Textbook Knowledge on Hospital Information Management as Linked Data

Konrad Höffner [a,*], Franziska Jahn [a], Birgit Schneider [a], Thomas Pause [a], Alfred Winter [a]
[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany*
E-mails: konrad.hoeffner@imise.uni-leipzig.de, franziska.jahn@imise.uni-leipzig.de,
birgit.schneider@imise.uni-leipzig.de, thomas.pause@imise.uni-leipzig.de, alfred.winter@imise.uni-leipzig.de

Abstract. Textbooks contain abstract knowledge about a domain. In the case of Hospital Information Management, much of this knowledge consists of information management functions, roles executing these functions and the information used or updated by these functions. We present a common data model for the domain and the result of applying the data model to three textbooks an interview and a standard. We publish the result over several interfaces that are useful for researchers, practitioners and students, depending on their objectives and their Semantic Web skills.

Keywords: information management, information systems, hospital information management

Table 1

x	y

1. Introduction

Given a care delivery organization, its health information system (HIS) can be defined as the part of the organization that processes and stores data, information, and knowledge. It usually consists of different application systems, computers, and network components [1]. Managing an HIS comprises planning, monitoring and directing activities. Due to the complexity and the unique conditions in health care, HIS management is an exceptionally challenging task. There is an enormous amount of frameworks, textbooks and articles describing the scope of HIS management from the perspective of medical informatics. A structured representation of the different perspectives leads to a holistic view on HIS management and helps help researchers and students connect their existing knowledge with further knowledge from other sources during research and learning. In order to integrate different knowledge sources and to provide the knowledge in a structured, machine-readable data format, we extracted knowledge about HIS management from three textbooks [1–3] and other sources. The combination of this knowledge [4**?** , 5] results in the SNIK ontology, the Semantic Network of Information Management in Hospitals ("Krankenhaus" in German).

In order to encourage and enable other researchers, students and health informatics professionals to use available knowledge of HIS management, We introduce different user interfaces to the SNIK ontology and discuss their suitability for different use cases of the target audiences. We conclude with plans for future work on interlinking and visualization.

Medical informatics students, who are trained for executive positions in information management departments of healthcare institutions, such as hospitals, need a clear terminology of their domain. Due to different frameworks and textbooks dealing with information management in healthcare, modelling the knowledge unravels the links between the different views on information management. These are only implicitly known or not known at all by experts in the field.

*Corresponding author. E-mail: konrad.hoeffner@imise.uni-leipzig.de.

We present a common data model for the domain and the result of applying the data model to three textbooks. We publish the result over several interfaces that are useful for researchers, practitioners and students, depending on their objectives and their Semantic Web skills.

Publishing textbook knowledge as Linked Data enables different ways of teaching. The data is the result of The extraction effort includes The data is continously revised: users constantly report wrong or missing data in the visualization, which is then corrected, respectively revised, by the researchers of the project.

This paper describes the current state of digitalized Hospital Information Management textbook knowledge.

Using a data model of Information Management Depending on the learning style and information need of the student, several tools are available to consume that knowledge.

FJ: Welchen Zweck hat dieser Teil in der Einleitung? Auch würde ich die Ziele vielleicht etwas kompakter als Liste beschreiben.

2. Data Model

In order to specify, which information should be extracted from the books and to facilitate comparisons, we use a common data model. Because processed textbooks contain abstract knowledge instead of information about any specific hospital, all concepts are modelled as classes. We thus call our data model the “meta model”, as it is an ontology.

The meta model (see fig. 1) provides a common vocabulary for the domain of HIS management and thus defines, which superclasses and property can be used. SNIK version 0.4.1 comprises five subontologies that are built upon the meta model, see Table 1. At the head of the class hierarchy is the “Top” class, which has exactly three disjunctive subclasses. Following the meta model, each class has to be a subclass of exactly one of them. The correct superclass of a new concept can be found by answering the question: Who (“Role”) does what (“Function”) and which information (“EntityType”) is needed? If a concept is neither of them, it cannot be modeled using the meta model. As the subclass relation is transitive, a new class can be placed further down the hierarchy and it can still be inferred, whether it is a Function, Role or EntityType (see Figure 3). Besides the subclass relationship, two classes can be connected with relations provided by the meta model. The generic “is associated with” relationship carries little information. For example, a role and a function can be connected as “is involved in” “is responsible for” and “approves”. Relations that are neither of them can either be modeled by using the generic “is associated with” relation or by creating and using a new sub relation of “is associated with”

3. Concessions to Practicality

SNIK is mainly used by students and researchers having no or little semantic web knowledge. This led to several concessions to usability and actual usage by sacrificing full conformance to the W3C OWL recommendation in the following cases:

OWL Restrictions The meta model offers several properties between the different subtopclasses. The definitions of `rdfs:domain` and `rdfs:range` state that the subjects, respectively objects of all triples with that property are instances of that class. The classes of SNIK are subclasses, not instances, of the subtopclasses. This means that using those properties directly in SNIK is technically wrong and OWL property restrictions should be used. This presented the following problems:

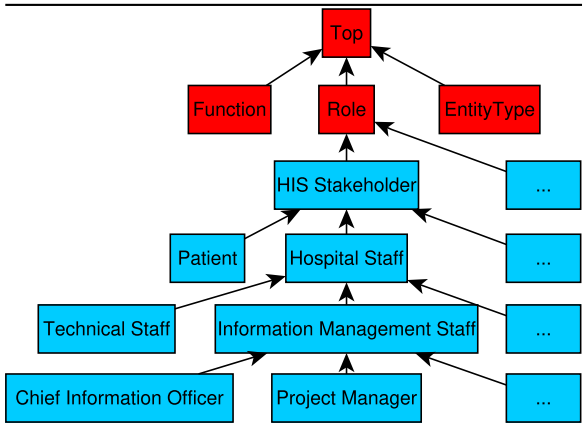
- Tools such as LodLive don’t understand those OWL constructs and only show the RDF form with blank nodes and helper constructs, which is confusing to the users
- Communicating OWL restrictions to extractors and non-Semantic Web expert project members is difficult.
- Extractors only enter the name of the property, not whether it is a "someValuesFrom" or "allValuesFrom" restriction. For example, it is unclear whether each CEO is involved in at least one project review, or whether each CEO is only involved in project reviews. It may even be the case that both of those statements are not what the author intended.
- Deleting or adding a connection between classes in the OntoWiki is no longer an atomic operation and may result in unreferenced blank nodes or partially restored undo results.

To avoid those problems, we use direct triples between the classes of SNIK as if they were instances, not classes.

FJ: Sollte man hier evtl. noch erklären, wie es denn tatsächlich gemeint ist? Eigentlich haben die in SNIK modellierten Aussagen jeweils eine Bedeutung wie: Laut Lehrbuch xy muss/sollte ein CEO im Kranken-

4. Unsorted

URL	http://www.snik.eu/ontology
Version	2018-10-10, 0.4.1
License	CC BY-NC-SA 4.0
SPARQL Endpoint	http://www.snik.eu/sparql
Visualization	http://www.snik.eu/graph
RDF Browser	http://www.snik.eu/ontology
Download	https://github.com/IMISE/snik-ontology/releases/download/0.8.0/snik-0.8.zip
# Triples	112 747
# Classes	4729
# Properties	329
# Interlinks	713



Ontology	Source
http://www.snik.eu/ontology/meta	Meta Model
http://www.snik.eu/ontology/bb	Textbook [1]
http://www.snik.eu/ontology/ob	Textbook [2]
http://www.snik.eu/ontology/he	Textbook [3]
http://www.snik.eu/ontology/ciox	CIO Interview
http://www.snik.eu/ontology/it4it	Standard [**?**]

5. Sources

Three textbooks provide different views on the domain of Hospital Information Management: [1] presents a broad view on "typical architectures of health information systems and their systematic strategic management". [2] concentrates on the *tactical* management of information systems in general and on healthcare in particular. The focus of tactical management lies on the planning and operation of projects. [3] explains information management beyond the scope of healthcare. Other sources are interviews and standards.

6. Data Model

[4] contains an initial structure of the meta model. The "meta model" defines three basic disjunctive classes and their possible relations: Roles (who), Function (does what) and Entity Types (and which information is therefore needed). A set of modular subontologies define subclasses of those three classes and their relations as described by a certain knowledge source about information management in hospitals: The **Semantic Network of Information Management in Hospitals** (SNIK[1]) is a modular OWL 2 DL ontology.

7. Data Set Description

The SNIK ontologies are available at http://www.snik.eu/ontology/ under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.

8. Lifecycle

When we add a new textbook to SNIK, we identify all concepts that fit the meta model and their relations to other concepts. Some project members are not proficient in RDF serializations, so we first fill out a meta model conforming spreadsheet template. We then use Tarql[2] with a mapping configuration file[3] to convert that spreadsheet to RDF. Next, we refine the data in Protégé and upload it to a Virtuso SPARQL endpoint.

After the initial publish phase, we describe the repeating steps of the reuse phrase with the SNIK lifecycle in Figure 2:

1. the Virtuoso SPARQL endpoint is used for querying and as data source for all our applications
2. an instance of the OntoWiki [6] is used for punctual changes and additions by non-Semantic-Web-experts, supplemented by SPARUL for changes involving large numbers of classes
3. LIMES [7] generates interlink candidates between the different textbooks, which can be approved or rejected by the users
4. LodView[4] lets the user browse among the classes of SNIK and view particular classes in detail

[1] Hospital means "Krankenhaus" in German.
[2] https://github.com/cygri/tarql
[3] https://github.com/IMISE/snik-csv2rdf
[4] https://lodview.it/

Fig. 1. The SNIK meta model

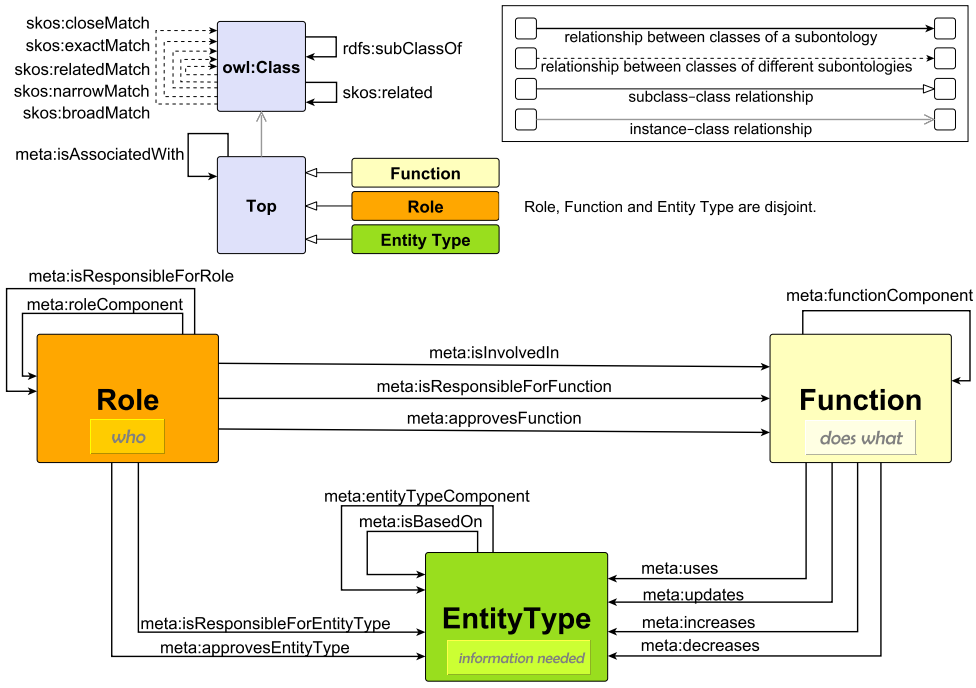


Table 2
use of established vocabularies

Vocabulary	Description

5. SNIK Graph, detailed in Section 9.1, supplements the RDF browser by visualizing the relationships between classes. Users often discover incorrect modelling or possibilities for enhancement in SNIK graph.
6. SNIK quality is a web application that uses SPARQL queries to find problems of varying degree, such as violations of domain, range or naming conventions, subclass cycles or missing definitions.
7. Problems suggested by users and SNIK quality are verified, solved, and published on the SPARQL endpoint using SPARUL queries, at which point the circle begins anew.

The technical environment of those services is described in [8]. Selected applications are described in the following section.

9. Applications

SNIK was initially intended [4] both as software to support hospital CIOs and to support teaching. For the former goal, two applications were designed and implemented as prototypes [9]: (1) the requirements engineering decision support system TOREOnto and (2) the knowledge exploration and navigation visualization CIONx. Due to internal regulations, efforts to integrate them into the informational infrastructure of the Uniklinikum Leipzig were stopped. Consequently, the other approaches are in the area of teaching support:

9.1. SNIK Graph

SNIK Graph is a web application[5] that transforms the classes and triples of SNIK to nodes and edges

[5] http://www.snik.eu/graph

reference franziskas poster in both figures

get a higher quality version of overview graphic

reference [10]

FJ: In

Fig. 2. The SNIK lifecycle

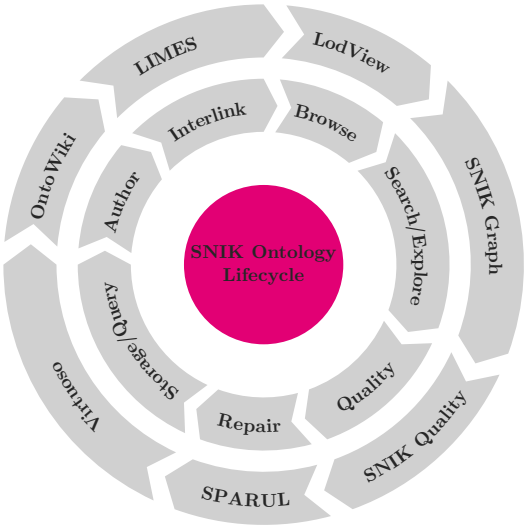
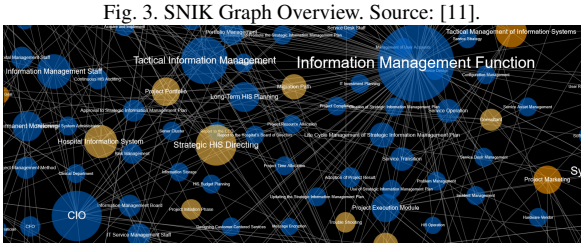


Fig. 3. SNIK Graph Overview. Source: [11].



of a graph. The graph is visualized using the Cytoscape.js [12] library with the force-directed Euler layout, see fig. 3. With several thousand classes, specific parts of SNIK can be hard to discern. Thus, there are several options to view subgraphs of SNIK, for example to show only a specific chapter of a book to prepare a lecture about a specific topic. Users can also search for and restrict the view to a single class and then show the neighbourhood of that class (see fig. 4) and subsequently the neighbourhood of selected neighbours of the previous step. SNIK Graph can also calculate the shortest path to between classes. We also tried to automatically find the most interesting path but this was not successfull as it is subjective and the depends on the goal of the user. Path and neighourhood operations are joined in the "spider worm" (see fig. 5), which consists of the shortest path between a start node and an end node together with the end node's neighbourhood, illustrate the context of a concept.

Fig. 4. A teacher prepares a lecture on strategic HIS directing using the "Circle Star" function of SNIK Graph. Source: [11].

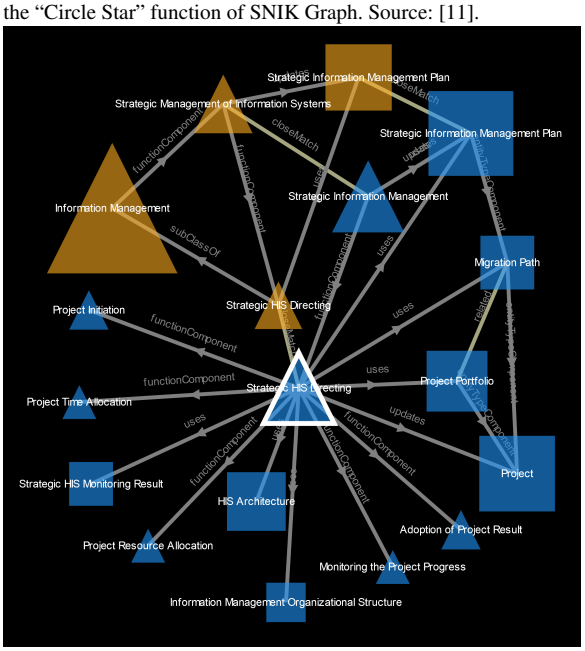
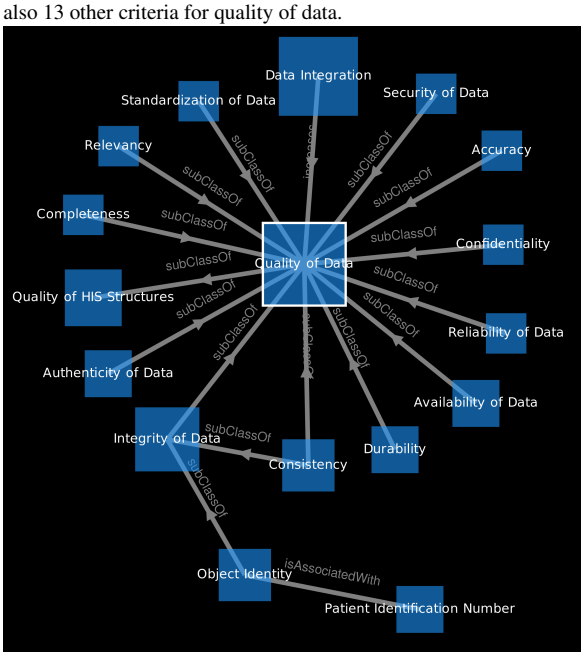


Fig. 5. Students learn new concepts about HIS quality by linking them to concepts already learned. A teacher asks a student to find out how the new concept "Quality of Data" is linked to the "Patient Identification Number". The student connects the two concepts by using the "spiderworm" visualization and learns that a patient identification number is associated with object identity. Object identity is a subclass of integrity of data. Besides integrity of data, there are also 13 other criteria for quality of data.



9.2. Quiz

The DBpedia-powered Clover Quiz [13] shows that ontologies can be used to automatically generate multiple-choice questions. We apply this approach to SNIK and generate 1231 English questions using the templates described in table 3. It was used by students from the Universities of Amsterdam, Heidelberg and Leipzig during the international Frank van Swieten lectures in 2019. To provide difficult wrong answers, the so called *distractors*, we use direct neighbours of the classes that represent the correct answers. Due to the limited nesting capabilties of SPARQL, which does not support loops but only subqueries that cannot access variables declared outside their scope, querying for distinct sets of 4 neighbours, where exactly one was connected over a certain property to a certain object, was not possible with our Virtuoso SPARQL endpoint initially. We circumvented the timeout by calculating the neighbour-relation as a first step and uploading it to the endpoint.

Because SNIK only contains the knowledge that the source textbooks describe, it does not contain the complete domain of Hospital Information Management. As such, negative questions are problematic, as the given relationship could hold in the real world but not be described in the textbook source of the class. The same problem concerns the distractors of positive questions. So this problem cannot be avoided. However, this is one of the reasons that we don't ask count questions like "How many functions is the CIO responsible for?", which don't help much for learning anyways. Relationships could also hold implicitly through the subclass hierarchy. For example, `bb:ChiefInformationOfficer` is a subclass of `bb:InformationManagementStaff`, which is responsible for `bb:OperationalInformationManagement`. It is unclear, whether the CIO is also responsible for operational information management. In the future, we will add more question types

10. Discussion

SNIK is based on the meta model from Figure . Thus, SNIK can be considered as an archetype for ontologies describing for a given domain (here "HIS management"), what functions a certain role has to carry out and what information a person of this role needs and what information she/he provides while carrying out the functions. This may be used for domains like the medical department of a hospital as well. This would be valuable knowledge, for instance as basis for systems analysis projects [1]. Different user groups can use the interfaces of SNIK that are most suitable to them. For example, the IT management of a hospital can use SNIK as a vocabulary to integrate data from different formats that result from application components from different vendors. Using the SPARQL1 query language, experts can also integrate applications with the SNIK ontology. Experts can also use the SPARQL query editor to answer specific questions, see Table 2. Students and teachers, on the other hand, benefit more from SNIK Graph, which intuitively presents knowledge without the need for technical skills.

Figure 4– A "Spider Worm" between the classes "Change Management" and "HIS Quality" in the SNIK Graph visualization web application. The content of SNIK is taken from textbooks which are protected by property rights of publishing houses or taken from an interview with a CIO holding property rights as well. Making the SNIK ontology openly available may therefore conflict with these property rights. In case of SNIK, we appreciate the approvals of the publishing houses Springer, de Gruyter and Schattauer and of the CIO. In general, missing approvals may prevent the creation of such open ontologies and thus the provision of textbook content as open knowledge. The approvals for SNIK were possible because for each class and relation being part of SNIK there is a clear indication of its source (see code "bb" and chapter information in Table 1). In future work we will use these indications for establishing hyperlinks into the sources such that users of SNIK may use SNIK as a smart index for retrieving more detailed knowledge and supporting illustrations from the original textbooks – if digitally available. We are convinced that this way an ontology like SNIK is not a competitor of textbooks but a smart guide for using a textbook. Thus, it might be attractive for publishing houses.

11. Conclusions

We showed that knowledge on the management of information systems in medicine and health care is made publicly available using open standards over several interfaces with different compromises between expressivity and accessibility. It can be combined with other knowledge in biomedical and health informatics and in other disciplines.

reference

FJ: Ich glaube, das müsste man ausführlicher erklären, das würde ich eher ans Ende der Diskussion nehmen.

verweis auf existierende papers

benutzung durch braunschweig

benutzung durch amsterdam

Table 3
Templates

Template	Description
Definition	Ask for the class that fits the given textbook definition.
Example	What is defined as “Examination of in and out patients in radiological department”?
Distractors	Labels of other classes (that have a path of length 2 or less to the correct class.)
Subject	Ask for the class that is related via a given relation to a given object.
Example	Who is *involved in* a *healthcare network*?
Distractors	Labels of other classes (of the same type) that *are not* related via the same relation to the same object.

12. Future Work

Connection with HITO

We plan to extend the text-based interlinks with semantic-similarity-based ontology matching [5] to find more correspondencies between the textbooks. This will enable users to switch to another textbook when a certain concept is explained there in more detail. The SNIK toolset, SNIK Graph in particular, is under permanent improvement. For instance, we are considering alternatives to the shortest path that are the most preferred by different audiences and whether those paths can be (semi-)automatically generated using suitable node or edge weight functions. We are also investigating candidates for additional subontologies and as interlink targets.

13. Acknowledgments

This work is supported by the DFG (German Research Foundation) under the Projects SNIK, grant no. 1605/7-1 and 1387/8-1, as well as HITO, grant no. WI 1605/11-1 and I3726-N31.

FJ: Hmmm, ziemlich selbstreferenziell. Brauchen wir da wirklich so viele Publikationen von uns? Welche anderen könnte man noch einbringen? Ich schaue auch mal, was mir so über den Weg läuft.

References

[1] A. Winter, R. Haux, E. Ammenwerth, B. Brigl, N. Hellrung and F. Jahn, *Health Information Systems: Architectures and Strategies*, Health Informatics, Springer London, 2011. ISBN 9781849964418. https://books.google.de/books?id=RzvmrgwCWncC.

[2] E. Ammenwerth and A. Bess, *IT-Projektmanagement im Gesundheitswesen*, 2nd edn, Schattauer, Stuttgart, Germany, 2014. ISBN 9783794530717.

[3] L.J. Heinrich, R. Riedl and D. Stelzer, *Informationsmanagement: Grundlagen, Aufgaben, Methoden*, De Gruyter, 2014.

[4] M. Schaaf, F. Jahn, K. Tahar, C. Kucherer, A. Winter and B. Paech, Entwicklung und Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus, *Informatik 2015* **246** (2015).

[5] K. Tahar, M. Schaaf, F. Jahn, C. Kücherer, B. Paech, H. Herre and A. Winter, An Approach to Support Collaborative Ontology Construction, *Studies in health technology and informatics* **228** (2016), 369–373.

[6] S. Auer, S. Dietzold, J. Lehmann and T. Riechert, OntoWiki: A Tool for Social, Semantic Collaboration, in: *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at the 16th International World Wide Web Conference (WWW2007) Banff, Canada, May 8, 2007*, N.F. Noy, H. Alani, G. Stumme, P. Mika, Y. Sure and D. Vrandecic, eds, CEUR Workshop Proceedings, Vol. 273, CEUR-WS.org, 2007.

[7] A.-C. Ngonga Ngomo and S. Auer, LIMES—A Time-Efficient Approach for Large-Scale Link Discovery on the Web of Data, in: *IJCAI–11, Proceedings of the 24th International Joint Conference on Artificial Intelligence*, AAAI Press, Palo Alto, California, USA, 2011.

[8] K. Höffner, F. Jahn, C. Kücherer, B. Paech, B. Schneider, M. Schöbel, S. Stäubert and A. Winter, Technical Environment for Developing the SNIK Ontology of Information Management in Hospitals, *Studies in Health Technology and Informatics* **243** (2017), 122–126.

[9] C. Kücherer, Domain-specific Adaptation of Requirements Engineering Methods, PhD thesis, 2018.

[10] M. Schaaf, F. Jahn, K. Tahar, C. Kücherer, A. Winter and B. Paech, Visualization of Large Ontologies in University Education from a Tool Point of View, in: *MIE*, Vol. 228, 2016, pp. 349–353. doi:10.3233/978-1-61499-678-1-349.

[11] J. F, H. K, S. B, L. A, P. T, A. E and W. A, The SNIK Graph: Visualization of a Medical Informatics Ontology, in: *MedInfo 2019, The 17th World Congress of Medical and Health Informatics, Lyon*, 2019.

[12] M. Franz, C.T. Lopes, G. Huck, Y. Dong, O. Sumer and G.D. Bader, Cytoscape.js: A graph theory library for visualisation and analysis, *Bioinformatics* **32**(2) (2015), 309–311.

[13] G. Vega-Gorgojo, Clover Quiz: a trivia game powered by DBpedia, *Semantic Web* (2018), 1–15.

14. Notes to Reviewers

This paper includes and extends text and figures from: [11]