

Semantic Correlation Rules as a Logic Layer between Content Management and Content Delivery

Wolfgang Ziegler^{1*}

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

Abstract. Semantic technologies have recently gained considerable influence and attention in the field of technical communication and information management. While metadata management is already a well-known field of content management technologies, its semantic extension addresses more recently, for example, problems of model-based product development and related content engineering processes. On the other hand, dynamic search technology and content delivery can benefit from semantic modelling by enhancing search functionalities or by integrating various data sources utilizing semantic mapping. In this evolving environment, we propose a logical layer of content correlations as so-called semantic correlation rules (SCR). This layer can be understood as an interface between content management systems, semantic modelling systems and content delivery portals. Semantic correlation rules serve as a light-weight ontology consisting primarily of untyped semantic relations between metadata classes. In doing so, class-to-class linking mechanisms can be implemented in content delivery and search environments while serving as a basis for the previously introduced microDoc concept

1 Introduction

In previous publications, microdocuments (microDocs) and their potential benefit in the field of technical communication (TC) have been introduced [1]. microDocs address the question on how to deliver requested specific information, including sufficient context within relevant use cases of information retrieval. In the domain of TC, the concept has gained interest due to the increase of content delivery applications, respectively, due to the number of vendors of content delivery portals (CDP) and the growing number of industrial implementations. Furthermore, there is a shift towards semantic technologies supporting model-based scenarios. They focus on the areas of content creation within component content management systems (CCMS), content retrieval and delivery within CDP or metadata mapping applying information hubs.

In this article, we propose an implementation and logical layer of microDocs as RDF-based semantic correlation rules (SCR) [2]. We claim that this SCR-representation of microDocs is an easy-to-implement framework for various types of CCMS systems considering CDP and can be modelled into ontology-related software. By using SCR as a light-weight ontology approach, the steps are facilitated towards more sophisticated ontology models.

2 From use cases to correlation rules

microDocs can be understood as a conceptual basis of an improved and contextually-enriched delivery of information. SCR then represent one of the technical implementations of microDocs. Naturally, one goal of this research is to propose SCR as a standardized representation independent from any specific content-related system.

In order to develop the minimum requirements for SCR, we investigated typical scenarios within industrial use cases of companies operating in the field of manufacturing, automation and software industries. Typical problems in the provisioning of information by manual-based documentation, in cases of user-driven information retrieval and various service support cases, were analyzed and discussed in detail with industrial process owners. The considered use cases, therefore, covered relevant phases of the product lifecycle, resp. of the customer journey: sales and sales support, customer support, installation, troubleshooting and help-desk. Thus, we gathered experience and gained insight into the requirements of our corresponding implementation approach.

A core condition on SCR was imposed, that no additional metadata should be required for the implementation. SCR should rely solely on the existing metadata environment given by CCMS or CDP. Naturally, the more complex the pre-defined industrial metadata environment is, the more detailed the contextual information can be defined. Furthermore, an additional decision affected the depth of context required: To limit model complexity and to ensure back-end user acceptance, only a single-level-approach was chosen. Thereby, only “nearest-neighbor” information in the logical vicinity of requested information should be modelled. Hierarchies can then be derived by iterating over subsequent hierarchies of correlated information objects.

A typical, but fictitious situation is depicted in Figure 1. Herein, we used the open-source “PI-Fan” content model as the reference model of two TC-related and industrially well-known metadata models [3]: The PI-Class model and the iRDS standard [4]. The latter is also derived from PI-

* Email: wolfgang.ziegler@hs-karlsruhe.de

Class but was extended for the purpose of packaged and standardized information exchange between CDP and CMS. The illustration shows, how simple correlations between a primary content object and contextually required secondary objects are constructed. The primary object, therefore, represents content, which has been retrieved or pushed within events in search systems like CDP. Hereby, the getting-started information of a fan is correlated with the postulated most-often required information of manual troubleshooting. Additionally, it

displayed error-codes and contact information of the producing company.

It is important to note, that all objects can be identified by corresponding metadata from the underlying CCMS. The metadata of the PI-Fan [3] is depicted alternatively in blue (PI-Class) and black (iiRDS). In the iiRDS environment, custom (blue) metadata values also appear as customer-resp. product-specific values in addition to standard values.

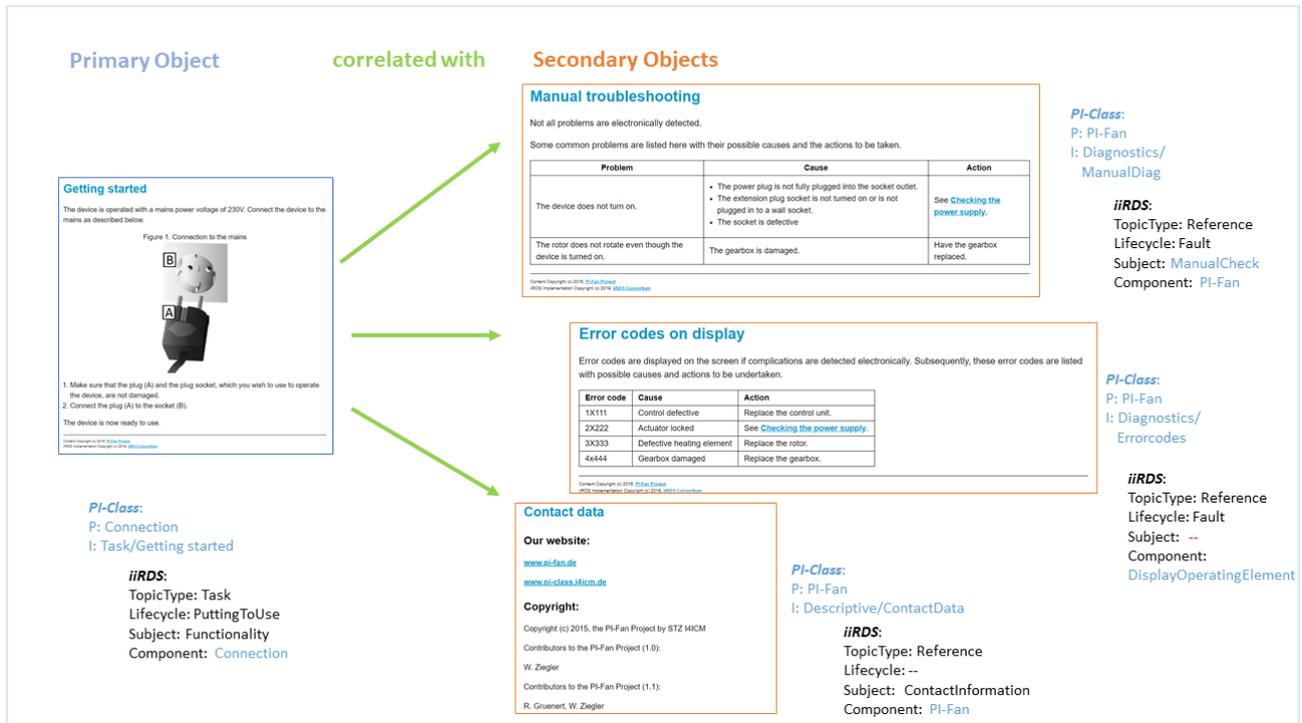


Fig. 1. Reference use case for the definition of microDocs and subsequent semantic correlation rules as described in the text. Metadata of the content objects involved can be expressed according to specific metadata models. Here, PI-Class and iiRDS are depicted as alternative models using different types and values of metadata.

3 Technical implementations

microDocs are based on use cases of context-enhanced information delivery. As mentioned in the previous section, correlation rules can be related to the retrieval and viewing events of information objects in corresponding search and delivery systems. In more detail, relevant use cases can be described as selection rules of primary and secondary information objects while they are connected by semantic relations.

The described concept has been implemented in a standardized RDFS/OWL representation [5,6], depicted by its main components in Fig 2: The primary information objects – in this reference use case once again one of the topics of the PI-Fan model – are characterized by so-called *InRule* instances and their relation (*scr:selects*) to given metadata (e.g. *pifan#connection*). The connection to secondary objects is given by 1:N-relations (*scr:hasCorrelation*). Finally, these secondary objects are

characterized by *OutRule* instances carrying once more a corresponding set of selectors (*scr:selects*), pointing each to the required object meta data, resp. instances (e.g. *iiRDS#Fault*). Hence, secondary objects can be identified in the CDP, displayed in the user interface or bundled for a web-service delivery.

Within this article, we only describe the basic concepts of SCR for the first development versions [7]. The full definition covers additional details, for example, the binding of secondary objects by a given correlation strength used for sequencing the visualization for content within microDocs. A further relation is given by *scr>equals* bound to *OutRule* instances, setting, for example, a generic product or user context for microDocs. The product context will be relevant in single topic delivery scenarios where no document or product context is given. The relation also allows for reusing secondary *OutRules* in various contexts.

In further versions of SCR, extensions of the model can cover additional features like dynamic rules. They can, for example, cover function-calls in the system environments or required user interactions leading to a better guided-search behavior of CDP.

It is important to note, that even though the relations (*scr:hasCorrelation*) are technically semantic, they are only weakly typed. Further detailing and sub-typing of the relations at this level of SCR usage is not necessary and keeps implementation simplicity.

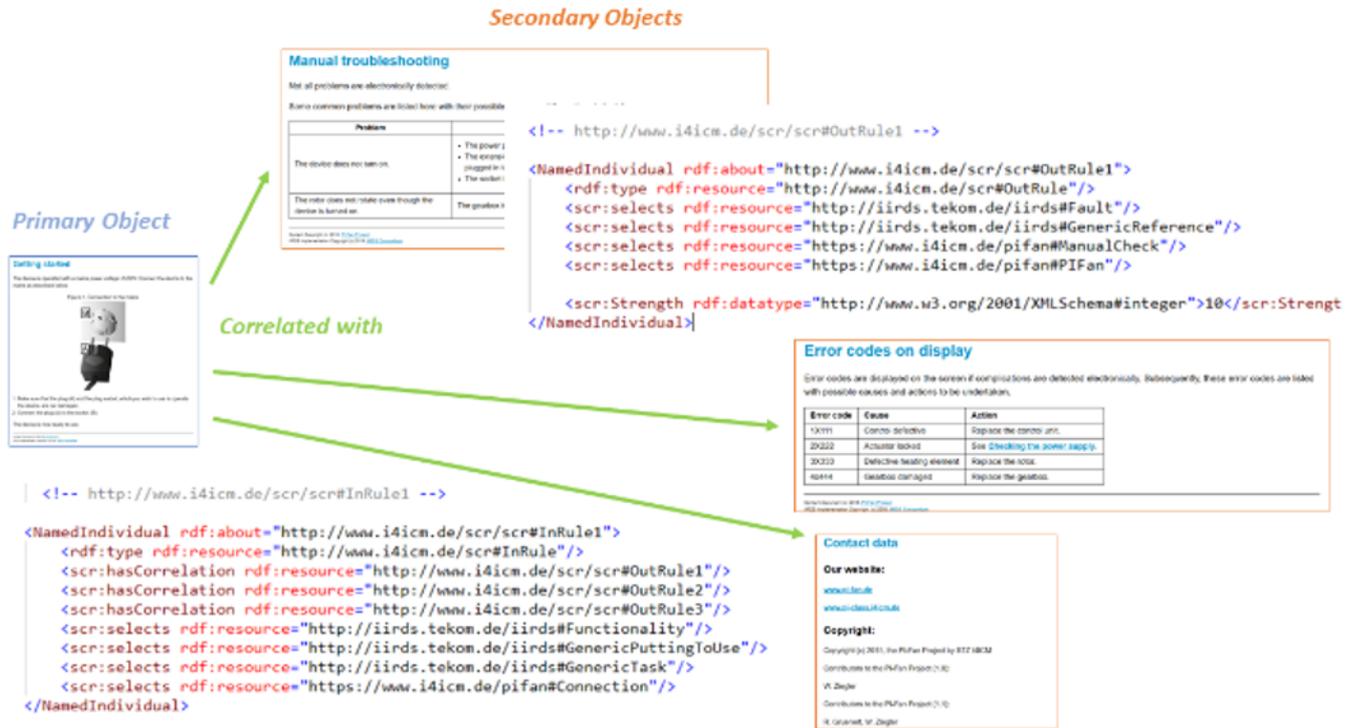


Fig. 2. Main components of RDFS/OWL representation of the semantic correlation rules using iirds and PI-Fan metadata.

The introduced formalism can also be understood as a constraint mechanism in the framework of ontology modelling. Compared to more elaborated constraint formalisms in this area like SHACL [8], it is a simpler approach, derived from and dedicated to, as emphasized, content delivery-related use cases.

4 System environments

System implementations of SCR can take place in different system types as shown schematically in Figure 3.

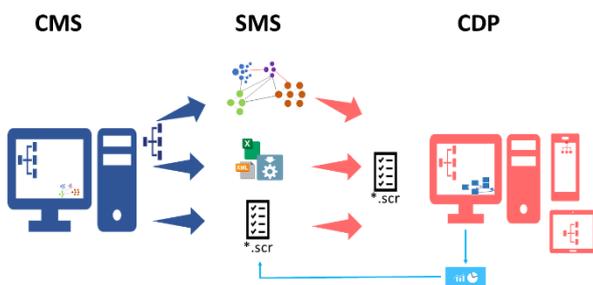


Fig. 3. Different system types for creating and processing SCR.

On one hand, SCR could be defined in the metadata and authoring environment on the system (C)CMS. In this case, of course, only information objects from this content

source will be affected by SCR-based search enhancements. On the other hand, within semantic modelling systems SMS (i.e. ontology models and graph databases), SCR can be included in all types of model-based (content) engineering approaches by using the existing semantic metadata.

Therefore, both system types can serve as environments for SCR-modelling and should support export-mechanisms for SCR in the standardized RDFS/OWL or compatible formats. Certainly, other ways of creating SCR as plain XML-files are possible. CDP must then perform the tasks of importing, managing, and processing SCR in users' search processes. Finally, they must visualize corresponding microDocs.

Prototypical examples of the above mentioned systems and SCR-implementations are shown in figure 4 to figure 7.

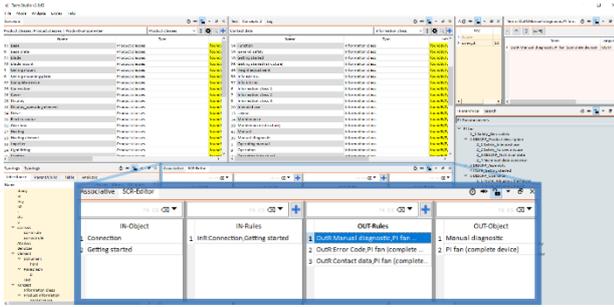


Fig. 4. SCR definition in an CCMS (Klar:suite) by selecting from (yellow marked) topic-metadata [9]. Resulting *In/Out Rules* are displayed in the zoomed viewlet on the bottom.

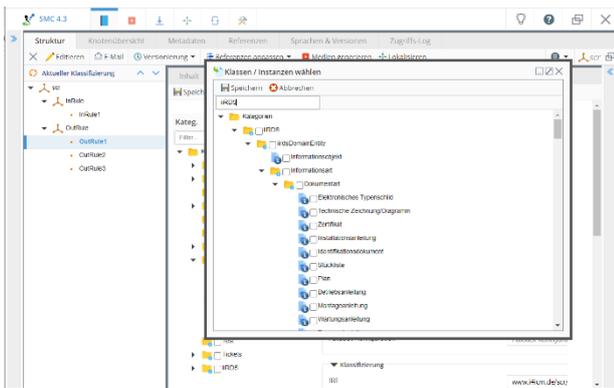


Fig. 5. SCR definition (left side) in an (C)CMS, here Smart Media Creator [10]. Selectors are constructed by selection from the metadata taxonomies.

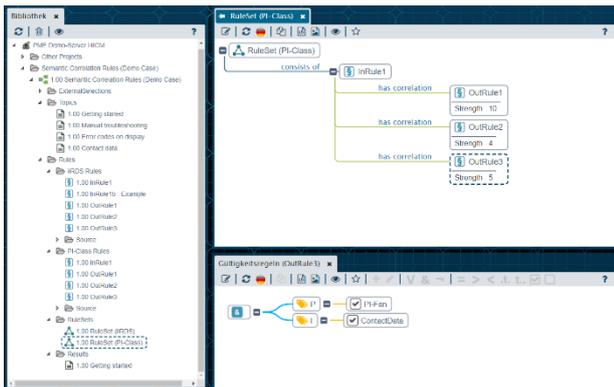


Fig. 6. SCR definition (upper right window) in an SMS, here Ontolis [11]. Selectors are defined within an ontology model as Boolean operators (lower right) and can be managed in the system.

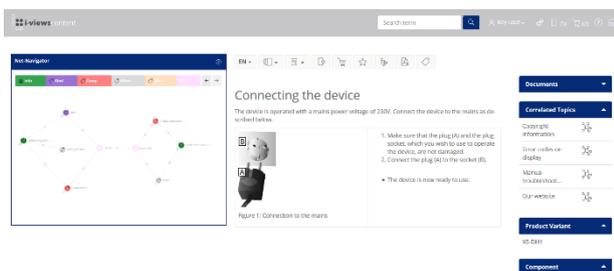


Fig. 7. Dynamic SCR processing within a CDP environment, here I-Views Content. Secondary objects are displayed as correlated objects on the right side. A semantic graph view of

SCR can be displayed (left side) as the system used has a primary SMS functionality [12].

Beside the implementation in user interfaces as shown in figure 7, the SCR can be accessed by API-calls and processed by corresponding web-services as in the c-rex environment [13] described in [2]. In general, the API-calls initiated by primary events, trigger graph traversals leading to the provisioning of metadata of correlated secondary objects. However, web services could also process API calls by returning complete microDocs as content packages. This could cover, for example, standardized formats like DITA maps, iIRDS packages, or SCORM learning packages.

5. Discussion and summary

In this article, we proposed a standardized RDFS/OWL notation as an implementation of microDocs for context-aware information delivery. The underlying concept of semantic correlation rules can be realized and implemented in various system types contributing to content-related processes. Using the described semantic technologies, we propose an easy-to-implement and lightweight ontology approach. Correlation rules should be derived from relevant use cases and can then be modelled in the described way. As depicted in figure 3, rules can be continuously improved by analyzing user interaction in the search ad delivery environment. This can be done, for example, by statistical web analytics, AI-driven pattern recognition, or direct user feedback.

SCR can initiate as a starting point more elaborated semantic modelling approaches going beyond content correlations for delivery only. But even within the simple SCR framework given, rules can be used with respect to other goals: For example, dependencies within objects or metadata arising from product configurations can be modeled to some extent. Other types of search applications, like document-based information retrieval systems, can use SCR formalism provided that objects are specified and retrieved by metadata. Also, in classical publishing environments for technical communication, linking between topics might be managed and processed with the help of SCR for all types of media. Similar approaches can be found as relationship tables in DITA environments. SCR can extend this to a more general class-to-class linking, using a semantic modelling formalism. Two aspects must be emphasized on, as they are crucial for future applications. First, the performance through preprocessing or dynamic processing of SCR in corresponding implementations. Second, the maintenance of SCR through updating and versioning. These two aspects are left to further work and publications on SCR.

In summary, SCR can be understood, at the introduced level, as a logic layer between content creation and content delivery, allowing to manage correlations between information objects independent of content creation. microDocs are one of the beneficiaries recently being discussed. Other industrial applications and

processes can define corresponding object correlations, not necessarily bound to content objects exclusively.

We acknowledge the support by system vendors implementing and prototyping SCR functionalities given in [2]. The following companies supported the subsequent research project with master's students at Karlsruhe University of Applied Sciences (HSKA): Expert Communication Systems, I-Views/Empolis, Klarso, Ontolis, c-rex.net. The underlying research was supported as sabbatical grant by HSKA.

References

1. Ziegler W. *Extending intelligent content delivery in technical communication by semantics: microdocuments and content services*, Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2020) <https://doi.org/10.1051/shsconf/20207703009>
2. W. Ziegler, *Rules-Based Content Correlations* tcworld online conference, Germany, (2020)
3. W. Ziegler, *Content management systems tested: Introducing the PI-Fan*, tcworld e-magazine 10/2016
4. *iIRDS – The International Standard for intelligent information Request and Delivery*, <https://iirds.org/material-downloads/iirds-version-1-0-1/>
5. W3C, *RDF Schema 1.1*, <https://www.w3.org/TR/rdf-schema/> (2014)
6. W3C, *Web Ontology Language (OWL)*, <https://www.w3.org/OWL/> (2012)
7. I4ICM, SCR 1.1.0, <https://www.i4icm.de/downloads/scr/>
8. W3C, *Shapes Constraint Language (SHACL)*, <https://www.w3.org/TR/shacl/>
9. Klarso, *Semantic software solutions*, <https://klarso.com/>
10. EC-Systems, *The Smart Media Creator*, <https://www.ec-systems.de/>
11. Ontolis, *Think It – Model It*, <https://www.ontolis.com/>
12. I-views, *The Smart Data Engine*, <https://i-views.com/en/>
13. c-rex.net, *Digital Transformation Service*, <https://c-rex.net/>