an overview difficult to obtain because of the enourmous amount of links. Some concepts have many supertype ancestors that are not placed in the same subhierarchies. An example of this is shown in ►Figure 7 where the single concept "Exploration of rectum" produces a very complex graph structure. Having many of these kind of concepts in an Interest Set will result in an intertwined jumble of links and nodes where other layout algorithms or manual rearrangement of node placements, may be needed to understand the structure.

However, the system is designed so that it is possible to manually drag and change the position of nodes. Thereby selected nodes of particularly entangled graphs can be manually fixed in positions which enhance the possibility to obtain an overview of the graph structures of interest.

►Figure 8 illustrates an Interest Set of a total of 16 concepts. These concepts are a part of an actual physical examination template from a Pediatric department. Even though this results in larger graph than the one in ►Figure 8, it is still possible to obtain an overview and understand the common graph structures.

2.2.4 Hierarchical exploration

Hierarchical exploration is aided by highlighting the concept in focus (i.e. by "mouse-over" a graph node). This is shown in ►Figure 9 where the concept *anatomical or acquired body structure* is in focus (orange shading). Color shading is used to support the identification of the direct parents (red shading) and direct children (yellow shading). A "mouse-over" can also show labels that have pre-

viously been hidden. These features can be helpful when navigating and analyzing large and complex graphs and is discussed in previous TermViz papers [18].

Displaying the children of a graph node is implemented as a list at the right side of the screen in the prototype. This feature also shows other direct children than the ones already visible in the graph. It is also possible to drag the children into the candidate and Interest Set areas. Additional information and node-retrieval features for this view are possible and planned for further development.

2.3 Tool Evaluation

The objective of the study is to assess the potential use of the prototype in the scope of mapping interface terms to SNOMED CT concepts. The evaluation was conducted as a qualitative study where 14 experts in system configuration and clinical terminology from different Danish regions (and vendors) evaluated the visualization tool by using the prototype to solve an assignment and discuss this experience with peers.

2.3.1 Participants

The 14 participants involved in this study represented three groups of stakeholders related to Terminology Implementation and Information Management being Terminology experts, Information Managers and System Developers. These groups were selected as they are all potential end users of the visualization tool.

- Terminology experts (4), from
  - the Danish SNOMED CT National Release Center (NRC) and
  - suppliers of national terminology software
- Information Managers (5) from
  - regional representatives (system owners)
- System developers and health IT managers (5) from
  - vendor representatives

A total of 19 people were invited to participate and 14 people attended the workshop.

2.3.2 Approach

Firstly, the participants worked with the system through a defined mapping assignment. Secondly, focus group interviews were conducted to reveal the participants user experiences.

Mapping assignment
The participants were divided into four small teams (of three to four people). This gave the possibility for the team-members to cooperate, discuss and support each other in the mapping process. As none of the participants were familiar with the visualization tool, being part of a team made the participants feel comfortable with the assignment. The duration of the assignment was one hour. The material of the assignment was a clinical documentation template from an actual EHR-system. The template was a physical examination – emergency department journal template. However, the results of the mapping were not intended to serve as result of the assignment. I.e. the purpose was not to compete in mapping as many concepts as possible or to compare the mappings between the groups, but merely to promote the participants to work with the system and gain experience with using the visualization to select concepts for the mapping. With the qualitative study approach, the purpose of the assignment was primarily to elicit the participant's immediate experience with using the tool and the terminological features. The other purpose of this assignment was to discuss visualization as a tool for achieving an overview of the terminology covered in a clinical template.

Each team was asked to map the clinical template using the developed visualization prototype. The mapping should support hierarchical comparison of related terms, as described as quality criteria for mapping in an earlier study [15]. To support hierarchical comparison, clinically related expressions should be represented by concepts from the same subtype hierarchy. Therefore, it was important to assess how the LCP feature met the quality criteria. This is illustrated in ▶ Figure 10, where two mapping proposals for the same clinical expressions, show the importance of selecting

concepts which are hierarchically related in order to obtain a LCP which is specific for the concepts it covers.

Since the visualization tool was developed as a prototype that visualizes sets of concepts and thus does not contain SNOMED CT browsing-functionality yet, the teams used the browser functionality provided in the '*HealthTerm*' *application*, developed by Carecom A/S [28], to search the SNOMED CT concepts and find the concept IDs which was used as input for the visualization tool. Health Term was used as it allow SNOMED CT search using Danish descriptions.

Focus group interviews

The interview was conducted immediately after the assignment session to obtain the participants' immediate response. A focus group interview was a qualitative research method in which the participants were asked about their perceptions and opinions [29, 30]. This approach also allowed the possibility to explore nuances within the answers and ask clarifying and supplementary questions.

Two parallel focus group interviews were conducted, to meet a recommended group size of not more than 10 people [30]. There were two overall themes of the interviews, namely 'Potential of the visualization tool' and 'potential improvements/refinements for the tool'. These very broad themes were selected to support an outspoken and un-constrained discussion. The data analysis was done by transcribing the interview and categorizing the topics of discussions, as described in [31]. The citations presented in the results sections were translated and meaning-condensed by the authors to present the statements of the participants' statements

3. Results

The participants experience with using the system was discussed during the interviews. The two overall themes resulted in three categories describing the discussed visualization potentials and another three that suggested refinements, see ▶Figure 11. Responses given by interviewees fall under the following topics:

3.1 Visualization potential

Generally, the participants expressed a positive attitude towards the tool and they liked visualization as a mean to get an overview of the terminology. However, the use cases and potentials for the tool varied across interviewees.

Visualization assisting SNOMED CT learning

One study participant stated that the tool was too difficult for SNOMED CT novices:

"…this tool is not for novices, I think it is a necessary tool for development… but I believe that we need a method for communicating with novices and this is probably not that tool…"

Another participant argued that using the tool supports people with a very visual learning approach. By using a graph to show the complex structure concepts and their interrelationships, interpretation and overview becomes much easier. The same person also stated that it is important, when working with such complex systems, that the functionalities in the systems support different learning styles.

Another interviewee stated that the tool would be useful as a learning tool for clinical end users.

"I believe it can be a good tool for providing learning for clinicians and supporting them to understand this very abstract terminology."

Visualization for discussing and defining the semantic content of CIS with end-users

In both interviews it was discussed whether the tool would be valuable to apply in a workshop session were clinicians discuss and define the content for the CIS. There were different perspectives

regarding this. Some stated the complexity and the time-factor to be a reason for not bringing the tool into such workshops:

> "I'm concerned that it'll be too time consuming to use it (the tool) actively in a workshop… I think that we can use it locally in the IT-department"

Others stated that it would be interesting to use the tool in a workshop with clinicians to underpin the discussion on the content applied in the systems.

> "… you can use the terminology as a control stick, as a control when you're out at workshops in relation to propose a sample space. I believe that it can provide some support for this, but more and more I believe that it must be prepared at the office."

Visualization supporting mapping

Numerous study participants agreed that visualization was valuable for the mapping process. It was pointed out that the tool very clearly shows when concepts are not related and, therefore, supports the process of selecting the concepts when the CIS is configured.

> "I believe the tool illustrates very well when concepts are not related. When concepts are placed in separate hierarchies, the system may help to clarify and select concepts."

To support this, another interviewee stated the potential to use the tool to support consistency and coherence among the concepts applied in the CIS.

> "You get the opportunity to see how many conceptual types you've thrown into this registration form and decide if that is what you want with this recording."

3.2 Tool refinement

Even though it was emphasized when the tool was introduced to the stakeholders as a prototype, they expressed different ideas and opinions about refinements and further development.

Add-on

During the group work assignment it was stated that the tool cannot stand alone in its current form, but it would work as an add-on to a typical SNOMED CT browser for displaying search results. This was also stated during the interview:

> "For me, it is like an add-on in a browser, visualizing the resulting concepts with respect to their placement in the hierarchies."

However, judging by the observations and the limited amount of questions which specifically concern the system features, the stakeholders quickly became familiar with the workflow supported by the system, and the terminological and clinical discussion was dominating the group discussions.

Graph layout

Some of the interviewees expressed some concern about the graph waggling and it was argued that the inter-activeness of the graph design actually made the system somewhat confusing to use.

> "But I think that it is difficult to use, too, because the concepts are so mobile, and then suddenly it appears in one way and all of a sudden it appears in another way. And, if you do not know what lies behind, you can still almost only get confused. So, I think it's a good tool, I think it's a great tool. But I also think that it requires a significant prerequisite to be able to understand it."

The above statement is due to the fact, that every time a node is entered, the surrounding nodes move to make place for the labels that becomes visible. For the same reasons, it was argued that manipulating the placement of the nodes should be comfortable for the system user. During the group work, it was suggested that the tool could be improved by providing a static view of the graph, so that it would be possible to switch between the interactive and a static layout.

Visualize more relationship types and use color for subtype hierarchies

One participant stated the desire to be able to see the different attribute relationship types related to each node, not only the | is a | relationships. He stated that common attribute relationships are also important for ensuring coherence between the concepts.

> "I think that one should make the tool – you have to further develop the tool so that the cross- relationships also are visualized. Also, one could maybe make a functionality where you could say – work with or without cross-relationships."

Another interviewee stated that it would be nice if more effort was put into using coloring for visualizing the different subtype hierarchies.

4. Discussion

The prototype presented in this paper was developed with the objective to explore applications for sets of concept visualization. The potential use of this SNOMED CT visualization had to our knowledge not previously been explored by expert evaluation. This section, presents three potential applications of the visualization derived from the tool evaluation along with possible system refinements.

4.1 Support SNOMED CT searches and training

A limitation of the developed prototype is that it doesn't allow searching. An increasing number of different and well-developed SNOMED CT search engines exist. In [17] 17 different SNOMED CT browsers were inspected, and the study emphasized the need for further research to determine what search techniques are required for specific use cases. Analysis of what search techniques would be interoperable with set of concepts visualization is beyond the scope of this study. However, the results of our evaluation suggest that visualization could support and enhance SNOMED CT searches. Today, searching for "diabetes" in most browsers gives a long list of lexical matches from multiple sub-type hierarchies. Choosing the appropriate term from this list can be a challenge if the terminological context is not apparent from the list. Using the list of lexical matches as an Interest Set, would show the context graphically (e.g. by visually separating the concepts that are specializations of clinical findings from concepts that are specializations of other top-hierarchies). Further research should therefore explore how set of concept visualization could be combined with SNOMED CT searches.

The results of evaluating the prototype also indicated that the prototype could be useful for learning about SNOMED CT, its content and structure. In contrast to existing SNOMED CT browsers that use a static graph layout, this tool uses an interactive forced based layout, allowing nodes to be moved, and highlighted. On one hand, this layout provides the possibility to visualize very large and informative graphs, however, on the other hand, the result of the evaluation indicated that the current version of the interactive, force based layout of the tool could be disruptive for interpreting complex graphs.

The ontology management system Protégé support visualization through Ontograf [32]. This tool supports various graph layouts, interactive hierarchy navigation and graph zooming. Moreover, Ontograf support selective visualization of different relationship types, which is one of the features requested in our evaluation. We believe that the range of features implemented in Ontograf can support SNOMED CT learning, as support different ways of examining the terminology content and structure. However, to our knowledge, Ontograf does not include set of concepts visualization nor the terminological features of LCP and common paths which is the primary focus of this study.

© Schattauer 2014

A.R. Højen, E. Sundvall, K.R. Gøeg: Methods and applications for visualization of
SNOMED CT concept sets.

Combining the terminological features of our tool with the features supported in Ontograf could therefore provide a powerful visualization framework for SNOMED CT learning and search.

4.2 Support terminological overview and consistent concept selection

Visualizing a set of concepts and their supertype ancestors can provide a graphic overview of the concepts of interest and their hierarchical interrelationships. In contrast, a list of concepts represented by the concept descriptions, as illustrated in the left side of ►Figure 12 does not provide any information about the semantic interpretation of the expression. Our study suggest that visualization can support concept selection when doing terminology implementation, because the relationships between the candidate concepts and the different conceptual types becomes clear. This finding corresponds to the finding of another study [33] comparing effectiveness and efficiency of two visualization techniques, respectively indented tree and graph visualization, it was found that graph visualization is more suitable than indented trees for providing an overview of concepts.

From our experience terminological overview and awareness about concept interpretation could potentially reduce coding variability. As described in [8, 14] coding variability occurs even amongst coding experts (the people who perform the actual mapping). The developed visualization tool, along with detailed mapping guidelines, described in [15], could potentially support consistent concept selection, by providing an overview of the concepts placement in the SNOMED CT hierarchy. If the coding experts agree that coherency should be supported, the tool could potentially support consistency because it clearly shows if a single concept is placed in another hierarchy than the related hierarchy. Thus, further studies should evaluate the implications of applying these tools and assess whether coding variability can be reduced.

4.3 Supporting content management and query design

The prototype takes an Interest Set consisting of any number of SNOMED CT conceptIds and merges their supertype ancestors into a single graph. This means that a meaningful link between CIS and terminology implementation can be established, because the visualization can support an overview of the terminology applied in a specific clinical data entry template, and support specification of what concepts to use in the template. To strengthen the link between clinical data entry templates and terminology further, using multiple simultaneous Interest Sets as input could be implemented as a feature. This would provide the possibility to compare concepts from different clinical templates with each other and would further make it possible to analyze on information shared across clinical templates. However, working with multiple Interest Sets will increase the graph size and complexity, and thus, future work should strive to improve visualization of large graphs, for example by applying clustering techniques [22].

Comparing two sets of concepts by using visualization is illustrated in ►Figure 13. The dark grey boxes are concepts used in one template and the light grey boxes are concepts from another template. The graphical overview shows the coherence between the content in the two templates and support reuse of data across these. For example, if we want to retrieve data recorded about a patients *Neurological findings* the illustration explicate what concepts in the two templates are subtypes of this, namely *Pain*, *Observation of sensation*, *Finding of sense of taste*, *Finding of taste of smell*. In [34] it is demonstrated how utilization of the SNOMED CT concept model can simplify query development and maintenance. The automated way of utilizing SNOMED CT in the example in ►Figure 13 is to query for all concepts, part of the two templates (or all of SNOMED CT-based clinical content), which are subtypes of *Finding of sensory nervous system*. In systems that do not support these types of queries, the visualization can provide a pragmatic way of utilizing the structure of SNOMED CT without being dependent on any retrieval functionalities. We therefore see a potential use of applying visualization to support content management and query design.

Conclusions

Ensuring coherent and consistent implementation of clinical terminologies may bring us a step closer to meeting the vision of interoperability between CISs, enabling the reuse and common interpretation of clinical data across system borders. Visualizing the hierarchical relationships between sets of concepts, to support coherent concept selection has been the main motivator behind this study. Hence, this study documents the development and evaluation of a system where sets of concepts and their interrelations are visualized related to the hierarchical composition in SNOMED CT. This prototype provides an overview of concepts and their hierarchical relations, and has the potential to support various purposes, such as SNOMED CT search and training, consistent selection of SNOMED CT concepts and content management.

One contribution of this study is the TermViz.js toolkit which is released as an open source toolkit and can be used in different setups for different purposes. This makes it possible for others to further develop tools to enhance terminology exploration and usage. Further work should improve usability and expand the range of terminological features supporting the explored applications of the system. In future work, evaluation of the explored potentials of applying this tool will be an important priority.

Clinical Relevance Statement

- It is important with technologies and methods that support implementation and use of complex clinical terminologies, as SNOMED CT, to enhance future applications of SNOMED CT. Visualization of complex graphs can be supported by specific features to support overview and interpretation of SNOMED CT concepts and sub-hierarchies. This study suggest four terminological features to support overview, comparison and interpretation of SNOMED CT concepts, i.e:
  - Merged graphs, Least common parents, Common and non-common parents and concept children
  - Interactive visualization of sets of SNOMED CT concepts has potential to support a range of use cases, such as: SNOMED CT search and training, consistent selection of SNOMED CT concepts and content management in clinical information systems.
- The TermViz.js toolkit based on D3.js is released as an open source toolkit aimed at terminology visualization and can be used in different setups for different purposes. This makes it possible for others to further develop and integrate tools to enhance terminology exploration and usage.
- Further tuning and development of the force based layout is needed in order to improve the user experience when browsing SNOMED CT, especially when interactively adding more nodes to an already crowded graph.

Statement on conflicts of interest
None.

Protection of Human and Animal Subjects
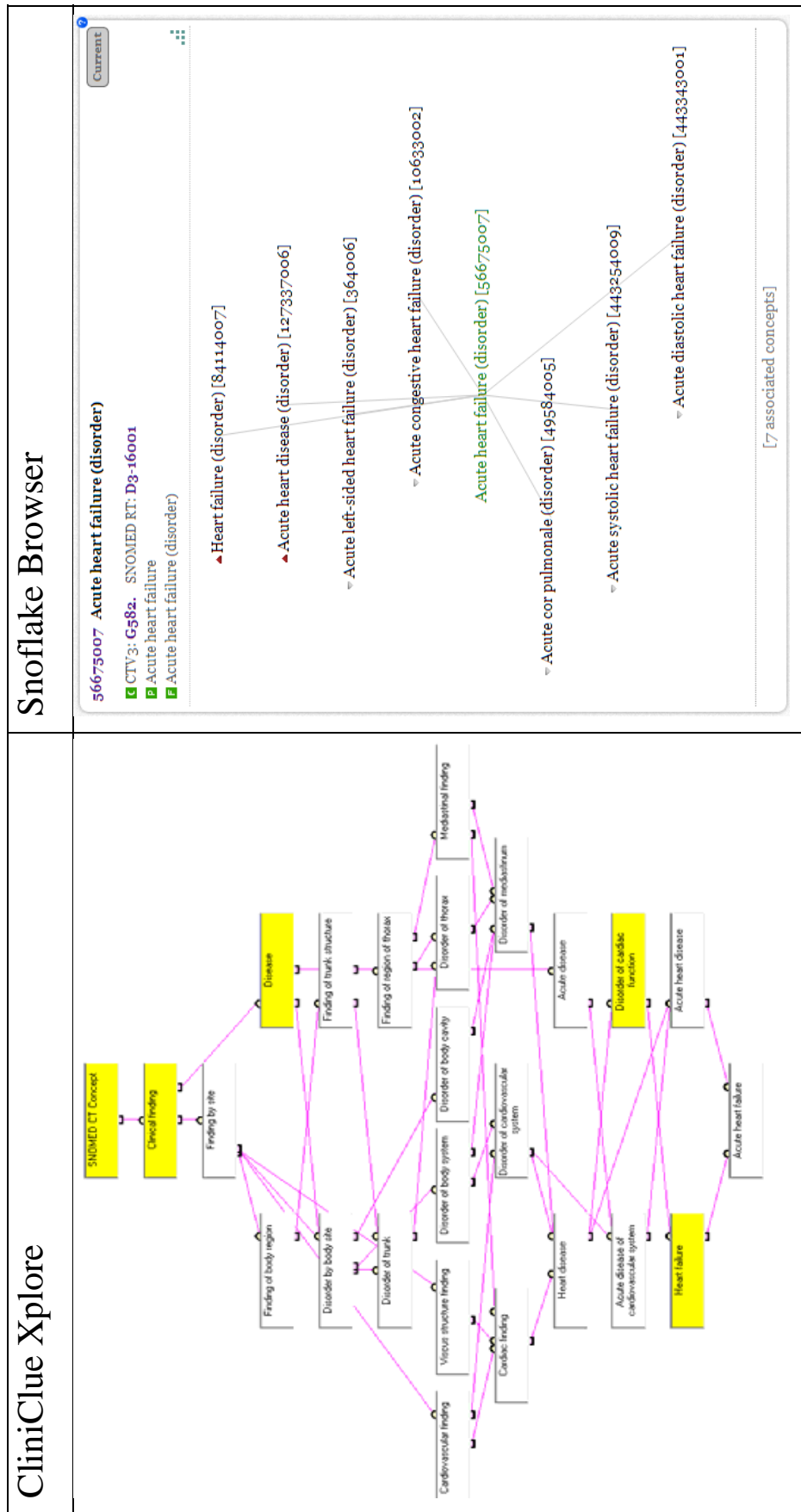No human and/or animal subjects were included in this project

Fig. 1 Illustration of visualizations provided in two existing SNOMED CT browsers. To the left the graphic view supported in CliniClue Xplore and to the right the visualization in the Snoflake Browser.
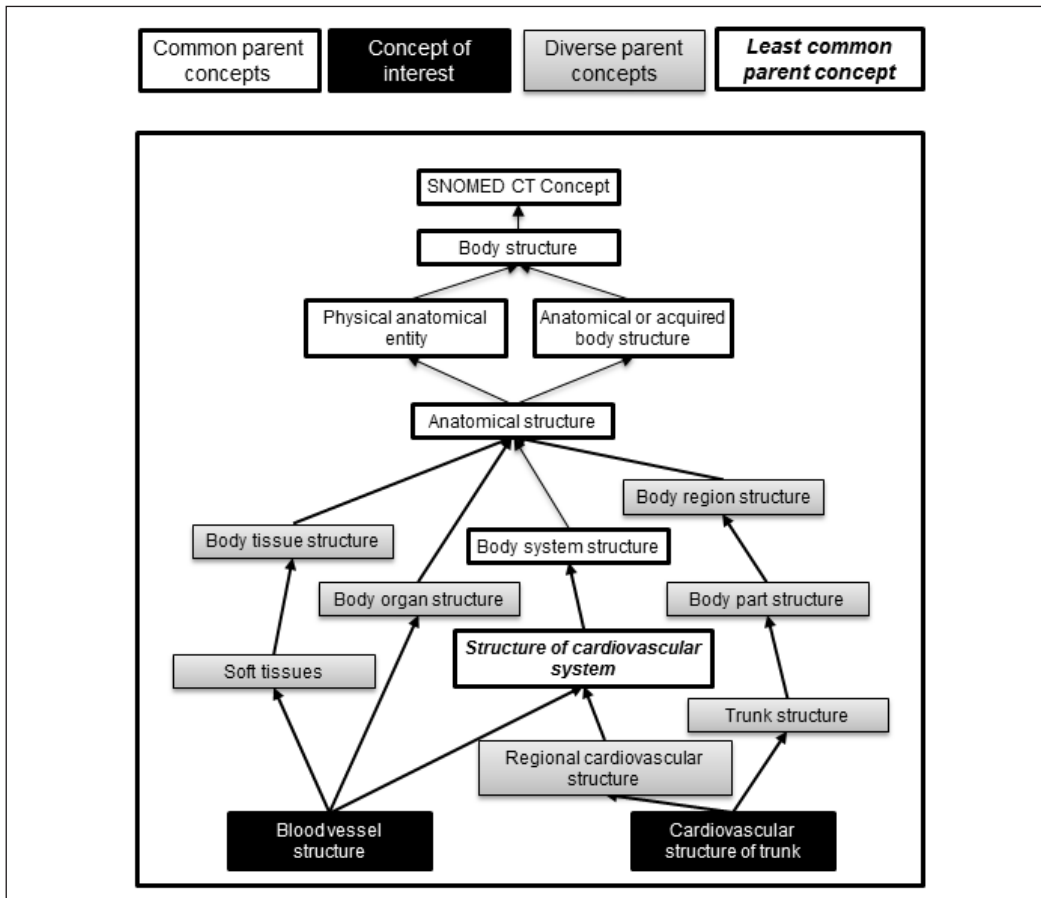
Fig. 2 Exemplification of the merged graph of two concepts of interest, 'Blood vessel structure' and 'Cardiovascular structure of trunk'.
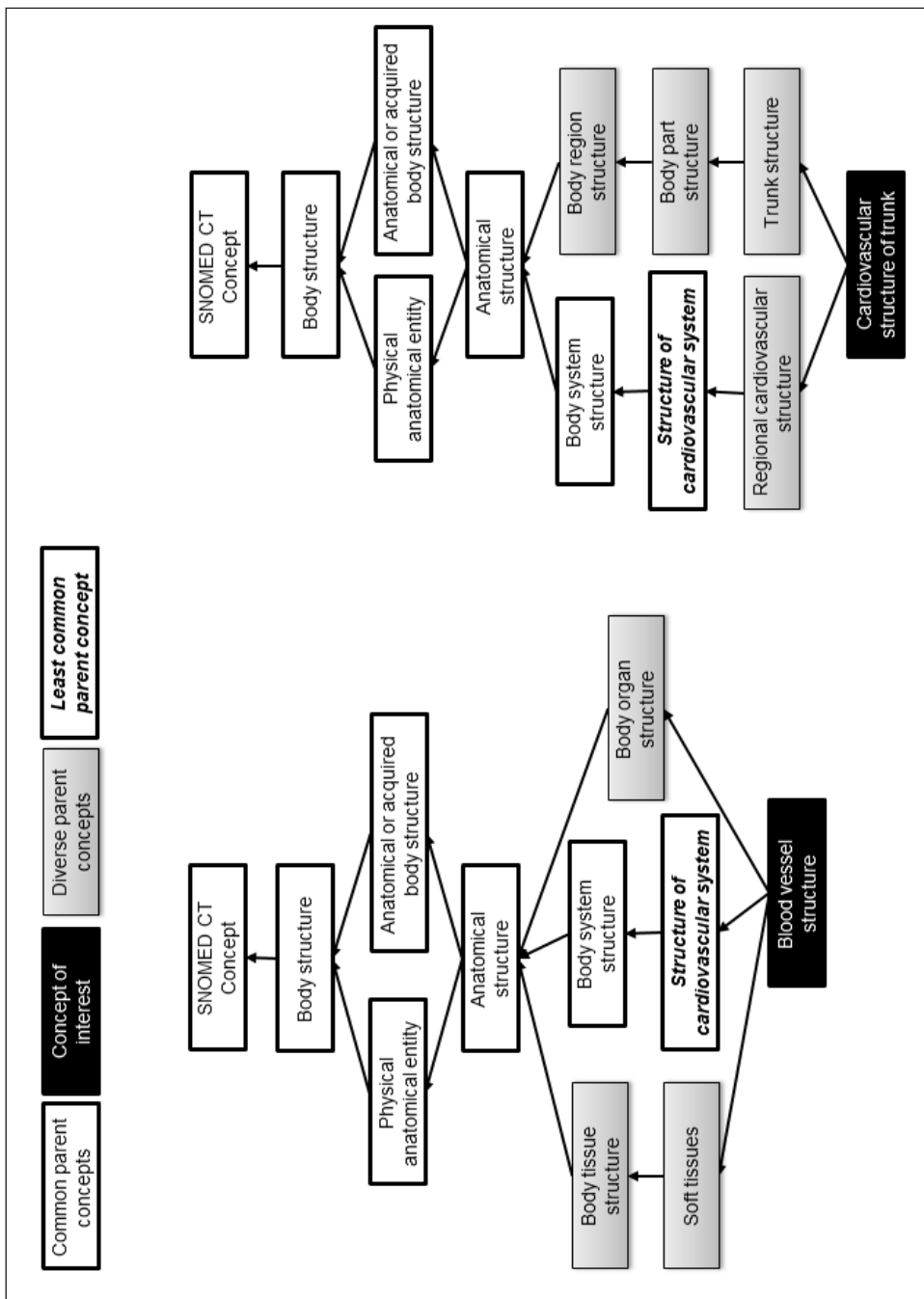
Fig. 3 Graphic views of the two concepts "Blood vessel structure" and "Cardiovascular structure of trunk", generated by CliniClue browser, exported and re-colored by the authors.
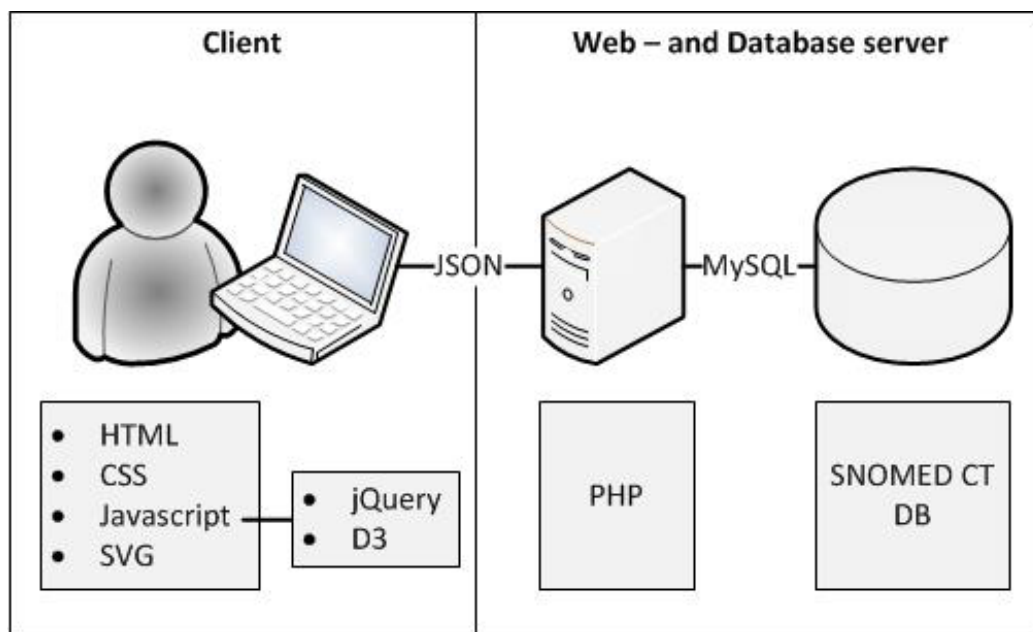
Fig. 4   Illustration of the overall architecture. The prototype is implemented as a web-application that retrieves SNOMED CT content from a MySQL database and processes the data using PHP, JavaScript and TermViz.js.
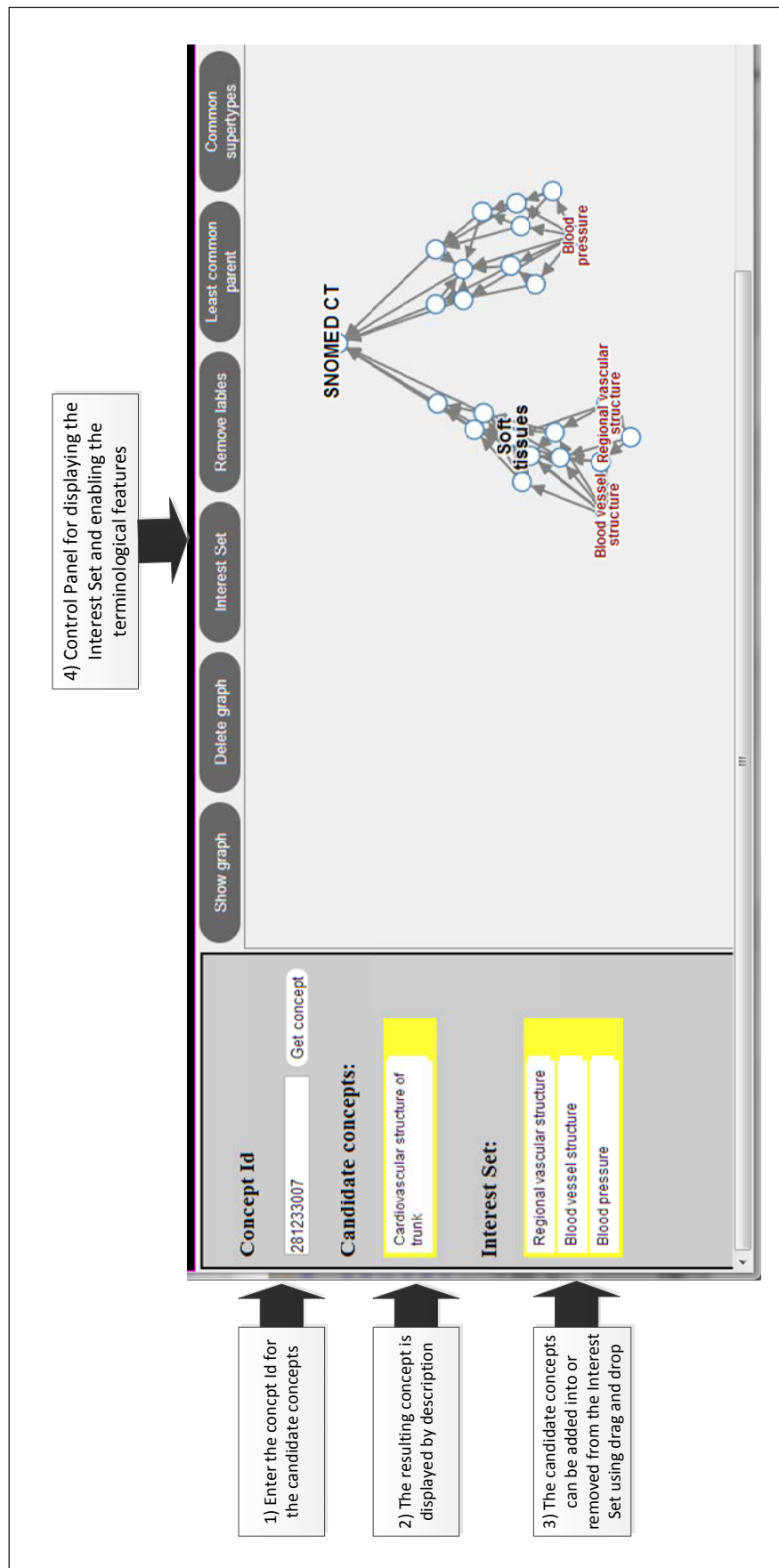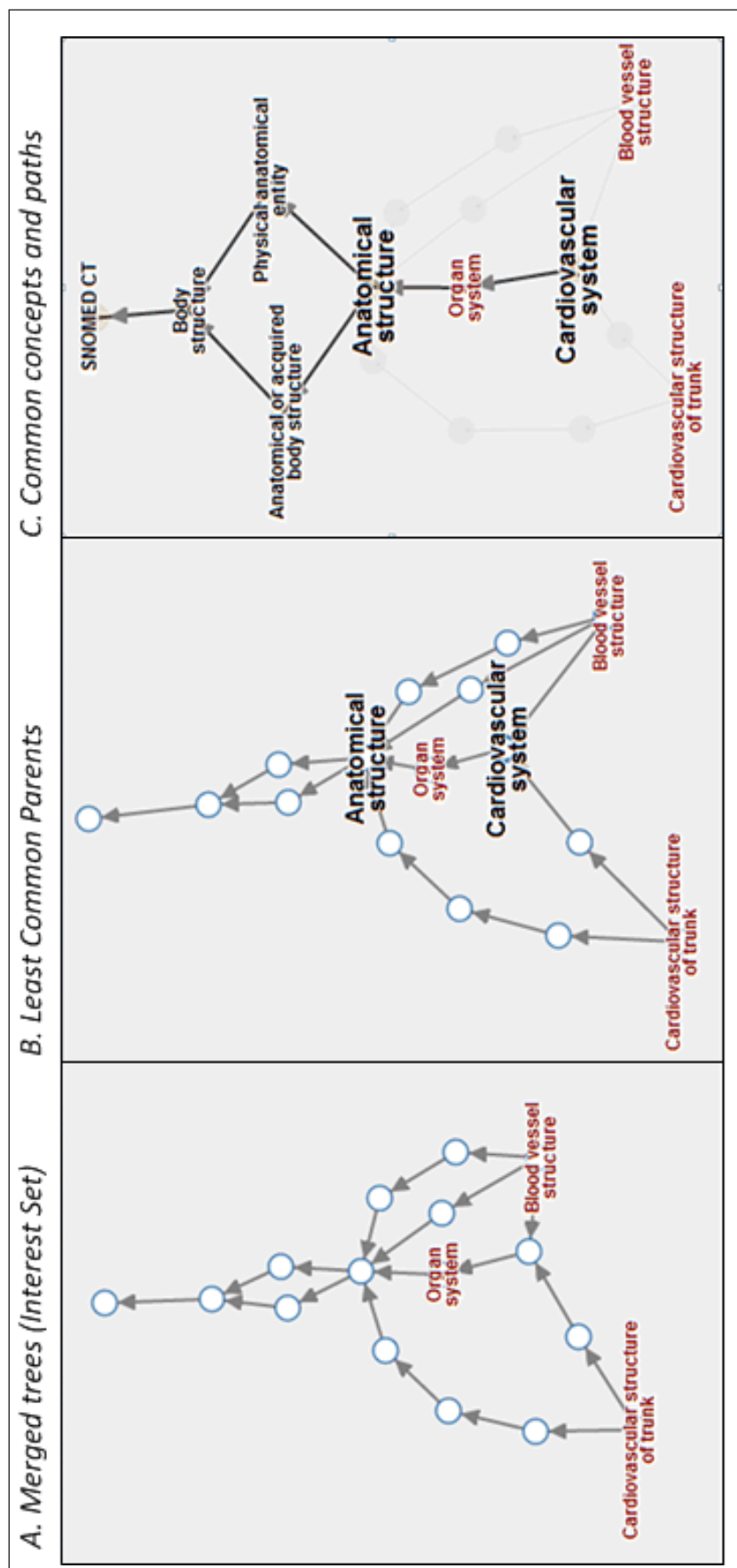
4) Control Panel for displaying the Interest Set and enabling the terminological features

1) Enter the concpt Id for the candidate concepts

2) The resulting concept is displayed by description

3) The candidate concepts can be added into or removed from the Interest Set using drag and drop

Fig. 5   Screen dump of the protoypes controlpanel

A.R. Højen, E. Sundvall, K.R. Gøeg: Methods and applications for visualization of SNOMED CT concept sets.

A. Merged trees (Interest Set)   **B. Least Common Parents**   **C. Common concepts and paths**

Fig. 6 Illustration of the graphical view of an Interest Set consisting of three Interest Concepts (marked red in the graphs). Figure A shows the merged tree and the highlighted Interest Concepts. Figure B shows the Least Common Parents, and figure C shows all common concepts and links in the graph. The labels of un-labelled nodes are revealed by "mouse-over".
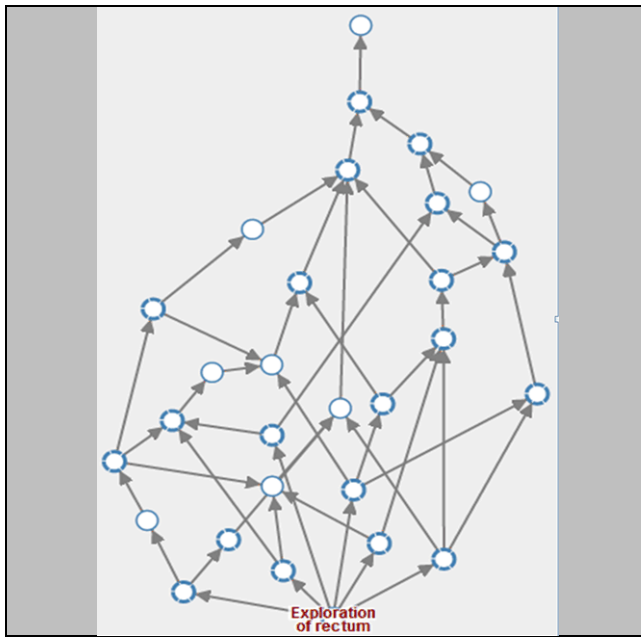
Fig. 7
Example of the concept 'Exploration of rectum' which has a very complex graph structure.
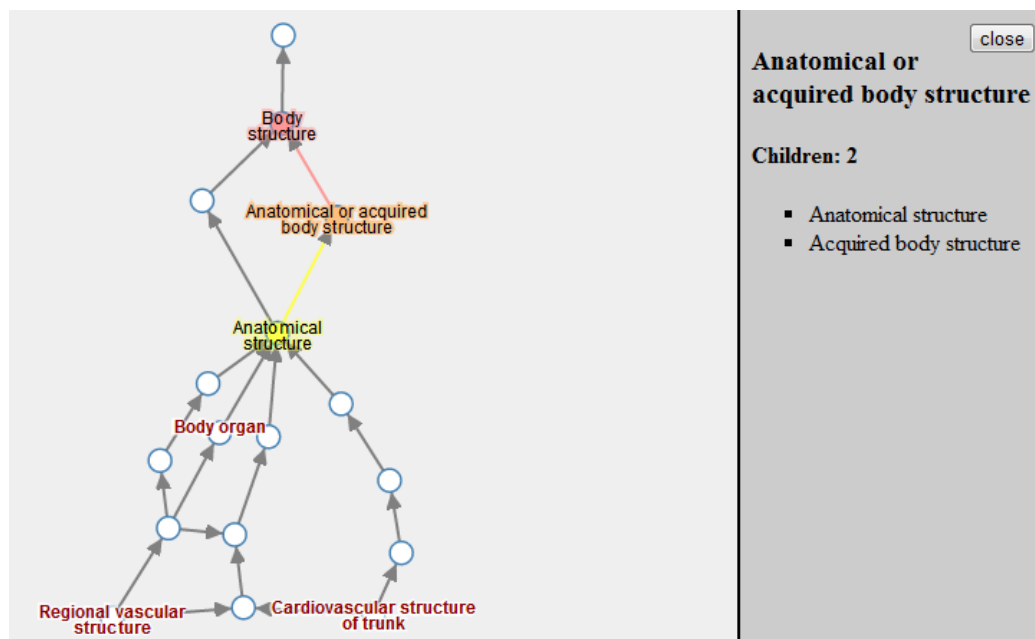
Fig. 8 Illustration of the graphical view of an Interest Set consisting of 16 Interest Concepts (marked red in the graphs). Figure A shows the merged tree and the highlighted Interest Concepts. Figure B shows all common concepts and links in the graph.

Fig. 9 Hierarchical navigation features in the prototype. Mouse-over functionalities applies colors on the concept in focus and its direct parent and child concepts. Also all direct children (not only the ones in the graph) is displayed in a list to the right of the graph
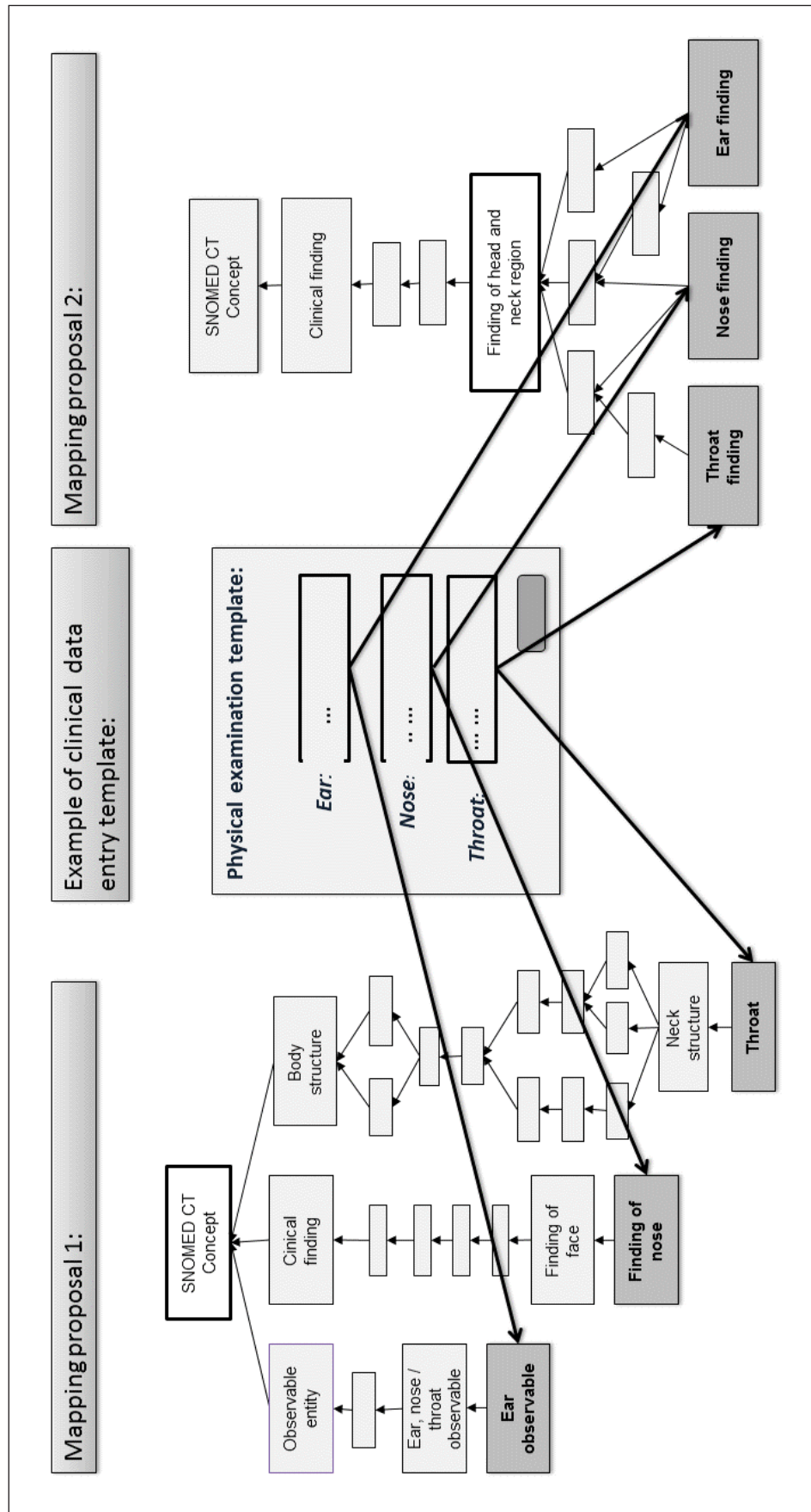
Fig. 10 An example of two different mapping proposals for the same clinical template. The LCP in 'mapping proposal 1' is the "SNOMED CT concept", whereas the LCP of 'mapping proposal 2' is "Finding of head and neck region". 'Mapping proposal 2' therefore meet the quality criteria with a very specific LCP describing for the three concepts. Modified from [15]
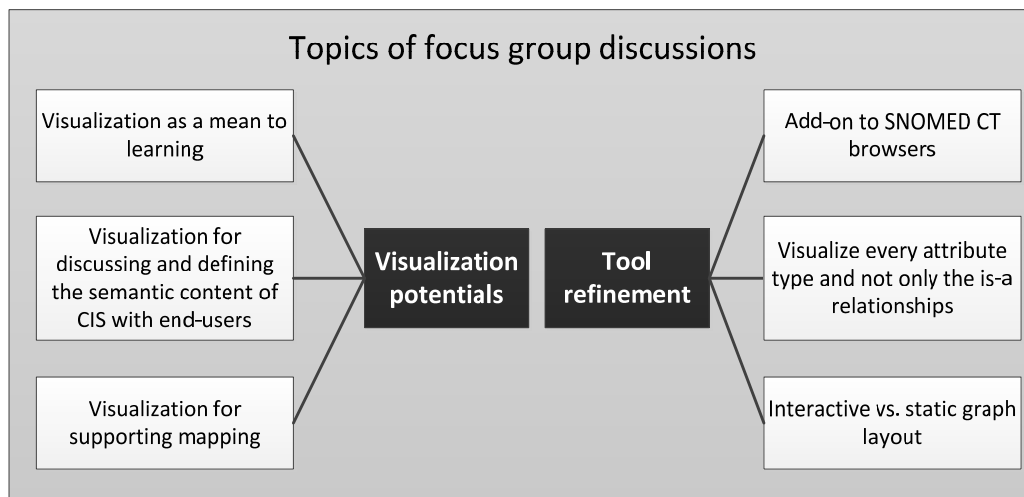
Fig. 11 Overview of categories derived from the two overall themes of the Focus Group discussions
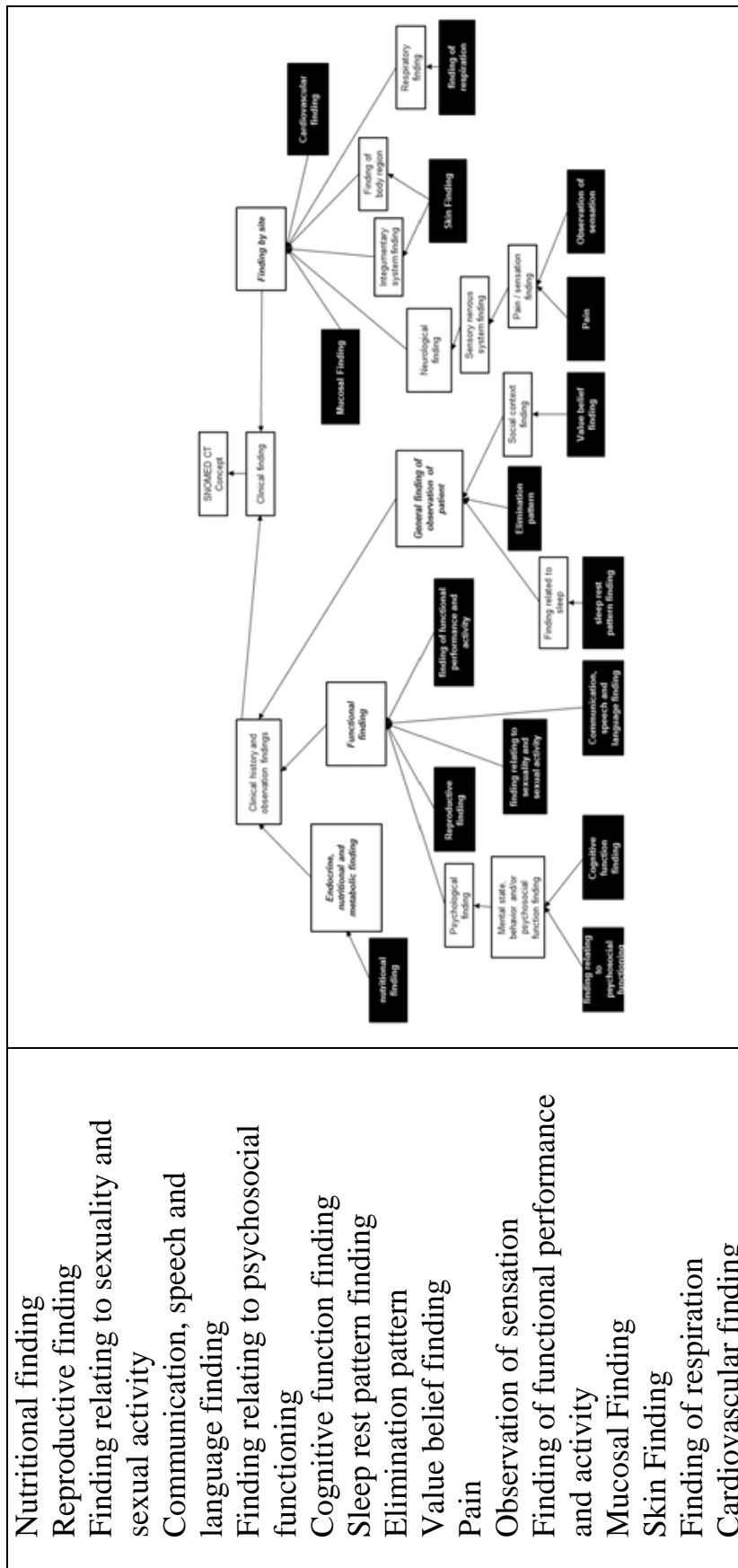
Fig. 12 List of concepts representing part of the content to be included in a physical examination template (Left). An illustration of graphical view of the same concepts (Right).
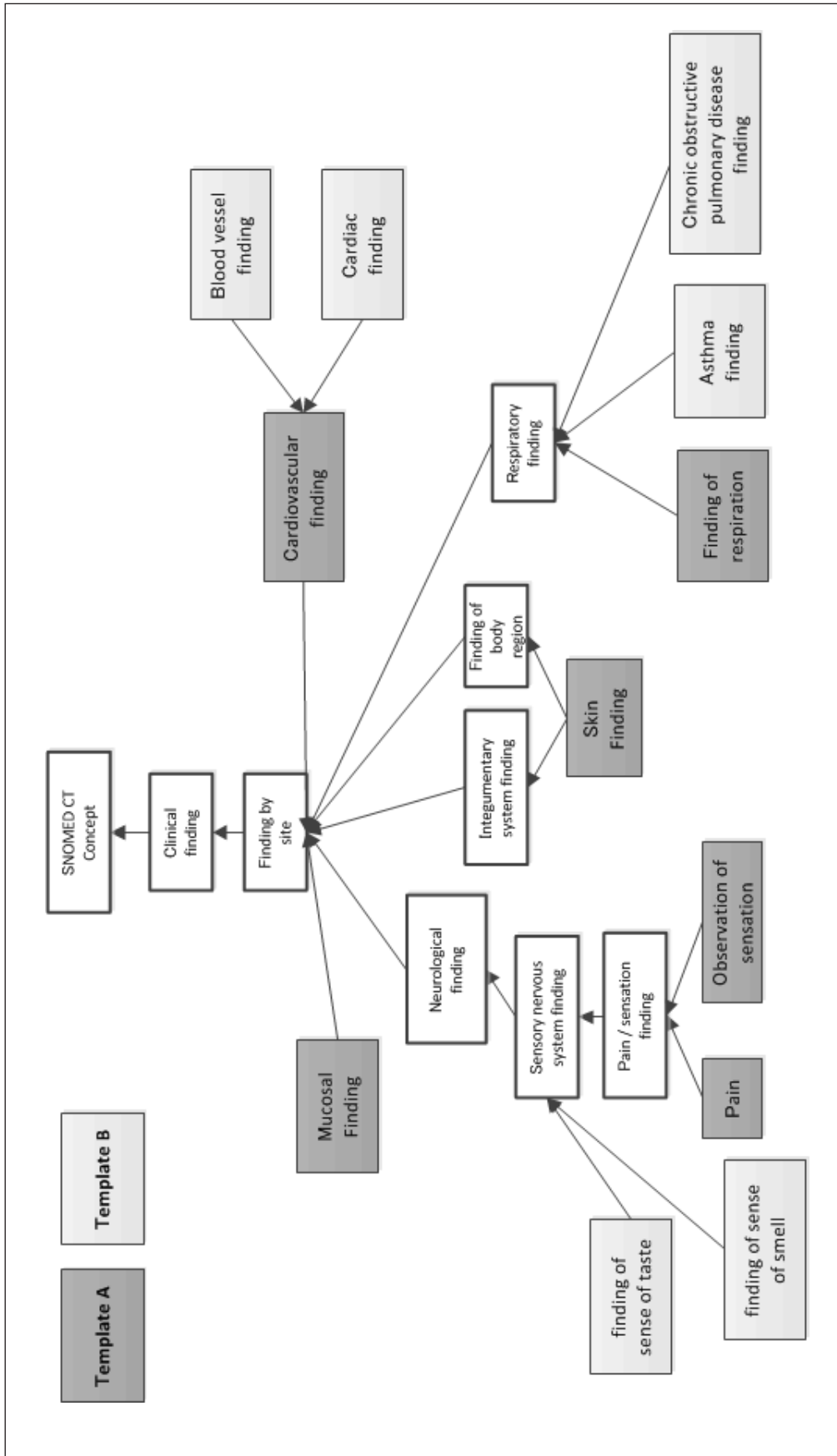
© Schattauer 2014

A.R. Højen, E. Sundvall, K.R. Gøeg: Methods and applications for visualization of SNOMED CT concept sets.

Fig. 13 Illustration concepts from two different templates.

References

1. Stroetmann VN, et al. Semantic interoperability for better health and safer healthcare, 2009 ISSN 978–92–79–11139–6.
2. Ryan A, Eklund EA. Framework for semantic interoperability in healthcare: A service oriented architecture based on health informatics standards. MIE ed. IOS Press, 2008.
3. Goossen WT. Intelligent semantic interoperability: Integrating knowledge, terminology and information models to support stroke care. Studies in Health Technology and Informatics 2006; 122: 435–439.
4. Garde S, Knaup P, Hovenga E JS, Heard S. Towards semantic interoperability for electronic health records - domain knowledge governance for open EHR archetypes. Methods of Information in Medicine 2007; 46: 332–343.
5. Chute CG. Clinical classification and terminology some history and current observations. Journal of the American Medical Informatics Association 2000; 7: 298–303.
6. Elkin PL et al. Secondary use of clinical data. Studies in Health Technology and Informatics 2010; 155: 14.
7. Cimino JJ. Collect once, use many: Enabling the reuse of clinical data through controlled terminologies. JOURNAL-AHIMA 2007; 78: 24.
8. Andrews JE, et al. Comparing heterogeneous SNOMED CT coding of clinical research concepts by examining normalized expressions. Journal of Biomedical Informatics 2008; 41: 1062–1069.
9. Bos L. SNOMED-CT: The advanced terminology and coding system for eHealth. Medical and Care Compunetics 2006; 121: 279.
10. Wade G. Implementing SNOMED CT for quality Rreporting: Avoiding pitfalls. Applied Clinical Informatics 2011; 2: 534.
11. Lee D, Lau F, Quan H. A method for encoding clinical datasets with SNOMED CT. BMC Medical Informatics and Decision Making 2010; 10: 53.
12. Wade G, Rosenbloom ST. Experiences mapping a legacy interface terminology to SNOMED CT. BMC Medical Informatics and Decision Making 2008; 8.
13. Hansen DP, et al. Building SNOMED CT Reference sets for use as interface terminologies. Electronic Journal of Health Informatics 2011; 6: e1.
14. Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. Journal of the American Medical Informatics Association 2007; 14: 497–506.
15. Højen AR, Goeg KR. SNOMED CT Implementation. Mapping guidelines facilitating reuse of data. Methods of Information in Medicine 2012; 51: 529–538.
16. Rasmussen RA, Rosenbeck K. SNOMED CT Implementation: Implications of choosing clinical findings or observable entities 2011.
17. Rogers J, Bodenreider O. SNOMED CT: Browsing the browsers. KR-MED 2008; 30–36.
18. Sundvall E et al. Interactive visualization and navigation of complex terminology systems, Exemplified by SNOMED CT. Studies in Health Technology and Informatics 2006; 124: 851.
19. The Clinical Information Consultancy Ltd. CliniClue – Clinical terminology services. 27.08.2012 Available from: http://www.cliniclue.com/.
20. Dataline Software Ltd. SNOFLAKE browser. Available from: http://www.snoflake.co.uk/#.
21. Rubin DL, Noy NF, Musen MA. Protege: A tool for managing and using terminology in radiology applications. Journal of Digital Imaging 2007; 20: 34–46.
22. Herman I, Melançon G, Marshall MS. Graph visualization and navigation in information visualization: A survey. Visualization and Computer Graphics, IEEE Transactions On 2000; 6: 24–43.
23. Sánchez D, Batet M. Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. Journal of Biomedical Informatics 2011; 44: 49–759.
24. Bostock M, Ogievetsky V, Heer J. D³ Data-driven documents. Visualization and Computer Graphics, IEEE Transactions On 2011; 17: 2301–2309.
25. Data-driven documents. Available from: http://d3js.org/.
26. Scalable vector graphics (SVG). Available from: http://www.w3.org/Graphics/SVG/.
27. IHTSDO. SNOMED clinical terms. Technical Implementation Guide. January, 2012–07–31.
28. Carecom A/S. Carecom. 2013 Available from: http://www.carecom.dk/.
29. Kvale S. Doing interviews. SAGE Publications Limited, 2008.
30. Barbour R. Doing focus groups. SAGE Publications Limited, 2008.
31. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. American Journal of Evaluation 2006; 27: 237–246.
32. Falconer S. OntoGraf, 2010.
33. Fu B, Noy NF, Storey M. The semantic web–ISWC 2013 Springer, 2013 Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation 117–134.

34. Lieberman MI, Ricciardi TN. The use of SNOMED© CT simplifies querying of a clinical data warehouse. American Medical Informatics Association, 2003.

© Schattauer 2014

A.R. Højen, E. Sundvall, K.R. Gøeg: Methods and applications for visualization of SNOMED CT concept sets.