# Linguistic Approach to Semantic Correlation Rules

*Charlotte* Effenberger[1]*

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

**Abstract.** As communication between humans and machines in natural language still seems essential, especially for end users, Natural Language Processing (NLP) methods are used to classify and interpret this. NLP, as a technology, combines grammatical, semantical, and pragmatical analyses with statistics or machine learning to make language logically understandable by machines and to allow new interpretations of data in contrast to predefined logical structures. Some NLP methods do not go far beyond a retrieving of the indexation of content. Therefore, indexation is considered as a very simple linguistic approach. Semantic correlation rules offer the possibility to retrieve easy semantic relations without a special tool by using a set of predefined rules. Therefore, this paper aims to examine, to which extend Semantic Correlation Rules (SCRs) will be able to retrieve linguistic semantic relations and to what extend a simple NLP method can be set up to allow further interpretation of data. In order to do so, an easy linguistic model was modelled by an indexation that is enriched with semantical relations to give data more context. These semantic relations were then queried by SCRs to set up an NLP method.

## 1 Introduction

The communication between humans and machines is essential. Even if more people learn how to input instructions, communicating in a natural language still seems key, especially for end users, particularly in the area natural language processing, for example, using speech recognition, searches, etc. Therefore, for many companies, the question arises on how to combine product and documentation. Especially for products like smart homes, communication with and about products completely changes. Here, natural language seems to be worth the research.

As language is dynamic and offers a broad variety of interpretation, special techniques have to be applied to make language understandable for machines. [1] Therefore, the understanding of natural language concentrates mostly on illustrating logical relations within natural language. Methods to classify and understand natural language are called Natural Language Understanding (NLU). NLU can enable computer-human-interactions by Natural Language Processing (NLP). NLP with NLU can then serve to understand and interpret natural language. Subsequently, in the following, only the term NLP is used, as NLU is a subset of NLP. [2]

## 2 Interpreting language

In the following, it will be analysed what methods can be used to tackle challenges in understanding and interpreting natural language.

### 2.1 NLP methods

NLP methods vary in their width and depth. Methods with a high width are, for instance, parsers, that are able to analyse sentences for their grammatical structures from any type and domain. They can be seen as a first step of linguistic division. Another broadly used method would be an indexation of documents where documents can be found by the indexed key words. Both methods cannot achieve more than exactly the expected and requested results. However, methods with high depth are use cases where only sentences with a very specific structure and wording can be processed with results allowing for interpretations that are usable for complex operations. The goal would then be to make deep methods broader and broad methods deeper. [1]

### 2.2 Semantics

The goal of this paper is to deepen a broad NLP method and provide better results for making it usable for complex operations.

One way to do so is by adding semantics to the data in a knowledge base. As knowledge can only be as smart as its integration in the bigger picture, data, itself, cannot necessarily be used for making assumptions and interpretations according to certain conditions. Data has to be brought into an interpretable context. Therefore, data and relations between data are semantically described which enriches them by context and meaning, such as that semantics help to turn data into explicit machine-readable knowledge. [3] Conceptualization like this can be specified by ontologies. An ontology can

\* Email: efch1012@hs-karlsruhe.de; charlotte.effenberger@gmail.com

be seen as a description of the concepts and their relationships. [4]

Semantic relations originally came up by W3O to create a semantic web by structured, machine-readable data with interpretable relations in RDF Schema (RDFS). Both RDF and RDFS expressions are a collection of triples, which means, that they contain a subject, an object, and a predicate that shows the relation between subject and object. [5] Besides RDFS, OWL, as little richer language, can be used to describe semantic relationships. [6]

Not only semantic web can profit by bringing data into a net structure but also technical communication and any kind of Content Management System (CMS). Semantic relations allow a query to jump from node to node using graph patterns to gain not only requested data, but also related data that was not explicitly asked for in a certain query. This creates dynamic and always up-to-date results [7] [8]. Data and semantic relations can be visualized in knowledge graphs. [9]

## 2.3 Semantic Correlation Rules

Semantic Correlation Rules (SCRs) offer the possibility to create semantic relations between a primary content object and a contextually required secondary object by using semantically described relations expressed with RDFS and OWL. This way, SCRs are able to find the "nearest-neighbour" data that is logically connected to the first data. Technically, SCRs use a combination of so called InRules and OutRules to request already existing metadata correlated to the content objects. In the following, a defined InRule/OutRule combination is called SCR.

As basic version of SCRs, InRules can be implemented with "select" relations. In this case, the InRule exactly selects the given parameters. For the full definition of SCRs, "equals" relations are included for more generic use cases to allow a dynamic selection of variable parameters.

As SCRs are implemented with RDFS/OWL, they work system-independently as long as system functionalities support semantic relations expressed with RDFS/OWL. In technical communication, they can be implemented into CMS to then be referred to content objects from the CMS as a source. This way, SCRs can be described as a light-weight ontology that can offer advantages of semantic relationships but without requiring much effort, preparation, or resources [10].

Therefore, this paper aims to examine, to which extend SCRs will be able to illustrate semantic relations from a linguistic context.

## 3 Building a linguistic model: indexation with ontology

In order to begin with the research on natural language, a linguistic model with semantic relations needed to be set up in order to be queried later by SCRs.

To build up a linguistic model, a variety of whole documents were indexed with keywords.

These keywords were brought into a linguistic taxonomic ontology with semantically described relations. Taxonomic means, that the classification was built hierarchically.
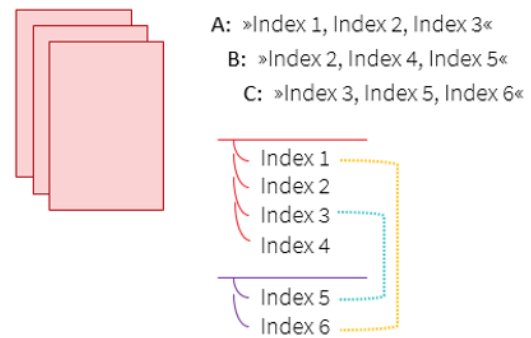


**Fig. 1.** Document A, B, and C were indexed with keywords being put into a taxonomic ontology
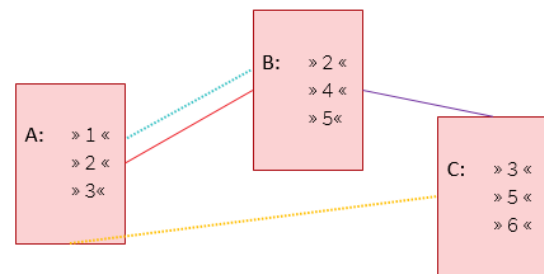


**Fig. 2.** Net structure coming from ontological relations

In figures 1 and 2, there is a red group with an index of 1, 2, 3, and 4, and a purple group with an index of 5 and 6.

There are two non-hierarchical relations from 3 to 5 and from 1 to 6, which forms a net structure for the documents.

This paper researches on whether a linguistic approach, like an indexation, can be supported with SCRs to allow further linguistic interpretation of data. If this is the case, the described method can be seen as a simple NLP.

## 3.1 Linguistic research on smart technologies

As an example use case for testing functionalities of SCRs, a knowledge network, was built on technical communication of smart technologies for end users. Within the subject of smart technologies, the knowledge network focusses on giving information about the selection of smart home systems, which includes mainly the field of applications of the smart home system, the connectivity, the measured physical parameters, etc. [11].

Within the knowledge network of smart technologies, a certain context was enriched with an own linguistic ontology and NLP functionalities to provide a linguistic benefit for the network. As a use case of the linguistic ontology, it was determined to focus on finding out more about one specific field of application, which is climate monitoring. It was considered, what similarities can be found in possible search terms of search requests. From a personal

research, searching for the field of application climate monitoring brought similar results compared with searching for devices to implement climate monitoring or searching for sensors that measure parameters like temperature, humidity, carbon monoxide, smoke, etc. Therefore, the combination of a field of application with actuator devices and physically measured parameters from the knowledge network turned out to be extremely interesting to be linguistically extended in an ontology.

As a net structure cannot be built by only 1 relation, the one semantic relation in the knowledge network "measures" had to be added with more non-hierarchical,

semantical relations. Thus, two more relations were needed between all 3 nodes of the knowledge network in order to get connected search results. The two additionally needed relations were climate monitoring "is set by" physical parameters and climate monitoring "is implemented with" devices.

In the linguistic approach of the smart technologies' knowledge network, sensor devices as logical layers were removed as it seemed to be more trivial if actuator devices, like thermostats, are measuring parameter directly.
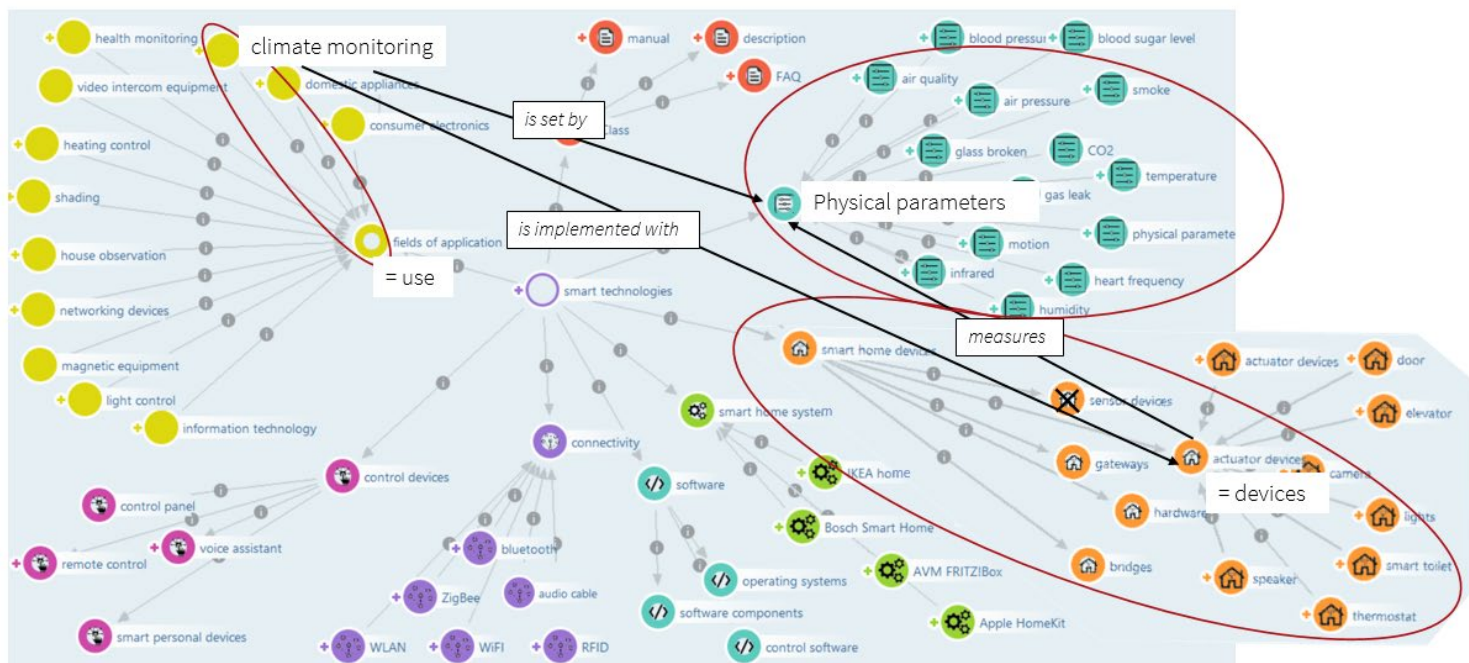


**Fig. 3.** Connecting 3 parts of the knowledge network to build a linguistic ontology (i-views [11])

### 3.2 Keywords from smart technologies

The semantic linguistic ontology was set up by basically extending the end points "climate monitoring", "physical parameters", and "devices" of the knowledge network smart technologies with the actual keywords.

The most used key search terms for climate monitoring that were found in the personal research, were used to build the ontology. These are the same words, that are later used to index the documents as they result in a search query for the documents.

For "climate monitoring", the following keywords covered climate, heating, cooling, heat protection, frost protection, heating ventilation air conditioning, HVAC, air quality, save energy costs, and saving energy.

For "devices", the keywords included smart thermostat, heating panel, heating control panel, radiator, radiator thermostat, blind management, ventilation, air conditioning, air purifier, blinds, door window contact, door management, windows, and window management.

For "physical parameters", the keywords consisted of temperature, change temperature, monitor

temperature, manage temperature, humidity, dust, pollen, gas, CO, carbon monoxide, fire, and smoke.

The keywords can be called associations as they are logically connected to each other and partly synonymously used as they consist of the same underlying meaning. For all keywords from the linguistic model, the overall meaning is, that they belong to a smart home which can monitor climate. Therefore, the keywords can be seen as linguistic utterances. Utterances in this context is a word, that is applied from NLP techniques, like chatbots, and it means words itself, as a linguistic form, can be separated from the underlying meaning. In NLP techniques, words are separated from their meaning to ensure, that users find what they are looking for, even if they have not used the exact same words. In addition, each group then contains of a further common meaning. For example, the meaning of words from the group devices, is, that they are devices used to implement climate monitoring. Within each hierarchical group, there can be words that are even more similar and can almost be seen as synonyms, for example, smart thermostat and heating panel. Logically, it might be correct, to clearly distinguish them into more subgroups. However, this was not done, as one topic of research should also be to

investigate, whether distinction can be done by SCRs with meaningful relations, so that more subgroups may not be needed at all.

### 3.3 Ontology and indexation in CMS

The next step was to bring the theoretical linguistic model into a corresponding system, which is able to handle semantic relations. Semantic relations can be built up in CMS in various forms. For this topic of research, the system klarso term:studio by klarso GmbH was used. This system seemed suitable for this use case, as it automatically sets up relations between newly created data entries: The system works as knowledge network structures. From naturally established relations, the required relations can easily be semantically described, queried, and used. Within the knowledge network in term:studio, more methods can be integrated such as (linguistic) ontologies, identity and access management, NLP methods like grammar parsing, concept maps, etc. [12].

Within term:studio, for climate monitoring, devices, and physical parameters, a so-called concept group was defined which includes all indexing keywords.



**Fig. 4.** Concept group device with hierarchical keywords

Next, the keywords needed to be connected with each other according to the relations in fig. 3: Climate monitoring as a field of application is connected to devices and to measured physical parameters. As an example, fig. 5 shows the semantic relations from the keyword "heating" (climate monitoring) to corresponding device keywords by establishing the associative relation "is implemented with". The column "Type" means a relation type and shows a semantic description.



**Fig. 5.** Relation climate monitoring to devices

After, the keywords were applied to documents. As an example, document with the IDs 8, 37, and 63 were tagged with the corresponding keywords by the semantic relation "rel_document_device" from document to device.



**Fig. 6.** Relation between document and device



**Fig. 7.** Overview of keywords with relations

## 4 Setting up NLPs with SCRs in the CMS system

Subsequently, the relations that were built, were interpreted with SCRs. Again, the goal is now to find results with SCRs that are not directly correlated with one search term or a content, but to gain data that was not explicitly asked for to receive NLP functionalities by it.

### 4.1 Expected results

In this article, two different ways of NLP are considered to be enabled with SCRs: Associations and subject.

Before, it was mentioned, that words from the same group are associated with each other. Therefore, one NLP method is to connect one word with the other words from the same content group for finding associations. For example, if one word from the group devices is requested, not only the requested word/device is offered, but all devices that allow climate monitoring, just in case users meant another word or technique to monitor their room climate. Furthermore, when indicating an interest for "smart thermostat", then the other devices are also offered.

It was also mentioned that all keywords from the ontology have one common meaning and the same subject that they belong to, in this case, a smart home, that can monitor the climate (subject climate monitoring). Therefore, the other NLP method was to connect one keyword with all other keywords from the ontology for finding the same subject. As this process might be overwhelming, keywords can give a certain strength to prioritize them. For example, those keywords, which are already given in the search request, can be prioritized.

## 4.2 Setting up the SCRs

For the association, one word from a content group was defined as InRule. As Outrules, all other words from the same group were defined.

| | IN-Select | | InR | | OUT-Rules | | | OUT-Select |
|---|---|---|---|---|---|---|---|---|
| 1 | heating | 1 | InR:heating | 1 | OutR:cooling | | 1 | cooling |
| | | | | 2 | OutR:heating ventilation ai... | | | |
| | | | | 3 | OutR:HVAC | | | |

**Fig. 8.** SCRs for associations with example "heating"

In this example, one InRule is created for one keyword "heating". The selected keywords are shown in the column IN-Select. This InRule (InR) is connected to as many OutRules as there are other keywords in the same content group. This is because they all contain exactly one word, as we can see for the selected OutRule from "cooling": The selected Out-Objects (OUT-Select) contain only "cooling".

For the associations, each OutRule only contains one word as the connection between the keywords should be "OR". Thus, for heating, the InRule/OutRule combination (SCR) should request all content with "cooling" OR "heating ventilation air conditioning" OR "HVAC", etc. If the keywords were provided altogether in one OutRule, the connection between them would be "AND". Therefore, only the content containing all of the given keywords would be provided, which is not the expected result.

For the subject use case, again, one keyword is defined in one InRule. For OutRules. All other words from the ontology were defined.

| | IN-Select | | InR | | OUT-Rules | | ^ | | OUT-Select |
|---|---|---|---|---|---|---|---|---|---|
| 1 | heating | 1 | InR:heating | 37 | OutR:heating | 10 | | 1 | heating |
| | | | | 1 | OutR:cooling | 5 | | | |
| | | | | 2 | OutR:heating ventilation ai... | 5 | | | |

**Fig. 9.** SCRs for climate monitoring with prioritized given word "heating"

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 11 | OutR:smart thermostat | 5 | |
| | | | 12 | OutR:heating panel | 5 | |

**Fig. 10.** OutRules were defined for all words from all concept groups

As shown in fig. 9, "heating" is prioritized with 10 while all other words from the ontology are only prioritized with 5. Fig. 10 shows, that all other words from the ontology are included as all words from all concept groups, and are selected in the OutRules, for example, also for devices.

## 4.3 Draft Content Delivery Portal (CDP) with implemented NLP method

In the CDP, the associations are shown in the "Did you mean this?" section. In this example, only associations for climate monitoring should be shown. Hence, the SCR described above, searches content described with the given keywords and shows the first 4 results in 4 tiles. As a result, users receive content, which is somehow connected to a field of application monitoring climate.

Showing all content from climate monitoring, but prioritizing already given content, is more difficult. For this, the section "Related links about your subject:" can be created with 3 tiles.

One idea is to create a SCR, like the one shown above, for every word from the ontology and connect all SCRs from one concept group with one tile. Thereby, one related content can be shown from each concept group. If there are too many related content or concept groups, tiles can be extended. Finding more information about a topic is a more generic use case for SCRs, that does not require exact selected words. Therefore, the "select" relation in this context could be replaced by an "equals" relation, which may provide better results. Therefore, instead of selecting exactly one word in the InRule, the InRule could select the keyword that equals already given metadata from content given as words, or chosen as content by the user.
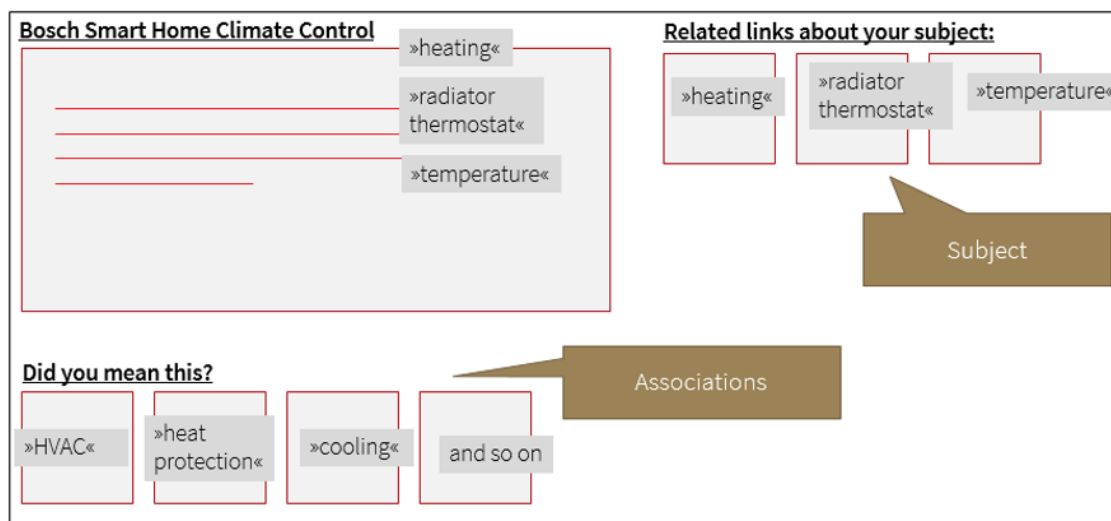
**Fig. 11.** Draft CDP with implemented SCRs

## 5 Outlook and conclusion

As the SCRs so far could not be tested in a real CDP system, the next step would be to export SCRs and linguistic ontology from term:studio to merge them to the already existing knowledge network and to import them to a CDP.

However, as a conclusion, it can be said, that SCRs in theory can provide an easy form of NLP for a finite number of words and predefined connection between documents and words. The described method can be seen as NLP, as SCRs offers the possibility to connect content to words and allow an out-of-the-box interpretation. In doing so, not only requested, but also not explicitly requested words and their content can be offered. As an example, SCRs can provide content connected to keywords that are always forgotten, or that can clearly help in certain circumstances. Additionally, not only words can become In-Objects, but also documents can be In-Objects, then it is possible to connect documents to associations that might have been forgotten otherwise.

However, every system is only as strong as its knowledge network and as smart as its logic. For the very open and imprecise use cases investigated in this article, SCRs can eventually provide a solution for this kind of NLP method, as well as many other techniques. However, it depends on the setup of data. Effective systems can define any relation, that are to be defined, ultimately it is logic, so the logic is fundamental.

For this topic of research, the logic could have been set up differently. Future studies can look at attributes given to words belonging to a concept group, which allows to call them all by one, instead of selecting all words one by one. In the same way, an attribute called "climate monitoring" could be defined for all words from the ontology, in order to call them much more dynamically.

## References

1. P. Liang, *Natural Language Understanding: Foundations and State-of-the-Art*, https://www.youtube.com/watch?v=mhHfnhh-pB4 (2017)

2. C. Dilmegani, *Natural Language Understanding in 2021: in-Depth Guide*, https://research.aimultiple.com/nlu/ (2021)

3. K. Munk, technische Kommunikation **6**, 27 (2019)

4. T. Gruber, *What is an Ontology,* http://www-ksl.stanford.edu/kst/what-is-an-ontology.html (1992)

5. W3C, *RDF Schema 1.1*, https://www.w3.org/TR/rdf-schema/ (2014)

6. Cambridge Semantics, *RDFS vs. Owl,* https://www.cambridgesemantics.com/blog/semantic-university/learn-owl-rdfs/rdfs-vs-owl/

7. U. Parson, tekom Schriften zur Technischen Kommunikation **22***, 26 (2017)

8. W3C, *SPARQL 1.1 Query Language*, https://www.w3.org/TR/sparql11-query/ (2013)

9. I-views, *About i-views,* https://i-views.com/en/smart-data-engine-2/about-i-views/

10. W. Ziegler, *Semantic Correlation Rules as a Logic Layer between Content Management and Content Delivery,* in Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2021)

11. D. Baro, L. Neundörfer, *Contrasting Ontology Modeling with Correlation Rules for Delivery Applications,* in Proceedings of the ETLTC ACM Chapter International Conference. Aizuwakamatsu, Japan. (2021)

12. Klarso, *Terminology Management with term:studio* https://klarso.com/terminology-management/