Contrasting Ontology Modeling with Correlation Rules for Delivery Applications

Dana Baro[1]*, Laura* Neundörfer [1]

[1]Karlsruhe University of Applied Sciences, Faculty of Information Management and Media, 76133 Karlsruhe, Germany

Abstract. With the increasing importance of knowledge management, variant management and the ever-growing quantity of data, ontologies emerged as a form of knowledge representation, especially in the field of technical communication for modelling metadata and to create correlations between them. In the area of delivery applications, the deliverable information objects receive a certain intelligence by semantic metadata. It is expected, that ontologies offer a higher level of intelligence which could lead to an improvement in classification, connection and delivery possibilities of content. On the contrary, creating those complex ontologies requires a time-consuming effort. Thus, the question arises, whether their use offers a decisive added benefit or if alternatives, such as untyped correlations, should be preferred. In that case, the concept of Semantic Correlation Rules can offer an opportunity to derive advantages from ontologies: By defining which classifications are connected to others, it is possible to present content tailored to user-specific information requirements. By developing use cases, we aim to evaluate the required level of intelligence of the metadata resulting from its modeling method to achieve this goal.

1 Introduction

Not only knowledge management and complex variant management requirements, but also the request for short customer-specific information, urge information architects to seek more suitable technologies for delivery applications [1]. For automatic systems like Content Delivery Portals (CDP), the question arises how suitable sections of the world can be delimited, analyzed, modeled, formalized and represented in a suitable way [2]. Ontologies offer a way to implement those requests. In the following, we aim to answer the questions whether use-case-specific information really requires the effort of creating a complex domain ontology or if fewer complex models, like Semantic Correlation Rules, already fulfill parts of the requirements.

2 Context of research

This paper is the result of the master's program at the University of Applied Sciences Karlsruhe, Germany, in collaboration with the University of Aizu, Japan. As part of the Semantic Information Management module, an ontology was developed as a metadata model for the provision of content in delivery applications on the subject of smart technologies.

3 Technological background

3.1 Ontologies

In a superior context, ontologies are a specification of conceptualization, which means the description of concepts and relations between them. In the context of artificial intelligence, an ontology can be created by defining representational terms such as classes, properties, and individuals.

Formally, an ontology can be described as the statement of a logical theory [3]. Ontological commitments are described by the common ontology to ensure that a set of agents can communicate about a domain of discourse. It can be stated that these are agreements to use the shared vocabulary of the ontology in a coherent and consistent manner. Therefore, the observable actions of the agents must be consistent with the definition in the ontology. Typically, these are then visualized by semantic networks [1].

Within this project, the ontology modeling was implemented in the tool Protégé and the subject of smart technologies. Because of the great dimension of this subject, a framework was agreed on which main parts of this subject should be displayed and how they should be displayed. The focus was placed on modeling fields of applications for smart home systems, soft- and hardware, physical parameters that can be measured, control options and the connectivity. Moreover, an information class (I-Class) was added to classify the content regarding its usage.

With Protégé, an open-source software to model ontologies, the framework of classes, subclasses, individuals, and properties offered by OWL was used to model the ontology. OWL is a semantic web language designed to represent comprehensive and complex knowledge about things, groups of things, and relations between them [4]. **Fig. 1** shows an extract of the scope of the classes and individuals of our ontology representing the metadata.
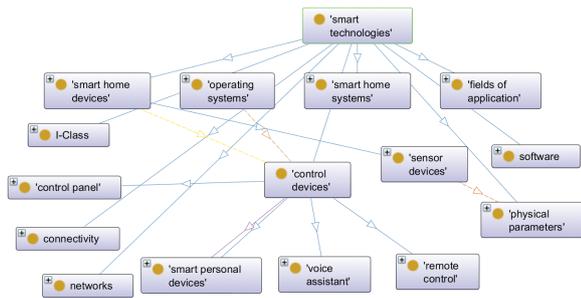
Fig. 1 Extract from the ontology model in Protégé

3.2 Semantic Correlation Rules

The key characteristic of an ontology-based system is its capability to handle complex structures of classes and individuals, pattern detection and the creation of correlations [3].

Driven by the desire to provide information as specific as possible, but at the necessary scale, Semantic Correlation Rules (SCR) deal with deriving the logical concept, the relevant context and the amount of content required.

This happens at different levels of rules and semantic models to provide all the required content for the user's case. The need for technical solutions, like SCR, arises from the lack of context of a topic and the overabundance of content in a large document [2]. The relevant content provided by the use case can be transmitted to the user with the necessary context as a microDoc, which is created by the networking of topics or other information units [1]. The prerequisite for the use of microDocs is intelligent information. The greater the semantic richness of the semantic model is, the more precise the logic is, when compiling the relevant content and its context. Ontologies provide the opportunity to create a (metadata) model with high semantic richness and to link it in a rule-based way. These rules can be used for the dynamic aggregation of the microDocs after defining the use cases [6].

SCR are derived as follows: The InRule results from the classification of the content searched or used; the OutRule results from the classification of the associated content, which should be presented to the user as necessary additional information. The linking of the In- and OutRule enables the scenario-specific provision of all information required by the user and thus, reduces the information retrieval process.

It is important, that SCR are independent from specific authoring environments and can be implemented and used within delivery and search systems. [10]

3.3 Prerequisites for using Semantic Correlation Rules

3.3.1 Metadata

To implement Semantic Correlation Rules, it is necessary to present metadata classifying the content. Represented by SCR, links can be implemented with these classifications.

Hereby, it is irrelevant whether the metadata are presented in a taxonomic or ontological form. Important is, that the content is tagged with metadata, so it can be found within the system.

In our case, the data from the modeled ontology was used as metadata for classification and definition of SCR.

3.3.2 Use case scenarios

The usage of Semantic Correlation Rules assumes that there are certain use cases in which users need specific information, available in different information objects and in addition to the information object already found. These information objects show the same classification and can, therefore, be defined in form of SCR.

4 Use case modeling

4.1 Definition

At first, it is important to clarify the definition of use cases in the specific environment of our ontology and the scope of this project. Use cases are artificially developed scenarios which can be used to record requirements for the ontology and implementation of SCR. They represent specific scenarios in which a user searches for certain information. The use cases are well suited for communication and for analysing system performance because they can be used to define the functionality of the system from the user's point of view.

In the given case, we used the use cases to test the Semantic Correlation Rules with different system providers, as they ensure all testing processes are conducted in the same way to make a final comparison possible.

The reasons why we decided to test the systems with use cases are described in the following. Since the topic of smart technologies is quite broad overall, use cases help to set a frame for the needed content and the way of usage of the ontology.

Developing use cases and getting a deeper knowledge of the overall topic, makes it possible to understand the needs of the target audience and find natural relations between classes which improves the structure of the "to model" ontology. As all ontologies that are implemented in different system ventures are tested using the same guidelines, the use cases dictate it as possible to achieve comparable results in the end.

4.2 Method of the use case modeling

4.2.1 Content research

The first step in the use case modeling process was to research relevant content as the basis for which information is displayed to the user in content delivery portals.

Therefore, we started to investigate in various information about the general topic of smart technologies, which can be pdf-documents, websites, or manuals for specific products.

Since this topic is very broad, we decided to sort the content we found in order to focus our research more on specific subject areas.

4.2.2 Content sorting

Similarities between information could already be found during the content research process. The overall topic of smart technologies could be reduced to specific topics such as smart home products, manuals, descriptions or areas of application.

It was then possible to sort and group the content allocated so far, which also showed relationships with one another.

4.2.3 Actor development

When researching and sorting content, it became apparent, that not all content is related to one another. For this reason, we developed different actors representing users who interact with the Content Delivery Portal. These actors facilitated the development of the user scenarios on basis of different problems and needs. This procedure additionally ensured that the necessary subjects were covered within the ontology. As we defined actors of different ages, environments, and knowledge about smart technologies, we were able to represent a wide-range target audience which will use the CDP and use cases for implementing different authentic use cases for SCR in future.

4.2.4 Scenario development

On basis of the actors needs and the content found, the development of an extensive user scenario was initiated. While interacting with the CDP, an actor pursues tasks that are involved in the user scenario. One example for a created use case is an actor, who already has an existing smart home system from a certain company at home and searches for additional devices. Since he is not often at home, the special need of this actor is the automatic adjustment of the temperature in his rooms. Therefore, the given scenario and main goal is to find a fitting climate monitoring for the already existing smart home system.

What our actor is looking for in the CDP is the brand and product of his system as well as the area of application to get more information about which climate monitoring system is compatible.

4.2.5 Content classification

After the development of the specific scenarios, all the documents were classified regarding their content, using the metadata from the ontology.

These classifications are needed to ensure that the content can be found in delivery applications.

4.2.6 Implementation of Semantic Correlation Rules

The given classifications in combination with the use cases built the basis for the implementation of SCR: The InRules represent the classification of the starting document which the user accessed in the given use case. By defining the OutRules, additional information on the user needs in the use case can be considered. Depending on the InRules, the related content from the associated OutRules is displayed in the delivery application to the actor and provides the appropriate information based on the use case.

4.2.7 Prioritization of results

As a final step, the order of the additional information from the OutRules was elaborated and prioritized by defining strengths for each rule. Prioritizing the OutRules ensures the most appropriate presentation of information possible.

5 Software implementation

5.1 Ontology system

As the focus of the project was also on testing the SCR in different delivery applications, the modeled ontology had to be imported into different software environments. We worked with i-views, a smart data engine to build knowledge graphs like ontologies to represent knowledge domains. The ontology modeling tool supports its own Content Delivery Portal and enables a direct connection between the ontology and the content displayed in the CDP.

5.1.1 Ontology import

The first obstacle depicted was the data format of the ontology: Although it is using parts of RDF, OWL is not yet supported by the system providers. Therefore, we developed a code to transform OWL to pure RDF which is supported by the system providers. RDF (Resource Description Framework) is a markup language, used for modeling metadata for resources in the internet [6].

To elaborate on the differences, we compared the OWL code with an exported code from i-views.

Based on the results, we took the differences into account by developing the transformation code (**Fig. 2**). After this transformation, the import was possible and the ontology could be presented and used in i-views.

```
<owl:NamedIndividual rdf:about="http://www.hs-karlsruhe.de/sim#CO2">
    <rdf:type rdf:resource="http://www.hs-karlsruhe.de/sim#physical_parameters"/>
</owl:NamedIndividual>

<sim:physical_parameters rdf:about="http://www.hs-karlsruhe.de/sim#CO2">
    <rdf:type rdf:resource="http://www.hs-karlsruhe.de/sim#physical_parameters"/>
</sim:physical_parameters>

<owl:ObjectProperty rdf:about="http://www.hs-karlsruhe.de/sim#is_applicable_for">
    <rdfs:domain rdf:resource="http://www.hs-karlsruhe.de/sim#smart_home_devices"/>
    <rdfs:domain rdf:resource="http://www.hs-karlsruhe.de/sim#smart_home_systems"/>
    <rdfs:range rdf:resource="http://www.hs-karlsruhe.de/sim#fields_of_application"/>
</owl:ObjectProperty>

<rdf:Property rdf:about="http://www.hs-karlsruhe.de/sim#is_applicable_for">
<rdfs:range rdf:resource="http://www.hs-karlsruhe.de/sim#fields_of_application"/>
<rdfs:domain rdf:resource="http://www.hs-karlsruhe.de/sim#smart_home_devices"/>
<rdfs:domain rdf:resource="http://www.hs-karlsruhe.de/sim#smart_home_systems"/>
</rdf:Property>
```

Fig. 2 Example of the OWL to RDF transformation

Fig. 3 shows an extract of the ontology in i-views. It is obvious, that the definition of classes, their subclasses, and the individuals are very close to hierarchical metadata.

What makes the difference and increases the intelligence of the metadata, are the relations (properties) that can be evaluated, as far as they are defined and applied. This finding in particular makes the ontology modeling very complex. This modeling effort increases continuously with the quantity of entities within the ontology.

Fig. 3 Extract from the ontology model in i-views

5.1.2 Content integration

To import the content to the CDP, it was necessary to use iiRDS. iiRDS is a standard which enables dynamic information request and delivery in the area of the Internet of Things and Industry 4.0 [8, 9].

Every information object was packaged into an iiRDS bundle and uploaded onto the CDP. We classified the content by using a representative content ID. This ID had to be assigned to the generated single object iiRDS packages to apply the metadata of our ontology to the physical content. Consequently, each content ID we assigned had to be changed to the generated iiRDS-ID within the package. This was also done with a transformation script. Extending the offered metadata from iiRDS with company or product specific metadata is only possible in a restricted manner. Within the usage of i-views, it was a suitable procedure to use iiRDS only as an exchange format to import content to the CDP. Our connecting points of the content were the subclasses from the class "information unit" from iiRDS. By using the iiRDS namespace and the associated tag, we assigned the content to the corresponding point, in this case "Topic" (**Fig. 4**, **Fig. 5**).

```
<iiRDS:Topic rdf:about="urn:uuid:40ee5cdf-5086-4711-9ae2-b39ed0c79b2d">
    <rdf:type rdf:resource="http://iirds.tekom.de/iirds#Topic"/>
</iiRDS:Topic>
```

Fig. 4 Extract from the code of an information unit

Fig. 5 Connection of an information unit to iiRDS and the classification

5.2 Conclusion

5.2.1 Software implementation and iiRDS

The workaround described in chapter 5.1.2 shows that iiRDS has a certain implementation standard in CDP, but the usage is depending on the integration of external metadata to iiRDS and restricted by the insufficient integration methods. This circumstance exists regardless of our procedure. Due to the fact that content retrieval in CDP can only work circumferentially with classifications beyond information specific metadata, iiRDS shows expandability regarding metadata extensions.

By using i-views as CDP, it was possible to classify the content with metadata from an ontology modeled in the software, but this work must be done manually as we previously did in Protégé.

Due to their system architecture, ontology-based systems offer a suitable possibility to implement and take advantage of the idea of SCR. But, since an ontology model is not mandatorily necessary, other technical solutions within content delivery applications are conceivable to implement SCR independently of ontologies.

5.2.2 Contrast SCR modeling with domain ontology modeling

The first difference between full domain ontologies and SCR is the field of application: Ontologies are originally used to represent knowledge domains whereas Semantic Correlation Rules pursue the target to implement correlations between classifications of information objects.

There is also an important difference to simple linking: SCR link classifications, so InRules are triggered by content, which shows the classification defined by the InRule. Content classified according to the OutRule specifications is presented additionally.

Another decisive difference is the fact, that the entire ontology and its components, meaning classes, individuals, and properties, must be defined. Whereas SCR offer a defined base by which the use cases can be implemented directly. This means an enormous reduction in the modeling effort. The extensive ontology model offers an interpretation of unlimited properties, while SCR needs only about five predefined relations. Hence, another difference can be deducted: An ontology offers an unlimited number of applications by the possibility of evaluating every defined property. On the contrary, SCR are used for a limited amount of information delivery for specific use cases.

By implementing the SCR, we concluded that an implementation without an ontology by using possibly existing metadata, such as classic hierarchical metadata, can also be used to define In- and OutRules. This could be important for companies, that do not have the capacities for modeling an entire product ontology, but use cases they would like to implement in delivery applications by defining SCR. A characteristic of these use cases is the representation of situations in which a user needs certain information that goes beyond obvious, or statistically derivable, rules and equal metadata. Moreover, during the testing of the SCR in different software environments, it became apparent, that the In- and OutRules must be defined as as precisely as possible. On one hand, to ensure that InRules do not overlap unintentionally and, on the other hand, to assure that the number of the presented related content does not rise uncontrollably through matching classifications.

6 Summary

To summarize the success of the project, especially the effort for the entire process of ontology and use case modeling should be emphasized.
For implementing use-case-specific information and their relations, SCR are very useful as they are representing the user's information needs in a certain use-case in a machine-processable way. The ontology formed the basis for the therefore necessary metadata model.
If a highly automated representation of related content is not required, the implementation of Semantic Correlation Rules is more efficient than the ontology modeling. Having a hierarchical metadata model in place significantly reduces the effort required to implement SCR compared to previous ontology modeling.
Nevertheless, ontologies offer many possibilities to increase the level of intelligence [7] of the content and implement (other) relations and dependency rules between content. They can also be used for SCR, however, it must be decided, whether there are more areas of application than SCR, since the creation of ontologies is not profitable for only semantic correlation rules.
While nowadays, the use of ontologies for annotating technical documentation content is quite popular, the ontologies actually used (such as iiRDS) usually do not

provide formal semantic rules in the sense of description logic, although the standards such as OWL support rules based on a description logic approach. The main reason is that description-logic-based rules are usually quite complex and not easy to understand and define. A combination of simple ontologies and a lightweight correlation approach could therefore be a proper approach to enable non-experts to define relationships between content.

References

1. W. Ziegler, *Extending intelligent content delivery in technical communication by semantics: microdocuments and content services*, The 2nd ACM Chapter International Conference on Educational Technology, Language and Technical Communication ETLTC (2020)

2. W. Ziegler, *Regelmäßig Inhalte vernetzen*, tc world online conference (2020)

3. T. Moser et al., *Semantic Event Correlation Using Ontologies* (2009)

4. OWL Working Group *OWL*, https://www.w3.org/OWL/ (2012) [last call 2021/01/25]

5. Ryte Wiki *RDF*, https://de.ryte.com/wiki/RDF (2020) [last call 2021/01/26]

6. W. Ziegler, *Delivery zwischen Kontext und Content,* technische kommunikation, Nr. 06/19. 58–61 (2019)

7. W. Ziegler, *Alles hängt mit allem zusammen - Content vom Anfang bis zum Ende gedacht,* tekom Jahrestagung (2019)

8. tekom Deutschland, *iiRDS – The International Standard for intelligent information Request and Delivery,* https://www.iirds.org (n.D.) [last call 2021/03/15]

9. D. Nägele, P. Vobl, *Ontology modelling and standardized information exchange with content delivery applications in technical communication,* The 3rd ACM Chapter International Conference on Educational Technology, Language and Technical Communication ETLTC (2021)

10. W. Ziegler, *Semantic Correlation Rules as a Logic Layer between Content Management and Content Delivery,* The 3rd ACM Chapter International Conference on Educational Technology, Language and Technical Communication ETLTC (2021)