Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2016) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 6th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2016)

Capturing, Annotating and Processing Practical Knowledge by using Decision Trees

Tobias Weller[a,*], Maria Maleshkova[a]

[a] *Institute of Applied Informatics and Formal Description Methods (KIT), Karlsruhe, Germany*

Abstract

Practical knowledge describes the knowledge resulted from experience of a person. Capturing and processing practical knowledge is usually difficult, because it is only available in a persons head. However, this knowledge would help inexperienced persons to obtain practical knowledge in a fast way. Approaches so far have only considered how already formalized knowledge can be shared and integrated across different systems. However, capturing and formalizing practical knowledge and working around with incomplete data have not yet been considered. To address this problem we 1.) developed an open-source extension for Semantic MediaWiki that supports the graphical modeling of practical knowledge; 2.) enable to enrich the formalized practical knowledge with semantics from ontologies and knowledge graphs with references to external data sources and rules and 3.) present a technical infrastructure to automatically execute the created decision trees and thus to retrieve recommendations of the practical knowledge in real-time.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Decision Support Systems in Healthcare, Information and Knowledge Processing, Knowledge Management and Representation, Semantic MediaWiki

1. Introduction

Modeling knowledge is an important aspect for processing, sharing and reusing knowledge. We distinguish between factual and practical knowledge [1]. Factual knowledge describes knowledge which has been written down into sources like educational books, reports in educational journals or studies. The knowledge, written down in these sources, allow people to access and gather the knowledge, become more educated and apply the knowledge into their familiar context.

Besides factual knowledge, practical knowledge results from experience, gathered by performing tasks. It describes insights that have independently been gathered by repeated performance of specific tasks. Usually, these insights consists of patterns that have been recognized during the performance or dependencies between different factors that influence each other.

* Corresponding author. Tel.: +49 721 60845770 ; fax: +49 (721) 608 46690
E-mail address: tobias.weller@kit.edu

1877-0509
Peer-review under responsibility of the Conference Program Chairs.

However, while factual knowledge is fair available (e.g. for hepatocellular carcinoma (HCC), there is an estimated amount of 3,000 contributions per year[2]), practical knowledge is usually hard to capture and share in a structured way. The problem is that practical knowledge is usually available in a persons head. It is hard for a person to write practical knowledge into quotable sources and formalize it in a structured way. Hence, the lack of expressiveness of persons hamper formalizing practical knowledge.

Although there are solutions in the Semantic Web domain that allows to formalize knowledge into rules like e.g. the recommended treatment plan of a patient, based on his factors, these rules are hard for domain experts to formalize. For domain experts, writing these rules, for instance in N3[1] or in SWRL[2], is only possible at great expense. However, the advantage of using such rule formats is that there are already engines available that allows to infer new knowledge by applying these rules on data, like e.g. on patient's data.

Another aspect that has to be considered is that information, needed to infer new knowledge by applying rules, might be incomplete. Therefore, new knowledge cannot be inferred. However, for supporting physicians in their decision-making process, it might also be useful to reduce the number of possible recommendations, based on the available data. E.g. an explicit treatment plan cannot be inferred, due to missing patient's data, however, several treatment options can be excluded due to specific available information about the patient.

We want to tackle the raised problems by presenting a Decision Support System in Healthcare that allows to model knowledge in decision trees, annotating the decision trees with rules and executing the modeled decision trees on patient's data. This Decision Support System will help physicians in their decision making process. We will demonstrate the applicability of our system by applying it in the medical domain, however, the presented system is not restricted to a specific domain. Hence, it can be used in all available domains to support people in entering, enriching with semantics, querying and processing of practical knowledge.

As a result, through the system, practical knowledge can be 1) captured and stored in order to be able to pass it on and share it with other people 2) visualized to simplify the comprehension of a decision 3) enriched with semantic information from ontologies and knowledge graphs to analyze the decision and its origin 4) applied on data from knowledge graphs (e.g. patient's data) in order to suggest the result of the execution of the modeled knowledge.

We demonstrate the applicability of our solution by modeling two concrete practical knowledge in form of decision trees in the medical domain. One decision tree is from the field of model-guided therapy for hepatocellular carcinoma (published by Berliner et al.[3]) and another decision tree is from the field of mitral valve reconstruction (published by Fedak et al.[4]). The used methods and an overview of the system, addressing the following research questions are described in section 3:

- How can we capture knowledge used in a decision-making process?
- How can we implement the infrastructure necessary for storing, accessing and processing knowledge?

Section 2 motivates the current scenario. Afterwards, we will show our approach to capture, annotate and process the practical knowledge (Section 3). The implementation of our approach is described in section 4. The evaluation of the system includes showing the average runtime of applying practical knowledge on data, stored in a knowledge base and the percentage of reduced recommendations, even if data was missing. Related Work about Decision Support Systems and applying Semantic Web Technologies is given in section 6. Finally, in section 7, we will conclude the paper.

2. Motivation

As mentioned, we distinguish in factual and practical knowledge. Currently, it is very hard to capture and pass on practical knowledge. By enabling domain experts to model knowledge and execute it on their data, we can directly pickup knowledge from domain experts in a very easy way, which is available for them and the Decision Support System at any time. Figure 1 shows our vision of the Decision Support System.

[1] http://www.w3.org/DesignIssues/Notation3
[2] http://www.w3.org/Submission/SWRL/

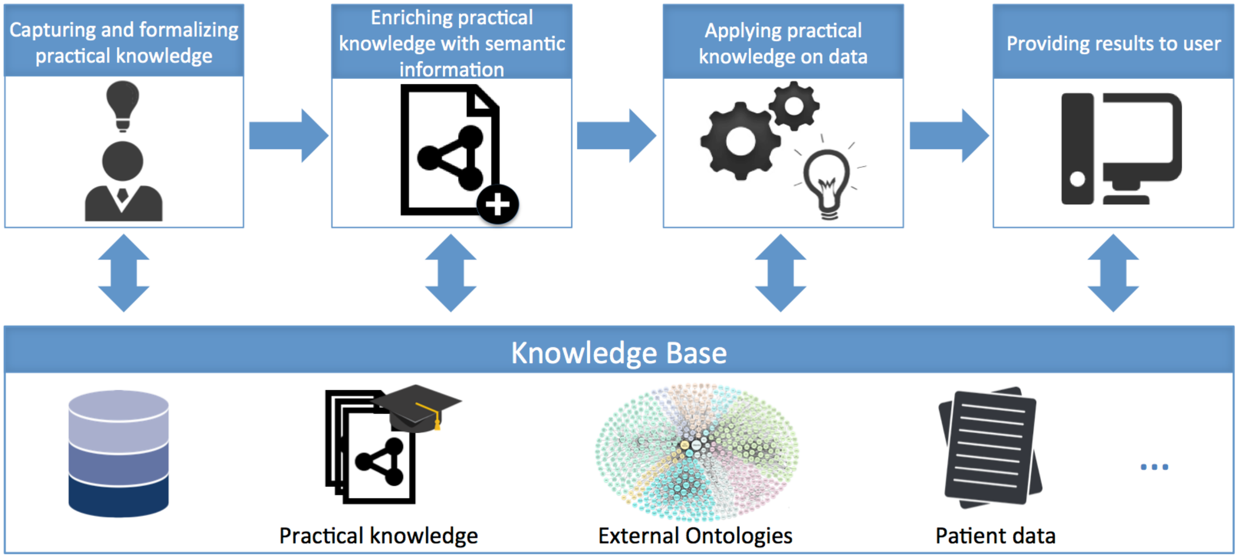

Fig. 1. Vision of our system to capture practical knowledge. The knowledge base consists of own entered data, as well as linked data to external sources. The system comprises four stages. 1) Capturing and formalizing practical knowledge: Practical knowledge is collected and converted into a machine-readable format and stored in the knowledge base. 2) Annotating with semantic information: The captured knowledge can be annotated with semantic information. 3) Applying knowledge on data: The captured and with semantic information enriched knowledge can be applied on data, stored in the knowledge base. 4) Providing results to user: The results of the application of knowledge on data will be presented to the user in order to provide recommendations based on the inferred knowledge.

In our opinion, it is very important to make knowledge, especially the practical knowledge, available to a vast amount of people and to enable sharing knowledge and using it in a Decision Support System. Inexperienced people can benefit from the provided practical knowledge by understanding and comprehending the provided practical knowledge from other people. Thereby, these formal inexperienced people gain additional knowledge, resulted from experience by other people, and become more educated. Thus, the learning curve for people, gained by providing practical knowledge, increases. We expect that these people can apply and link the provided knowledge with own experience and factual knowledge, written in quotable sources, and thus learn faster. Therefore, they get more educated in a faster way and gain their own practical experience, resulted from experience, which can be formalized in turn and provided to other people. This could lead to a quadratic increasing of the learning curve of people and can be used in the Decision Support System.

As mentioned above, sharing practical knowledge is an important aspect to pass on experience to successor and other people that might benefit from the gained practical knowledge. An approach to capture and ensure the formalizing of practical knowledge from people is preferable. One aspect to model knowledge and the corresponding decision of a rule is to model the knowledge in decision trees. This form of representing knowledge is very intuitive and comprehensible to people.

However, a crucial aspect in supporting physicians via a Decision Support System, is to handle incomplete data. Not to provide any recommendation to a physician by the Decision Support System, we want at least to reduce the number of possible outputs and provide these recommendation with a remark of incomplete execution, so a physician is aware of the fact that due to missing data, an explicit decision could not be made. Our method to tackle this problem is presented in section 3.

The presented problems of capturing and using practical knowledge, as well as handling incomplete data during applying the captured knowledge on data, is occur in many domains. E.g. in the medical domain, the barcelona criteria[3] provides a recommendation of how a patient should be treated according to his factors. However, as mentioned, the patient data might be incomplete, so an explicit decision cannot be made by the Decision Support System. Similar circumstances might occur also in other domains, like e.g. in the manufacturing industry, where decisions of the further processing of produced materials cannot be made due to missing data.

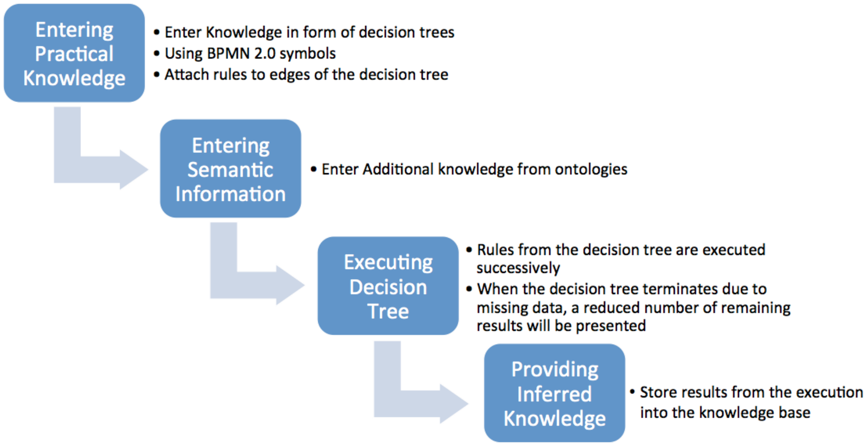

Fig. 2. Procedure of entering and executing knowledge in form of decision trees from a user's perspective.

3. Proposed Methods

An important aspect for capturing practical knowledge is to find a suitable representation that allows to be comprehensible for humans but also accessible for machines in order to apply them on data. We decided to model knowledge by using decision trees. Decision trees can be modeled by using flowchart symbols. The root indicates the start of the decision tree. Each node represents a decision point. The edges contain the rules that specify the direction of the further processing. For applying decision trees on data, one does start at the root of the decision tree and choose the path according to the rules on the flows. The decision tree is traversed until a decision cannot be made or the leaf is reached. A leaf represents the recommendation, which should be applied according to the performed decision tree.

This way of representation and processing is very comprehensible for humans. In addition, due to the few elements, needed to model decision trees, it is suitable that decision trees can easily be created by domain experts. Furthermore, decision trees can easily be adapted according to new insights. Flows can be rearranged to latest experience and thus allow to bring practical knowledge up to date.

The formalization of the entered practical knowledge should be done automatically without interventions by the user. We will keep the user interface as simple as possible, so the user can concentrate on entering his practical knowledge. Once we formalized the practical knowledge, it is available for querying and processing. Additional annotations to the practical knowledge help users to understand and filter for certain practical knowledge.

All available information, like e.g. patient data and the captured practical knowledge and its annotations, are stored in a common knowledge base. Thus, the knowledge base serves as a central data hub for all relevant data.

Figure 2 illustrates the procedure of the system from a user's perspective. It encapsulate our vision of the Decision Support System from figure 1. In the following, we will present the procedure and used methods in more detail.

1. Step – Entering Practical Knowledge: The provided tools, in order to capture practical knowledge, must be easy usable, hence also domain experts, not familiar with semantics, can use the tools to capture their practical knowledge. We will use symbols, provided by BPMN 2.0 [3], to model decision trees. BPMN is proposed as a standard by the Object Management Group (OMG) in 2008. The current available version of BPMN is 2.0.2, published in ISO/IEC 19510 [5]. We will attach the corresponding rules on the edges of the decision tree, which specify the direction of processing. We will use N3 rules[4] to formalize the decision-making process. BPMN is a graphical modeling language. Thus, users should be able to enter the decision trees, which represent their practical knowledge, via a graphical user interface. The decision trees will be stored in the knowledge base.

[3] http://www.omg.org/spec/BPMN/2.0/
[4] http://www.w3.org/DesignIssues/Notation3

2. Step: Entering Semantic Information: In order to exploit the semantics, we have to annotate the entered practical knowledge with semantic information. We would like to allow users to enrich practical knowledge with semantic information for instance from the Foundational Model of Anatomy (FMA)[5], which is an ontology that describes structures and relationships of a human's body. However, also references to unstructured information like text documents should be possible. These information can be used to 1) enhance processing of practical knowledge and 2) comprehend practical knowledge and get more detailed information, as well as allowing to filter for certain practical knowledge.

3. Step: Executing Decision Tree: Usually, decision trees cannot be executed when data for processing the tree is missing. However, we created an engine that process the nodes successively. The advantage of using decision trees, having attached rules on the edges and processing the data successively, is that whenever the tree terminates, because of missing data, we know the node in the tree and can therefore return, according to the current node in the tree, the remaining possible results from the leafs that are reachable from this node. The reduced outcome is marked to show physicians the uncertain result of the decision tree. Thus, we can at least reduce the number of possible outcomes of the decision tree. We believe that this approach is very useful in a Decision Support System and better than having no outcome.

4. Step: Providing Inferred Knowledge: The results of the executed decision tree on data is stored in the knowledge base and is thus available for further processing and querying by the user. If the decision tree represents the practical knowledge of another person, he has the result of what experienced people would do in similar circumstances. Users can comprehend the results by looking up the formalized knowledge and reproduce the results of the application on the data.

4. Implementation

The medical domain is very suitable for our approach, because in this domain exists a lot of experience by physicians that can be captured with our approach. Inexperienced physicians such as assistant doctors, can benefit from the entered practical knowledge in order to increase their knowledge rapidly. Figure 3 illustrates our infrastructure to capture and annotate practical knowledge with semantic information, applying the practical knowledge on data, stored in our knowledge base, and provide the results to users.

The infrastructure is a three tier architecture to store, query and process the data. The basis of our architecture is a Semantic MediaWiki[6]. Semantic MediaWiki is a powerful collaborative knowledge management system to store and query data. Data resources, concepts and properties can be annotated internally, as well as linked to external data sources like e.g., DBpedia[6]. The possibility of linking concepts and properties enables the integration of well-known ontologies such as Dublin Core[7] and SNOMED[8]. The information of the Semantic MediaWiki is stored in an Open Virtuoso database.

We model decision trees using BPMN as modeling language. We used Cognitive Process Designer[7,8] to model the knowledge in form of decision trees. Cognitive Process Designer is an extension to Semantic MediaWiki that allows to model BPMN networks and checks the syntax. Each BPMN flow and connecting object is represented by its own wiki page.

Besides modeling practical knowledge, we enable users to link and provide additional semantic information. Each flow and connecting object can be enriched with further semantic information. To this end, different semantic forms, entered into Semantic MediaWiki, are available to annotate BPMN flows and connecting objects. The forms are opened in a front layer, before the graphical user interface. Afterwards, users can enter additional semantic information such as references to external data sources, comments, labels and uploading rule files in N3. This allows to attach rules to the edges of the decision trees that describe the direction for processing the data.

For executing the modeled decision trees (practical knowledge), we provide a Web API. The execution starts at the root element of the decision tree and follows its path until it reaches a decision node. In this case it checks all rules,

[5] http://sig.biostr.washington.edu/projects/fm/index.html
[6] http://dbpedia.org
[7] http://dublincore.org
[8] http://www.ihtsdo.org/snomed-ct

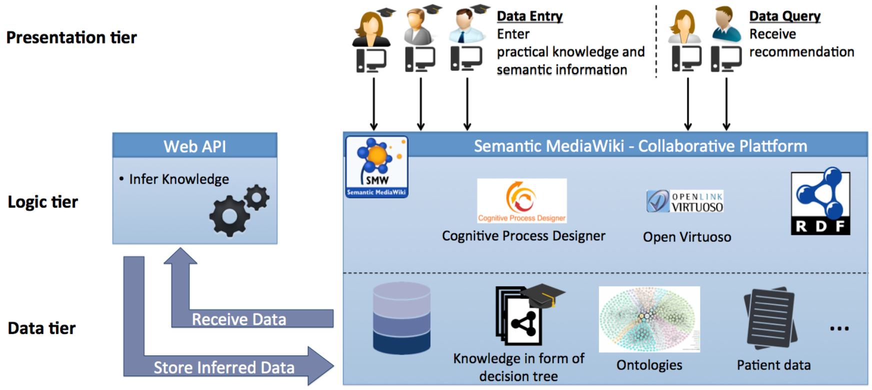

Fig. 3. Infrastructure to capture, annotate and applying practical knowledge on data, stored in our knowledge base and provide the results to users.

according to their applicability of the outgoing links of the actual element, and follows those links that are applicable. The procedure is repeated by following the corresponding paths until reaching the end of the decision tree (BPMN End Event) or a node from which no path can be followed due to missing data. The end event contains the outcome of the decision tree, stored in Semantic MediaWiki.

Our deterministic algorithm supports incomplete patient data. If there is patient data missing, the algorithm does not abort. Instead, it returns the smallest possible subset of outcomes, based on the reached point of the decision tree. Therefore, our approach is very robust in minimizing the appropriate number of possible outcomes. The Web API uses cwm[9] as reasoner to infer knowledge. The inferred knowledge will be stored to the Semantic MediaWiki database.

5. Experiments

In order to evaluate our approach, we chose two existing decision trees that represent practical knowledge. Figure 4 shows the two chosen decision trees. We chose the Barcelona criteria, published by Berliner et al.[3], that recommends a therapy plan for patients that suffer for hepatocellular carcinoma and a decision tree from Fedak et al.[4] that recommends annuloplasty rings, which is an artificial annuloplasty ring for treating mitral regurgitation.

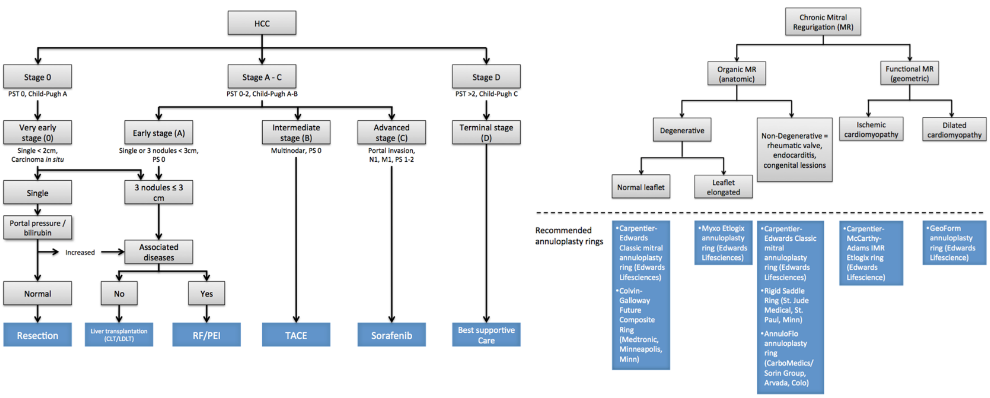

Fig. 4. Two decision trees that is used for evaluating the system. The left decision tree shows the recommendations for the therapy plan for patients suffering upon heptocellular carcinoma according to Berliner et al.[3]. The right decision tree shows Fedak et al.[4] recommendations for annuloplasty rings. The blue boxes below outline the recommendations according to a patient's factors.

[9] `http://www.w3.org/2000/10/swap/doc/cwm.html`

We captured both decision trees by using Cognitive Process Designer[7,8] and annotated the decision trees with references to PubMed[10] and rules in N3 format.

In order to apply the decision trees on data, we generated 1,000 random patient data. The patient data include factors, needed to pass through the decision trees. However, because we generated the data randomly, it may occur that the decision trees may not pass trough completely because decisions cannot be made due to fact that conditions may not be true. However, the API returns the smallest possible subset of outcomes that is possible.

We ran the experiments on a local machine with 2.9 GHz Intel Core i5 processor and 16 GB 1867 MHz DDR3 memory. For the annuloplasty decision tree by Fedak et al.[4], the decision tree could always terminate till the leaves. The average runtime for receiving the patient's data was 0.5054 seconds. The average runtime for executing the annuloplasty decision tree was 1.5067 seconds. The inferred knowledge was stored back into the knowledge base. The average runtime of the Barcelona criteria[3] was 1.508 seconds. However, this decision tree did not always terminate until its leaves but could reduce the number of possible treatment options. In 526 cases, the decision tree could not infer a unique treatment option but reduce the number of possible treatment options to the half. In average, we could reduce the number of possible treatment options about 50.1873 %.

6. Related Work

Current Decision Support Systems often use Machine Learning Methods to support people in their decision-making process[9,10]. These methods need historic data in order to train the models. Algorithms exist to create decision trees out of the available attributes, entered by the user[11]. However, entering is complicated by having a huge amount of different decision attributes. In addition, the decision trees are created automatically and cannot be influenced by the user. Decision trees need all data in order to terminate properly, however, one can retrieve a reduced set of results by knowing the node, for which no decision could be made. Besides classical Machine Learning methods, there are also approaches that incorporates practical knowledge into machine learning models[12]. Thereby, Naïve Bayes methods[13] are used but physicians select the initial variables that will be used from the available data.

Besides these methods, which provides the decision, previous works have mainly focused on capturing knowledge that was generated during project work. Thereby, influences in capturing and sharing knowledge were considered[14]. However, solutions such as broadcasting emails, regular meetings between project groups and mentoring as knowledge management strategies were given, but none which allows to model knowledge in a structured way. In order to capture knowledge, techniques like Computer Supported Cooperative Work (CSCW) are used to support cooperative communication, activities and coordination[15]. However, proposed techniques index the knowledge, generated in meetings, communications, and decisions, to enhance the retrieval of documents but do not capture experience from persons in a structured way. Therefore, the knowledge is still unstructured but the retrieval is easier due to indexing.

Platforms that integrates large-scale reasoning services have been published[16]. They deduce further knowledge by applying rules on data. Thereby, the expressiveness of the rules for inferring new knowledge has to be considered[17].

For annotating decision trees, we used existing ontologies and vocabularies. Published conventions, such as Dublin Core Schema[11] and FOAF[12], provide a set of metadata that can be used to annotate resources. The advantage is that such conventions do not restrict on specific structures. They can easily be integrated and reused into existing ontologies in order to annotate resources. There are also some ontology-based annotations available for process models[18,19], which might also be used to annotate practical knowledge in form of decision trees. The process models are semi-automatically annotated.

7. Conclusions

We modeled practical knowledge by using decision trees and used them in a Decision Support System in Healthcare. Decision trees are suitable to structure practical knowledge due to their transparency and the fact that they can

[10] `http://www.ncbi.nlm.nih.gov/pubmed/`
[11] `http://dublincore.org`
[12] `http://www.foaf-project.org`

be interpreted very easily. Annotations, such as references to external sources and meta-information of the modeled knowledge allow advanced processing of the captured practical knowledge and the reutilization of existing knowledge from various sources.

As rule engine, we used cwm[13], a semantic reasoning engine. This reasoner, integrated in a web service, provides the functionality to infer further knowledge and deciding the pathway on the tree. The execution of the tree is very robust, because it does not abort if specific factors, necessary to decide the path, are missing, but returns a minimized set of outcomes of the decision tree. A validation study performed with 1,000 patients, using the Barcelona Criteria [3] and the Fedak et al. [4] decision tree, indicates that our concept is applicable and paves the way for capturing, annotating and executing practical knowledge.

Future work comprises among others the simplified creation of semantic rules like e.g. N3. So far, rules have to be created manually and then linked to the modeled practical knowledge. However, domain experts, which are not familiar with such rule languages, cannot create these rules. Therefore, we plan to simplify the creation of these rules and automatically link them to the modeled practical knowledge.

In conclusion, we have taken a first step towards capturing, annotating and processing practical knowledge that can be shared and used in a Decision Support System to assist inexperienced people.

References

1. März, K., Hafezi, M., Weller, T., Nolden, T., Zelzer, S., Maleshkova, M., Wagner, M., Engelhardt, S., Kenngott, H., Rettinger, A., Studer, R., Mehrabi, A., Maier-Hein, L.: Towards Knowledge-based Liver Surgery: Holistic information processing for surgical decision support. International Journal of Computer Assisted Radiology and Surgery 2015 (April); 10(6).
2. Gores, G.J. Hepatology: A Home for Hepatocellular Cancer Publications, Hepatology. July 2009; 50(1): 12.
3. Berliner, L., Lemke, H.U., van Sonnenberg, E., Ashamalla, H., Mattes, M.D., Dosik, D., Hazin, H., Shah, S., Mohanty, S., Verma, S., Esposito, G., Bargellini, I., Battaglia, V., Caramella, D., Bartolozzi, C., Morrison, P.: Model-guided therapy for hepatocellular carcinoma: a role for information technology in predictive, preventive and personalized medicine. The EPMA Journal 2014 5:16.
4. Fedak, P.W.M., McCarthy, P.M., Bonow, R.O.: Contemporary Reviews in Cardiovascular Medicine: Evolving Concepts and Technologies in Mitral Valve Repair. Circulation 2008;117:963-974
5. Information technology – Object Management Group Business Process Model and Notation (ISO/IEC 19510:2013), July 2013
6. Krötzsch, M.,Vrandečić, D., Völkel, M. Semantic MediaWiki, The Semantic Web - ISWC 2006: 5th International Semantic Web Conference, ISWC 2006, vol. 4273, pp. 935-942, Athens, November 2006
7. Weller, T., Maleshkova, M. Towards a Collaborative Process Platform: Publishing Processes according to the Linked Data Principles, Proceedings www, pages: 8, ACM Order Department, Montreal, Canada, April, 2016
8. Weller, T., Maleshkova, M. Cognitive Process Designer – An Open-Source Tool to Capture Processes according to the Linked Data Principles, The Semantic Web: ESWC 2016 Satellite Events, Springer
9. Alickovic, E., Subasi, A. Medical Decision Support System for Diagnosis of Heart Arrhythmia using DWT and Random Forests Classifier, Journal of Medical Systems, 40(4), pp. 1–12
10. Kaya, Y., Uyar, M. A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease, Applied Soft Computing, Volume 13, Issue 8, August 2013, Pages 3429-3438, ISSN 1568-4946, http://dx.doi.org/10.1016/j.asoc.2013.03.008.
11. Vlahovic, N. Discovering Tacit Knowledge in Business Decision Making. WSEAS Transactions on Business and Economics, Vol.4 (5), 2008, pp.64-73.
12. Kuusisto, F., Dutra, I., Elezaby, M., Mendona, E. A., Shavlik, J., & Burnside, E. S. (2015). Leveraging Expert Knowledge to Improve Machine-Learned Decision Support Systems. AMIA Summits on Translational Science Proceedings, 2015, pp. 8791.
13. Wang, Xiao-Hui, et al. Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. International journal of medical informatics, 1999, 54. Jg., Nr. 2, pp. 115-126.
14. Polyaninova, T.: Knowledge Management in a Project Environment: Organisational CT and Project Influences. Vine,Vol. 41(3), 2011.
15. Matta, M., Dai, X., Rauscher, F., Atifi, H., Ducellier, G. How to Capture Knowledge from Project Environment? IFIP WG 5.7 International Conference, APMS 2015, Tokyo, September 2015
16. Assel, M., Cheptsov, A., Gallizo, G., Celino, I., Dell'Aglio, D., BradeÅko, L., Witbrock, M., Della Valle, E. Large Knowledge Collider - a Service-Oriented Platform for Large-Scale Semantic Reasoning, Proceedings of the International Conference on Web Intelligence, Mining and Semantics (WIMS '11), New York, doi: 10.1145/1988688.1988737
17. Heymans, S., Vermeir, D. Integrating Semantic Web Reasoning and Answer Set Programming, Proceedings of the 2nd Intl. ASP'03 Workshop, Italy, September 2003
18. Lin, Y., Ding, H. Ontology-based Semantic Annotation for Semantic Interoperability of Process Models. Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, Vol. 1, pp.162-167, November 2005, doi: 10.1109/CIMCA.2005.1631259
19. Lin, Y., Strasunskas, D. Ontology-based Semantic Annotation of Process Templates for Reuse, 10th Int. Workshop on Exploring Modeling Methods in System Analysis and Design (EMMSAD05), Porto, Portugal, 2005

[13] `http://www.w3.org/2000/10/swap/doc/cwm.html`