Semantic Visualization of Patient Information

Sonja Zillner\textsuperscript{2,4}, Tamás Hauer\textsuperscript{1,3}, Dmitry Rogulin\textsuperscript{1,3}, Alexey Tsymbal\textsuperscript{2,4}, Martin Huber\textsuperscript{2,4}, Tony Solomonides\textsuperscript{1,5}

\textsuperscript{1}CCS Research Centre, CEMS Faculty, University of the West of England
\textsuperscript{2}Corporate Technology Division, Siemens AG, Germany
\textsuperscript{3}first.last@{cern.ch, siemens.com}
\textsuperscript{4}first.last@uwe.ac.uk

Abstract

Clinical practice and research rely increasingly on analytic approaches to patient data. Visualization enables the comparative exploration of similar patients, a key requirement in certain clinical decision support systems. Patient data is complex and heterogeneous, may have different formats, reside in various structures and carry different semantics. This makes the comparison and analysis of clinical data a challenging task. Most medical applications visualize patient data without integrating additional semantic information to structure the analysis. Our objective is to map patient data onto relevant fragments of ontologies and inferred ontological structures as a basis for improved patient data visualization, comparison, and analysis. Two visualization scenarios that we have implemented using the patient data acquired in the Health-e-Child project will be presented and their clinical evaluation will be provided.

1. Introduction

Clinical care and research rely increasingly on digitized patient information. There is a growing need to store and organize all patient data that may contribute to case evaluation or clinical study. While information is getting richer with recording more and different kinds of patient attributes, the visualization of the overall picture becomes increasingly harder for the clinician or the researcher. Discussions with clinicians collaborating in the EU project Health-e-Child\textsuperscript{1} revealed that, the clearly arranged presentation of similar patients becomes a key requirement for medical decision support systems. For clinicians, the comparison of similar patients is a valuable source of information in the process of decision finding.

Health-e-Child is an EU-funded Framework Programme 6 project aimed at improving personalized healthcare in selected areas of pediatrics, especially focusing on integrating data across disciplines, modalities, and vertical levels such as molecular, organ, individual and population. The project of 14 academic, clinical, and industry partners aims at developing an integrated healthcare platform for European pediatrics and testing it in the domain of selected representative diseases in three different areas: congenital heart defects, inflammatory diseases, and brain tumours.

The integration of medical data along multiple dimensions of heterogeneity is rather challenging, yet it is crucial that clinicians have access to a coherent view of the information that has been collected. The results presented in this paper contribute to the development of data integration and visualization components of the Health-e-Child system. They provide the clinicians with a tool to easily browse patient data and help visualizing complex information such as correlations in order to establish new clinical hypotheses.

With traditional applications, users may browse and visualize patient data but little to no help is given when it comes to interpretation because the required semantics is implicit and thus inaccessible to the system. In this paper, we will show that the incorporation of medical ontological knowledge does contribute to improved and concise patient data visualization. Our objective is to align patient data with relevant (fragments of) ontologies and to infer a more descriptive patient ontology as a basis for improved patient data visualization, comparison and analysis. By mapping patient data attributes onto ontological concepts, the inherent structuring information of ontologies will be used for structuring and classifying the patient records.

\textsuperscript{1}See www.health-e-child.org
We have addressed the following concrete problem. Let a database (such as a hospital information system or epidemiological research database) contain patient records with attributes, such as e.g. tumor site for cancer patients. We also have an ontology that captures the semantics of the range set of such attributes, e.g. the hierarchical meronomy of anatomical regions. Using this semantics, we visualize the induced hierarchical classification of patients.

Our proposed solution uses an OWL DL view of the patient database enhanced with external semantics allowing for the patient record classification by a reasoner, from where the inferred hierarchy is directly fed into an appropriate visualization tool.

This paper is organized as follows. Section 2 discusses related approaches and Section 3 the alignment of external semantics for clinical applications. In Section 4, we sketch the required steps for implementing two ontology-based visualization scenarios: a semantic facet browser and a semantic treemap visualization application. Both scenarios are introduced and discussed in Section 5, followed by their evaluation by clinicians in Section 6. We conclude with a summary, open issues and further research.

2. Related Work

For visualization, navigation, and analysis of complex and heterogeneous datasets, such as patient records, a certain structuring and annotation of data, present in the dataset, is desirable. A natural way of imposing a structure on the data is the establishment of a hierarchical order of cases, based on the attribute values or on available background knowledge. We distinguish two approaches to such hierarchy [7]: anonymous hierarchical clustering and so-called lexicographic hierarchies. In the former case, a similarity distance is defined, based on (some of) the attributes; and the dataset organizes itself into hierarchical clusters based on proximity. In contrast, lexicographic hierarchies are created by imposing an a priori hierarchical structure on the attribute values which in turn induces a hierarchy on the dataset. In [7], we discussed the visualization of inter-patient similarity and anonymous hierarchical clustering using a number of visualization techniques including treemaps, relative neighbourhood graphs, correlation plots and heatmaps. Based on sufficient datasets, anonymous hierarchical clustering can lead to the detection of previously unknown clusters. A certain disadvantage of such hierarchies, however, is that the meaning of the anonymous clusters is often difficult to capture. In this paper, our focus is on visualizing lexicographic hierarchies, where each “cluster” carries a label, reflecting the corresponding semantics. By selecting one or a small number of relevant attributes and by mapping these attributes to concepts of an appropriate ontology, the inherent structuring information of the ontology may determine the hierarchical structure of the patient dataset.

There has been much work on ontology visualization [1–4] that helps the user navigate ontological concepts. In contrast, the present work proposes not the navigation of the ontology directly, but rather to visualize instance data with the help of the knowledge that is represented by the ontology using reasoning with a medical ontology [5, 6].

The work presented in this paper relates visualization to knowledge discovery. The motivating example is that of the physician who believes that an extensive set of patient data would reveal subtle patterns if only it could be viewed the right way. The “right way” is here taken to mean analyzed in terms of established or even tentative concepts. The ontology carries what may have already been accepted or added as a research hypothesis, and the data is visualized on that basis, as a means of strengthening the evidence or as a means of bolstering the new hypothesis. There is a significant volume of work on visualization with (and without) ontologies. It may be differentiated from the work presented here on the basis of various criteria, such as: what is visualized, why and how. Visualization may be of data from a homogeneous database in order to display relative volume or from a heterogeneous source to expose some other feature or for exploratory data analysis [16]; it may be of the structure of the data (rather than the data itself) in order to reveal class-level relationships [3]. Visualization has been used to facilitate query formulation or to order threads of data in some schematic way (e.g. temporally); to display a data schema or to inform navigation through the data [17]. In particular cases, ontology-based visualization has been used to support queries based on temporal abstractions [18]; to enrich maps with additional geographic information [19]; to reveal multiple levels of abstraction in decision-tree generation [20] and to assist in information mining [21]; very popularly, to map social networks and communities of common interest [22], [23], [24]. Ontologies have also been used for knowledge discovery without visualization, especially in the integration of heterogeneous scientific repositories ([25]). For an up to date survey of ontology visualization methods reference must be made to [14], which provides a near exhaustive discussion of methods of classification, of visualization techniques (Euclidean, hyperbolic and spherical space, node-and-link, and other less obviously geometric methods), of representation, overview and focus methods, but it is distinctive in offering no discussion of content. There is a body of work dating back to the mid-90s on database visualization (see [16]) some of it devised to assist
with data mining tasks. Arguably the work closest to ours in spirit is [26], although their approach is designed more to manage the heterogeneity of web sources in order to summarize the information provided. Their visual method is based on node-and-link representation and resembles mind-mapping. On the other hand, [27] use semantic visualization to support the development process for database-oriented systems, and in particular phases of the process questions of data visualization must be tackled.

3. Incorporation of External Semantics

We have implemented two patient data visualization applications, demonstrating the benefits of incorporating external semantics. These have been demonstrated using the brain tumour patient data acquired within the Health-e-Child project and they provide experts in brain tumour diagnosis and treatment with an improved method for patient data visualization and comparison. The Health-e-Child database captures, for each patient record, more than 100 attributes of medical history and status data. Our hypothesis is that the using the semantics of some of the attributes, the patients can be better classified and, in turn, visualized. The incorporation of semantics is accomplished using inference; in particular we employ DL reasoning on an OWL DL view of the patient data aligned with an OWL DL medical ontology.

For a beneficial integration of external semantics, one has to decide which external knowledge sources are appropriate for the purpose in mind, i.e. which external knowledge source captures relevant and helpful knowledge for a particular context. In our scenario, we aim at classifying patients with respect to their diagnosed tumour location or WHO tumour classification. With regard to the attribute “tumour location”, we identified the Foundational Model of Anatomy (FMA) [8] covering the partitive hierarchy of brain regions as relevant and valuable medical knowledge. As the FMA is most comprehensive, covering approximately 70 thousand distinct anatomical concepts with more than 1.5 million relationships of 170 relationship types, we rely on fragments covering the concepts and relationships relevant to a particular visualization use-case. In our scenario, the established fragment encompasses all anatomical concepts describing possible brain tumour locations hierarchically structured by the regional-part-of relationship.

The WHO classification of Tumours of the Nervous System establishes a classification and grading of human tumours that is accepted and used worldwide [11]. Its constituted entities establish a hierarchical structuring of histological typed tumours covering a multiplicity of factors, such as the immunohisto-chemistry aspects, genetic profiles, epidemiology, clinical signs and symptoms, imaging, prognosis, and predictive factors. The WHO classification of tumours refers to the ICD-O (International Classification of diseases for oncology) code and includes a WHO-grading scheme that is used for predicting response to therapy and outcome. For improved patient data visualization, we revert to its hierarchical structuring and its grading scheme for hierarchically representing patients’ data.

4. The Steps towards improved Patient Data Visualization

We demonstrate how external semantics can be used for the hierarchical classification of patients based on attribute values. In our example the part-of semantics of brain regions induces a subsumption relationship on patient classes and our goal is to make this knowledge explicit available. The automatic classification of patient records is achieved by creating an OWL DL view of the patient database, aligned with the FMA partitive hierarchy fragment (in OWL DL) and by inferring the hierarchy using a reasoner.

In practice, we have the choice to use either A-Box or T-Box reasoning. A-Box reasoning looks for class membership, i.e. in our case an OWL DL view of the database would represent patients as instances. T-Box reasoning infers subsumption relationship; thus for enabling patient classification, patients are mapped to (singleton) classes. The pros and cons of either approach will be discussed elsewhere; in this paper our discussion is based on the demonstrator where we use class-based reasoning.

In the following, we illustrate all the steps we took for preparing the patient data to allow for improved visualization:

First, we identified the relevant ontology fragment, i.e. the fragments encompassing all concepts relevant to describing our particular visualization task, e.g. the visualization of brain tumour patient records. We make use of an FMA fragments covering the regional-part-of hierarchy of brain regions and the WHO classification of Tumours of the Nervous System as relevant medical background knowledge. We engineered our WHO ontology by hand based on the corresponding text document, and the anatomy semantics by creating a fragment of the FMA frames version and subsequently translating it to OWL DL.  

Second, we transformed the Health-e-Child patient data into OWL DL representation. Emanating from a flat view of the patient data covering a set of elected attributes, we create an OWL view with each patient being a class, each elected attribute corresponding to a

---

2 There are OWL versions of the FMA available which can be segmented; here we took the simpler route for the purpose of the demonstration.
role being represented as a restriction class axiom. Thus the information including the relevant attributes of the patient data gets transformed into OWL DL representation.

Third, we established a classification ontology determining the patient data classification within the reasoning process. The classification ontology consists of a set of defined classes (OWL classes described by necessary and sufficient constraints) capturing all patient attributes, such as tumour location, WHO-grade, or WHO-classification, that are governing the classification process. According to the semantics of each patient class, it will be classified within the reasoning process.

Fourth, we integrated all created ontologies, i.e. the medical ontology fragments, the patient ontology, and the classification ontology and executed the reasoning process on top of the integrated ontological model. The reasoning process enables the classification of patient classes by integrating knowledge captured by the external medical ontologies. The resulting ontological model explicitly captures the inferred hierarchical classification of patient records.

Fifth, we deployed the inferred ontological model into the visualization software by transforming the OWL DL representation into the format conformant to the API of the visualization software.

5. Ontology-driven Visualization Scenarios

5.1. Semantic Facet Browsing

Ontology-based facet browsing establishes browsing facilities over the set of all patient records. It provides means for identifying similar patients’ records with regard to selected and relevant patient attributes. Due to the incorporated ontological structures and the help of OWL DL reasoner, such as Pellet or Racer, the variety of data regarding detail and precision can be reflected. For instance, the clinical protocol of each brain tumour patient covers the attributes “tumour location” and “WHO tumour classification”. By integrating the partitive hierarchy of brain regions into the facet browser, clinicians can identify all patients with similar brain tumour locations, whereby the similarity measure reflects different levels of detail in patient data description. For instance, the search for patients similar to patients with tumour location “Left Cerebellum” will yield all patients with tumour location specified at higher granularity, i.e. all patients with tumour location specified by the concept “Left Cerebellum” or by any of its parts, such as “Temporal Left Cerebellum”.

5.2. Semantic Treemap Visualization

Treemaps [15] are an efficient two-dimensional technique for visualizing hierarchical data structures. It is particularly popular for disc storage view of hierarchical filesystems because file size is an aggregate attribute of files. It is a space filling technique, i.e. one that uses the entire screen area by dividing it up between leaf nodes which are subsequently grouped into enclosing rectangles [14]. The image is effectively a rectangular Venn-diagram of nonintersecting sets. Besides the set semantics, there are other attributes such as colour, choice of font and label which can represent attributes of the data beyond that of the hierarchy.

To allow for improved patient data visualization, we represent patients as rectangles of equal size such that the cardinality of patient classes can be easily seen. Figure 2 shows the user interface based on the patient taxonomy that has been inferred from the anatomical meronomy. The ontological structuring information, in our example the hierarchical structured
brain tumour locations establishes enclosing and nested areas labelled by regions of the brain. Colour may be used to visualize further attributes, in our example the status at the end of treatment for each patient. It is meant to show that the ontology-based space-filling hierarchies can be useful for visualizing correlation, in this example one finds that patients with tumour in the Hindbrain tend to have better prognosis than in other areas (especially the Left Temporal Lobe).

Fig. 2: A treemap view of the ontology assisted data representation

6. Clinical Evaluation

We presented and discussed the two ontology-based visualization scenarios with our clinical project partners, the I.R.C.C.S. Giannina Gaslini in Genoa, the University College London, Great Ormond Street Children’s Hospital in London, and the Assistance Publique Hopitaux de Paris. The response towards both visualization applications was very positive. It is important to note that the system does not give the clinician new information that she would not know herself, yet the doctors acknowledged that, in many scenarios, the advanced visual organization is very helpful in navigating the information: “out of sight, out of mind”.

The clinicians emphasized the relevance of inferred knowledge for visualization purposes: although medical experts might be aware of a patient’s WHO-Grade by reading the patient record, the inferred and explicitly represented knowledge about a patient’s WHO-Grade, as it is implemented by the facet browser, helps the clinicians to find patients along this axis. Moreover, the clinicians appreciated to easily locate (browse for) previous patients that are similar with regard to some selected classification axis (facets) for subsequently accessing their treatment process and progress to benefit from past, but context relevant medical decisions and experiences.

The treemap approach was acknowledged by the clinicians as basis for analyzing and interrelating multiple characteristics of patients. In their daily routine, clinician researchers might come up with hypotheses that can be now checked by using the treemap approach allowing for easy visualization of correlation between relevant clinical attributes. Moreover, the clinicians noted that the visualization of patient data correlation could be helpful in education and training situations.

7. Conclusion and Future Steps

The clearly arranged presentation of similar patients with respect to the complex and heterogeneous patient data becomes a key requirement for medical clinical decision support systems. Our approach aims for improved patient record visualization by the alignment of external semantics.

We have presented the steps required for realizing improved patient data visualization by means of incorporated external semantics, as well as described two realized applications, i.e. semantic facet browsing and semantic treemap visualization, along with their clinical evaluation. Both applications were accomplished using the OWL DL description language and class-based reasoning.

Within our future work, we will focus on the comparison of T-Box versus A-Box reasoning with regard to scalability and performance.

Acknowledgments

This work has been partially funded by the EU project Health-e-Child (IST 2004-027749). The authors wish to acknowledge support provided by all the members of the Health-e-Child consortium in the preparation of this paper. We are especially thankful to our clinical partners Anwar Baban, Maria-Luisa Garre, Giacomo Pongiglione and Monica Marini. We are grateful to Bernd Lindner for implementing the facet browser demonstrator.

10. References


