Decomposition of post-coordinated SNOMED CT expressions for storage as HL7 FHIR resources (PCEtoFHIR)

Tessa Ohlsen, Cora Drenkhahn, Josef Ingenerf

Submitted to: JMIR Medical Informatics
on: February 28, 2024

Table of Contents

Decomposition of post-coordinated SNOMED CT expressions for storage as HL7 FHIR resources (PCEtoFHIR)

Tessa Ohlsen[1, 2] BSci, MSci; Cora Drenkhahn[2] BSci, MSci; Josef Ingenerf[1, 2] Dr rer nat, Prof Dr

[1]IT Center for Clinical Research University of Luebeck Luebeck DE
[2]Institute of Medical Informatics University of Luebeck Luebeck DE

Corresponding Author:
Tessa Ohlsen BSci, MSci
IT Center for Clinical Research
University of Luebeck
Ratzeburger Allee 160
Luebeck
DE

Abstract

Background: To guarantee interoperability, structural and semantic standards must be adhered to. For exchanging medical data between information systems, the structural standard FHIR (Fast Healthcare Interoperability Resource) is recently gaining traction. In terms of semantic interoperability, the reference terminology SNOMED CT as a semantic standard enables a post-coordination in comparison to many other vocabularies. These post-coordinated expressions (PCEs) result in SNOMED CT being an expressive and flexible interlingua, which enables precise coding of medical facts, but at the expense of increased complexity and challenges with storage and processing. In addition, the boundary between the scope of semantic (terminology) and structural (information model) standards blurs, the so-called TermInfo problem.

Objective: Although often discussed critically, the TermInfo overlap can also be explored for its beneficial potential by enabling a flexible transformation of parts of the PCEs. In this paper, an alternative solution for the storage of PCEs is presented, i. e., in combination with the FHIR data model. In the end, all components of a PCE should be expressible exclusively by pre-coordinated concepts that are linked to suitable elements of the information model.

Methods: The approach is based on storing PCEs and/or parts of them conforming to FHIR resources. Using the Web Ontology Language for generating an OWL ClassExpression in combination with an external reasoner and semantic similarity measures, a pre-coordinated SNOMED CT concept that describes the PCE most precisely is determined as a superconcept. In addition, the non-matching attribute relationships between the superconcept and the PCE are determined as the so-called delta. Once SNOMED CT attributes have been mapped to FHIR elements manually, FHIRPath expressions can be determined for the superconcept and the delta, which in turn enable the identified pre-coordinated codes to be stored in FHIR resources.

Results: A web application called PCEtoFHIR was developed to implement this approach. In a validation with 600 randomly selected pre-coordinated concepts, the correctness of the generated OWL ClassExpression could be confirmed. Additionally, two different approaches for calculating semantic similarities were validated, with the approach by Sanches et al. demonstrably yielding more precise results. In a validation of the entire approach, considering 33 already existing PCEs, the correct functionality could also be demonstrated.

Conclusions: PCEtoFHIR provides services to decompose PCEs for storing them in FHIR resources. When creating structure mappings concerning certain subdomains of SNOMED CT concepts (e.g. allergies) to desired FHIR profiles, the use of SNOMED CT Expression Templates has proven to be very useful. Domain experts can prepare templates with suitable mappings that can be reused in a constrained way by end users more easily.

(JMIR Preprints 28/02/2024:57853)
DOI: https://doi.org/10.2196/preprints.57853

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Decomposition of post-coordinated SNOMED CT expressions for storage as HL7 FHIR resources (PCEtoFHIR)

Abstract

Background: To guarantee interoperability, structural and semantic standards must be adhered to. For exchanging medical data between information systems, the structural standard FHIR (Fast Healthcare Interoperability Resource) is recently gaining traction. In terms of semantic interoperability, the reference terminology SNOMED CT as a semantic standard enables a post-coordination in comparison to many other vocabularies. These post-coordinated expressions (PCEs) result in SNOMED CT being an expressive and flexible interlingua, which enables precise coding of medical facts, but at the expense of increased complexity and challenges with storage and processing. In addition, the boundary between the scope of semantic (terminology) and structural (information model) standards blurs, the so-called TermInfo problem.

Objective: Although often discussed critically, the TermInfo overlap can also be explored for its beneficial potential by enabling a flexible transformation of parts of the PCEs. In this paper, an alternative solution for the storage of PCEs is presented, i. e., in combination with the FHIR data model. In the end, all components of a PCE should be expressible exclusively by pre-coordinated concepts that are linked to suitable elements of the information model.

Methods: The approach is based on storing PCEs and/or parts of them conforming to FHIR resources. Using the Web Ontology Language for generating an OWL ClassExpression in combination with an external reasoner and semantic similarity measures, a pre-coordinated SNOMED CT concept that describes the PCE most precisely is determined as a superconcept. In addition, the non-matching attribute relationships between the superconcept and the PCE are determined as the so-called delta. Once SNOMED CT attributes have been mapped to FHIR elements manually, FHIRPath expressions can be determined for the superconcept and the delta, which in turn enable the identified pre-coordinated codes to be stored in FHIR resources.

Result: A web application called PCEtoFHIR was developed to implement this approach. In a validation with 600 randomly selected pre-coordinated concepts, the correctness of the generated OWL ClassExpression could be confirmed. Additionally, two different approaches for calculating semantic similarities were validated, with the approach by Sanches et al. demonstrably yielding more precise results. In a validation of the entire approach, considering 33 already existing PCEs, the correct functionality could also be demonstrated.

Conclusion: PCEtoFHIR provides services to decompose PCEs for storing them in FHIR resources. When creating structure mappings concerning certain subdomains of SNOMED CT concepts (e.g. allergies) to desired FHIR profiles, the use of SNOMED CT Expression Templates has proven to be very useful. Domain experts can prepare templates with suitable mappings that can be reused in a constrained way by end users more easily.

Keywords: SNOMED CT; HL7 FHIR; TermInfo; post-coordination; semantic interoperability; terminology; OWL; semantic similarity

Introduction

Background

The increasing digitization of medical records has resulted in a larger amount of patient data available for analysis in healthcare. There is a need to utilize this data to enhance medical care and provide more targeted treatments. However, to achieve this, it is necessary to have the ability to

exchange and process data automatically between different systems. This not only requires technical compatibility, but also semantic interoperability, which ensures that the meaning of the data is retained when importing it into another computer system. The ability to transfer and work with data across systems is essential for leveraging the full potential of digital medical records and enhancing patient care [1].
To guarantee semantic interoperability, structural and semantic standards must be adhered to. Structural standards define the syntax for accessing data fields in information models. In recent years, the newly developed HL7 standard FHIR (Fast Healthcare Interoperability Resources) is gaining international popularity in this regard due to its focus on simplified implementation and up-to-date technologies [2]. Semantic standards, on the other hand, involve terminologies with language-independent codes to represent the data's meaning in an interoperable way. Here, SNOMED CT is known as the most expressive terminology in medicine that is used to improve semantic interoperability [3]. In 2021, Germany obtained a national license for SNOMED CT leading to increased interest and usage of the terminology. SNOMED CT concepts are being incorporated into data modeling, such as the Medical Information Objects of the German National Association of Statutory Health Insurance Physicians (NASHIP) [4] and the core data set of the Medical Informatics Initiative (MII) [5]. The use of SNOMED CT – over 350.000 concepts exist here – is to provide a machine-usable interlingua that minimizes coding issues specific to countries and medical fields. Due to the complexity of natural language, not all medical circumstances can be precisely coded using SNOMED CT's large number of pre-coordinated concepts. To avoid a rapid increase in the number of new concepts, and in contrast to many other vocabularies, SNOMED CT enables post-coordination, which allows the combination of pre-coordinated concepts into new expressions using a formal grammar.

While being of great value for the precise recording of medical facts, SNOMED CT‘s ability for post-coordination further blurs the line between the scope of structural and semantic standards. This so-called TermInfo problem [3, 6, 7] largely emerged due to the independent development of these standards leading to overlaps and inconsistencies in data representation. Thus, there is often more than one possibility how to record a medical idea, employing either the semantic or the structural side more heavily. For example, "Fracture of the left femur" can either be represented by using a single composite post-coordinated expression (PCE) as the FHIR element *Condition.code* like

- *Condition.code*: *71620000*:{*363698007=722738000}* *(Fracture of femur : {Finding site = Structure of bone of left femur})*

or by splitting the semantic meaning up into two precoordinated SNOMED CT concepts using further element-code-combinations of the FHIR resource *Condition*:

- *Condition.code*: *71620000 |Fracture of femur (disorder)|*
- *Condition.bodySite: 722738000 |Bone structure of left femur (body structure)|*.

The variability in the presentation of facts increases when using interlinked FHIR resources. For example, "Allergic reaction caused by pistachio nut" can either be represented by using a PCE with *419452009 |Allergic reaction caused by food (disorder)|* and *246075003 |Causative agent| = 227512001 |Pistachio nut|* as a refinement:

- *Condition.code: 419452009:{246075003=227512001}*

or the specific nut type is expressed via the FHIR information model, i.e. via the referenced resource *AllergyIntolerance*.

- *Condition.code: 419452009 |Allergic reaction caused by food (disorder)|*
- *AllergyIntolerance.code: 227512001 |Pistachio nut|*.

Motivation

As already mentioned, post-coordination is a unique feature that largely increases SNOMED CT's expressive power. However, the use of post-coordination is currently lagging due to various difficulties. Some of these hurdles have already been addressed [8], yet the storage of post-coordinated SNOMED CT expressions (PCEs) remains challenging due to several reasons:

1. *Adherence to familiar data structures*
   The medical circumstances are typically described using individual codes, and there are established methods for storing and processing these. Restrictions such as length-limited data types can be tolerated when using simple codes, but not with arbitrarily large formal expressions. Because of this, there are concerns about the use of PCEs [9].
2. *Lack of technical support for the evaluation*
   In most cases, the evaluation of PCEs requires a terminology server, such as the CSIRO Ontoserver [10], which is the only terminology server supporting post-coordination as of now [11]. Furthermore, the evaluation of PCEs requires the use of FHIR services, such as determining subtype relations or using the Expression Constraint Language (ECL) of SNOMED CT or the use of the Web Ontology Language (OWL) and a reasoner.
3. *Difficulties with FHIR Search*
   When searching for information in FHIR resources, post-coordination is only supported if a PCE is explicitly defined in a FHIR CodeSystem Supplement [12].

One option is to store PCEs using a FHIR CodeSystem Supplement [12] or a “post-coordinated expression library” [13]. Although the PCEs can be collected in a FHIR CodeSystem Supplement, it does not allow storage in electronic patient records.

In this paper, an alternative solution for the storage of post-coordinated SNOMED CT expressions will be presented, which is based exemplary on storage as FHIR resources. The approach can also be transferred to other information models, such as openEHR or relational databases. For selected FHIR resources, such as *Condition* and *Observation*, a mapping of some SNOMED CT attributes to FHIR is currently provided by HL7 International (see [13] and [14]). In addition, the Confluence pages “SNOMED on FHIR” [15] show various options for mapping SNOMED CT attributes to FHIR while avoiding semantic overlaps. This information is considered here.

Related work

While an increasing number of publications address the post-coordination of SNOMED CT concepts [8, 16, 17, 18, 19], a literature review revealed that publications on the storage of post-coordinated SNOMED CT expressions are very scarce. In a previous project, the authors developed WASP, a web application that supports the guided creation of PCEs based on the Concept Model and the Compositional Grammar and enables the storage of the resulting PCE in a CodeSystem Supplement [8].

There are some papers addressing the use of SNOMED CT in combination with standardized information models based on HL7 Reference Information Model (RIM), HL7 Clinical Document Architecture (CDA), HL7 FHIR resources. A project by Perez-Rey et al. [20] is generally concerned with linking the normal form of pre-coordinated SNOMED CT concept definitions. SNOMED CT concepts are normalized and then bound to the HL7 RIM classes. In general, this approach could be extended to PCEs. However, the HL7 v3 standard has not been widely adopted due to its complexity [1]. A project by Arguello-Casteleito et al. [21] deals with a mapping of pre-coordinated SNOMED CT concepts or post-coordinated SNOMED CT expressions from Consolidated Clinical Document Architecture (C-CDA) to FHIR resources. The objective of this approach resembles the approach taken in this work. However, the approach by Arguello-Casteleito et al. intensely focuses on an ontology. In contrast, this work primarily utilizes the widely adopted FHIR and SNOMED CT native specifications [1]. Furthermore, the specific ocular diseases used by Arguello-Casteleito et al. only pertain to a very specific subset of SNOMED CT expressions. Additionally, the publication does not provide data regarding to what extent all information included in the PCEs could be transferred into the FHIR representation.

Methods

Overview

The aim of this work is to develop and implement an approach that enables the storage of a SNOMED CT post-coordinated expression (PCE) within FHIR resources by using only pre-coordinated codes. For this alternative representation, the PCE shall be decomposed into pre-coordinated concepts, which can then be stored in suitable elements of matching FHIR resources. An overview of the envisioned approach is shown in Figure 1.

A PCE, which is first checked for syntactic and semantic correctness, serves as input. This can be classified within SNOMED CT by using OWL and a reasoner, whereby the direct superconcepts can be determined. Of these concepts, the most similar concept to the PCE is determined (superconcept). The delta can then be calculated between the superconcept and the PCE, which includes all the information of the PCE that the superconcept does not represent. In the final step, appropriate elements of matching FHIR resources must be identified to store the information of the superconcept and the delta. Therefore, FHIR StructureMaps defining these associations on a general level need to be created beforehand.

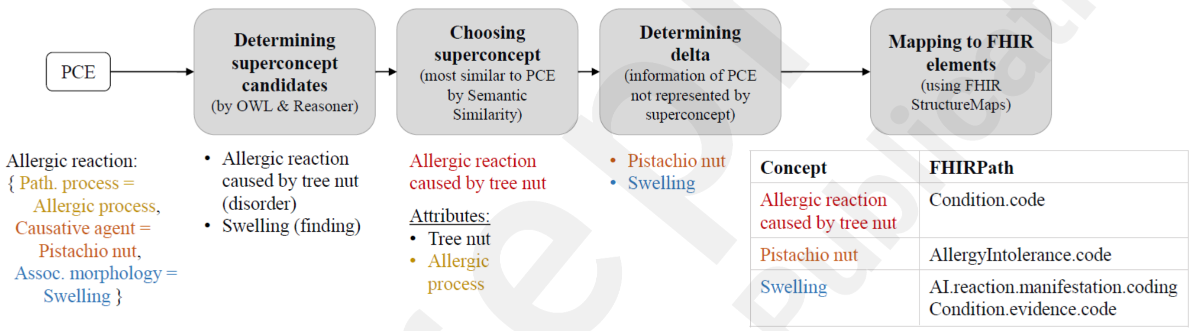


Figure 1. The decomposition of a PCE into elements of (profiled) FHIR resources consists of four steps.

Validation of PCE

To ensure that the flexible PCEs can be interpreted and evaluated, they must comply with the syntactic requirements of the Compositional Grammar and the semantic rules of the Concept Model. In the first step, the input PCE is checked for syntactic and semantic correctness. The HL7 FHIR services *$validate-code* and Ontoserver are used for this. Ontoserver is a FHIR-based terminology server provided by the Australian company CSIRO. Ontoserver supports and facilitates work with important coding systems such as SNOMED CT or LOINC [10]. Only if a PCE is syntactically and semantically correct, then it will automatically continue with the next steps.

Determining the superconcept candidates

The Web Ontology Language is used to determine the superconcepts. The Web Ontology Language (OWL) is an exchange format for ontologies. By using OWL, complex knowledge about concepts and relationships between concepts can be represented [22, 23]. SNOMED CT can also be represented as an OWL ontology, and individual concepts of SNOMED CT can be represented as OWL expressions. For example, in the SNOMED CT Browser and in the monthly release package of SNOMED CT, the concept definition for each SNOMED CT concept is shown as an OWL expression in addition to other formats [24].

To obtain SNOMED CT as an OWL ontology, the SNOMED OWL Toolkit [25] was used on the

Release Format 2 of the International Edition, 2022-04-30 of SNOMED CT [26]. The generated ontology contains, among other things, the OWL ClassExpression in the form of the functional syntax for the individual pre-coordinated SNOMED CT concepts.

PCEs on the other hand are based on the syntax of the Compositional Grammar. Thus, a PCE needs to be transformed into an OWL ClassExpression for further processing. For each component of the PCE the respective OWL counterpart is determined according to the existing concept definitions and as shown in Table 1. An algorithm is developed to automatically carry out the transformation.The result of the algorithm is an OWL ClassExpression that is structured in the same way as the OWL ClassExpression created by SNOMED International for the definitions of SNOMED CT concepts. An OWL expression for a PCE is shown in Figure 2. For better readability, the Fully Specified Names are also shown in addition to the SNOMED CT Identifiers of the SNOMED CT concepts, although normally only the SNOMED CT Identifiers are used. The symbol ":" is a placeholder for the defined namespace, here for "*http://snomed.info/id/*" [24].

Table 1. Various OWL constructs are used for the representation of the components of a PCE.

PCE component	**OWL construct**
SNOMED CT concept	OWL Class
Linking of ungrouped attribute relationships	OWL ObjectIntersectionOf
Linking between individual attribute relationships in a Role Group	
Linking between focus concept and all attribute relationships of ungrouped attributes and Role Groups	
Attribute relationship of attribute and SNOMED CT Identifier as attribute value	OWL ObjectSomeValuesFrom
All grouped attribute relationships and Role Group Identifier	
Attribute relationship of attribute and concrete value as attribute value	OWL DataHasValue
SNOMED CT attribute, attribute value: SNOMED CT concept	OWL ObjectPropertyExpression
SNOMED CT attribute, attribute value: concrete value	OWL DataPropertyExpression
Concrete value as attribute value	OWL Literal

In the next step, the previously formed OWL ontology of SNOMED CT, the created PCE-specific OWL ClassExpression and a reasoner are used to classify the PCE into the existing SNOMED CT hierarchy. By using a reasoner, new knowledge can be generated through logical inferences from existing contents of an ontology, such as the determination of superrelationships between concepts [22]. The reasoner ELK, which was developed by the University of Oxford and the University of Ulm and is used in this work, is also pretty much the only reasoner that can process SNOMED CT at all due to the size of SNOMED CT [27].

The direct superconcepts of PCEs are determined via the reasoner ELK, leading to at least one OWL Class corresponding to a pre-coordinated SNOMED CT concept. All identified pre-coordinated SNOMED CT concepts are possible concepts for the superconcept.

PCE

```
419076005 |Allergic reaction| :
{ 370135005 |Pathological process| =
    472964009 |Allergic process| ,
  246075003 |Causative agent| =
    227512001 |Pistachio nut| ,
  116676008 |Associated morphology| =
    442672001 |Swelling|
}
```

Associated OWL Class Expression

```
ObjectIntersectionOf(
 : 419076005 |Allergic reaction|
 ObjectSomeValuesFrom(
  : 609096000 |Role group|
  ObjectIntersectionOf(
   ObjectSomeValuesFrom(
    : 370135005 |Pathological process|
    : 472964009 |Allergic process|
   )
   ObjectSomeValuesFrom(
    : 246075003 |Causative agent|
    : 44257100012418 |Tree nut|
   )
   ObjectSomeValuesFrom(
    : 116676008 |Associated morphology|
    : 442672001 |Swelling|
   )
  )
 )
)
```

Figure 2. PCE on the left side and the associated OWL ClassExpression based on functional syntax on the right side. The symbol ":" is a placeholder for the defined namespace, here for "*http://snomed.info/id/*".

Choosing the superconcept

One of the previously determined superconcept candidates needs to be selected as the superconcept. The superconcept is the pre-coordinated SNOMED CT concept, which is most similar to the PCE. Similar means that the superconcept covers the most information of the PCE compared to other pre-coordinated SNOMED CT concepts. Determining the semantic similarity between concepts is used for this purpose. It is a measure for calculating the taxonomic proximity between two elements within a knowledge base, such as SNOMED CT. If two elements are more similar, the semantic similarity is higher [28]. An information content-based approach and a path-based approach were considered to find the best approach for determining semantic similarity.

Information content-based approach

Information content-based approaches are intended to provide accurate results by taking a lot of semantic information into account. The values of information content (IC) mirror the grade of specification of concepts, where higher IC values indicate concepts lower in the hierarchy and are therefore more specific. A central component of this approach is the consideration of subconcepts [29]. Information content is calculated based on the probability of occurrence of a concept ($p(c)$) in a knowledge base:

$$IC(c) = -\log(p(c)).$$

Resnik et al. [29] have proposed the following formula for calculating the probability of occurrence $p(c)$:

$$p(c) = \sum_{n \in subsumer(c)} \frac{count(n)}{N}.$$

Here, the ratio of the number of subconcepts of concept c and the number of all concepts in the hierarchical knowledge base is calculated.

One approach for calculating the semantic similarity based on information content by Resnik et al. [29] considered in this paper is the measure of Lin et al. [30]:

$$¿(c_1, c_2) = \frac{2 \times (IC(LCA(c_1, c_2)))}{IC(c_1) + IC(c_2)}.$$

Here, the semantic similarity is calculated from the ratio of the Information content of the Least Common Ancestor (LCA) of the two concepts and the sum of the Information content of the individual concepts.

Path-based approach

In the path-based calculation, the semantic similarity is calculated based on subtype relationships or paths within a knowledge base. In some approaches, such as Choi and Kim's approach [31], only the shortest paths are considered. As an alternative, Sanches et al. [28] developed a new method for calculating semantic similarity. Sanches et al. describe that in large knowledge bases, such as SNOMED CT, a concept inherits information from several hierarchies and is connected to many concepts simultaneously. This means that it is insufficient to consider only the shortest path between two concepts. Therefore, Sanches et al. use all ancestors of a concept. The equation for the calculation is as follows:

$$¿(c_1, c_2) = -log_2\left(\frac{|T(c_1) \cup T(c_2)| - |T(c_1) \cap T(c_2)|}{|T(c_1) \cup T(c_2)|}\right).$$

Here, $T(c_i)$ is the set of direct superconcepts of a concept c_i and the concept c_i itself. To calculate the semantic similarity between the two concepts, the ratio between the set of non-shared superconcepts and the set of all superconcepts of both concepts is considered [28].

Implementation

An algorithm was developed which selects the most suitable or most similar SNOMED CT concept from the superconcept candidates. The calculation of the semantic similarity is based on a graph-based approach. For this, SNOMED CT must be represented one-time as a directed acyclic graph (DAG). The Release Format 2, version 2023-04-30 [26] is used for this purpose. The MRCM ReferenceSet contains, among other things, the relationships between two SNOMED CT Identifiers or concrete values. The SNOMED CT Identifiers or concrete values are the nodes in the DAG and the relationship types are represented via edges.

To calculate the semantic similarity based on one of the two measures between the PCE and a superconcept candidate, the PCE must be inserted into the DAG. For this purpose, a node called "pce" is temporarily introduced into the graph as a subnode of the focus concept and the previously determined superconcept candidates (see Figure 3). The algorithm iterates over the individual attribute values of the PCE and inserts the edges between these and the node "pce".

The semantic similarity between the PCE and all possible superconcept candidates is calculated. The superconcept candidate with the highest semantic similarity describes the PCE most accurately of all pre-coordinated concepts. This concept is defined as the superconcept.

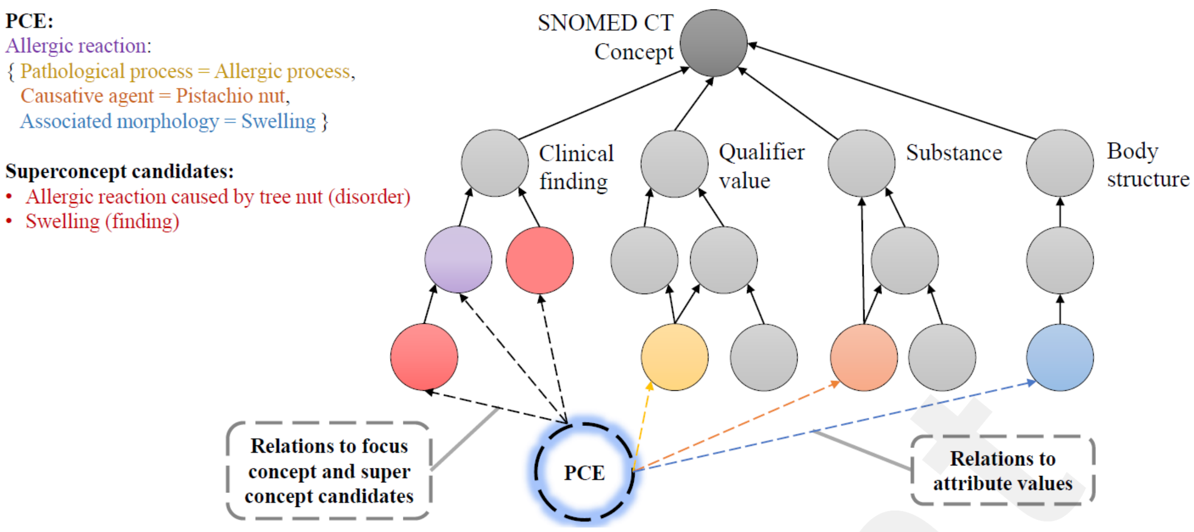


Figure 3. The PCE is represented as a sub-concept of its focus concept and the super-concept candidates. It also has edges to its attribute values.

Calculation of the delta

In most cases, the superconcept does not contain all the information contained in the PCE. The aim is now to determine the missing information in terms of attribute relationships. This is also referred to as the delta between the PCE and the superconcept. A graph-based approach was developed to calculate the delta. For this purpose, both the PCE and the superconcept are represented in separate graphs. The root node of both graphs must have the same name so that the delta can be calculated. The name "dummy" is chosen as the common name. The attribute relations of the PCE or the superconcept form further nodes in the graphs. These are connected to the node "dummy" via edges (Figure 4). The delta is calculated by subtracting the edges of the two graphs. This determines all edges that do not match between the PCE and the superconcept. The determined edges contain attribute relations that were not covered by the superconcept but are necessary for the precise representation of the PCE. These are either new attribute relations of the PCE compared to the superconcept or to attribute relations with certain attribute values (Figure 4).

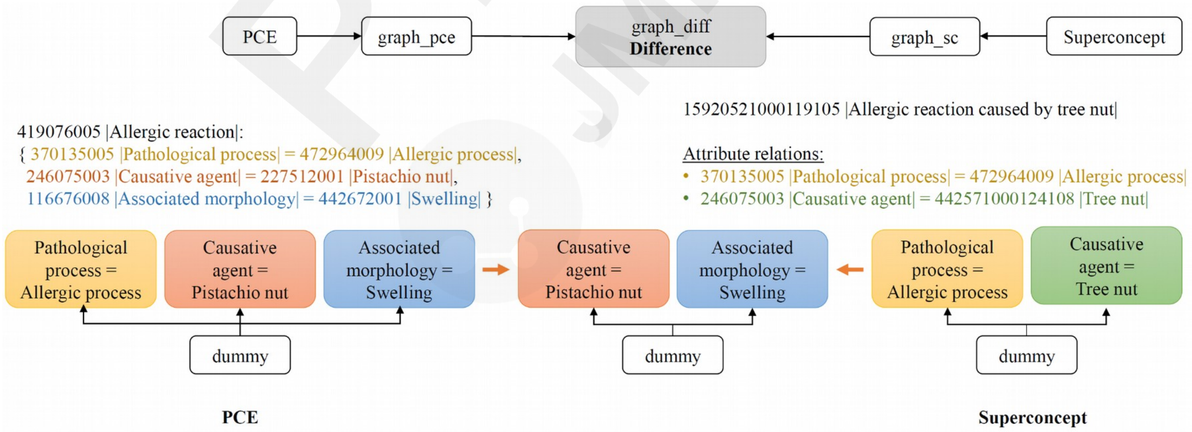


Figure 4. When calculating the delta, the graph of the superconcept is subtracted from that of the PCE. The determined edges contain attribute relations that were not covered by the superconcept but are necessary for the precise representation of the PCE.

Mapping to FHIR elements

In a final step, the superconcept and the delta must be stored as pre-coordinated concepts in suitable

FHIR elements. Either the FHIR basic resources from HL7 International or FHIR profiles can be used for this. FHIR profiles are certain resources that are customized to national characteristics and legislation as well as special use cases [32].

Two target representations were considered to illustrate the mapping:

- Basisprofile of the German National Association of Statutory Health Insurance Physicians (NASHIP, version 1.4.0) [4]
- profiles of the core data set of the German Medical Informatics Initiative (MII, version 1.0) [5].

Mapping rules are required to map the superconcept and the delta to the correct FHIR elements. It is not necessary to define a mapping rule for each PCE; instead, PCEs with similar content can be grouped into a respective category. The categories selected in this work are based on existing SNOMED CT Expression Templates. The Top-level hierarchies *Procedure* and *Clinical finding* are considered as examples. The hierarchy *Procedure* is considered as a whole, whereas the hierarchy *Clinical finding* can be subdivided into different thematic areas. The following three categories are considered: Allergy, Disease due to allergy, Allergic reaction.

The mapping rules are realized via FHIR StructureMaps [33], whereby a FHIR StructureMap is created for each category based on the FHIR profiles. The FHIR StructureMaps are stored on a local HAPI FHIR server [34].

The FHIR StructureMap contains exactly one entry for the superconcepts and SNOMED CT attributes (Figure 5, left side). For this, the SNOMED CT attributes of the SNOMED CT Expression Templates [35] which belong to one of the categories just mentioned, as well as other selected attributes that occur in pre-coordinated concepts of the categories mentioned, are considered. The selected attributes were chosen as relevant because PCEs are generally oriented towards the existing pre-coordinated SNOMED CT concepts and their concept definitions. The superconcept and the SNOMED CT attributes serve as source elements in the FHIR StructureMap, the corresponding FHIRPaths are used as target elements. FHIRPath is an extraction and navigation language for navigating through the individual elements and elements in an XML or JSON document. Thus, certain sub-elements of a FHIR resource can be traversed [2]. An example of a FHIRPath is *AllergyIntolerance.code*. This specifies that the associated value is stored in the element *code* of the resource *AllergyIntolerance*. The FHIR StructureMap also contains references (when several linked profiles are used) and the names of the respective profiles (Figure 5, right side).

Based on the FHIR StructureMaps, the corresponding FHIRPaths are created for the superconcept and the delta.

Figure 5. Extract of the FHIR StructureMap for allergic reactions in JSON format. It contains an entry for each SNOMED CT attribute and the superconcept as well as references and names of the

profiles.

In addition, a value range is defined for each FHIR StructureMap, whereby a PCE can be assigned to a FHIR StructureMap. The value range contains a subset of SNOMED CT concepts that fit the respective category. To determine the appropriate category for a PCE, the Ontoserver and the FHIR service *$expand* are used to check whether the superconcept is contained in the subset. It is a prerequisite that the value ranges of the individual FHIR StructureMaps are disjunct to each other.

Results

Web application

Bringing all the previously explained preliminary considerations and processing steps together, a web application called "PCEtoFHIR" (see Figure 6) was developed. The web development framework Angular (version 15.2.2) and Spring Boot (version 2.7.2, Java version 17) were used for this.

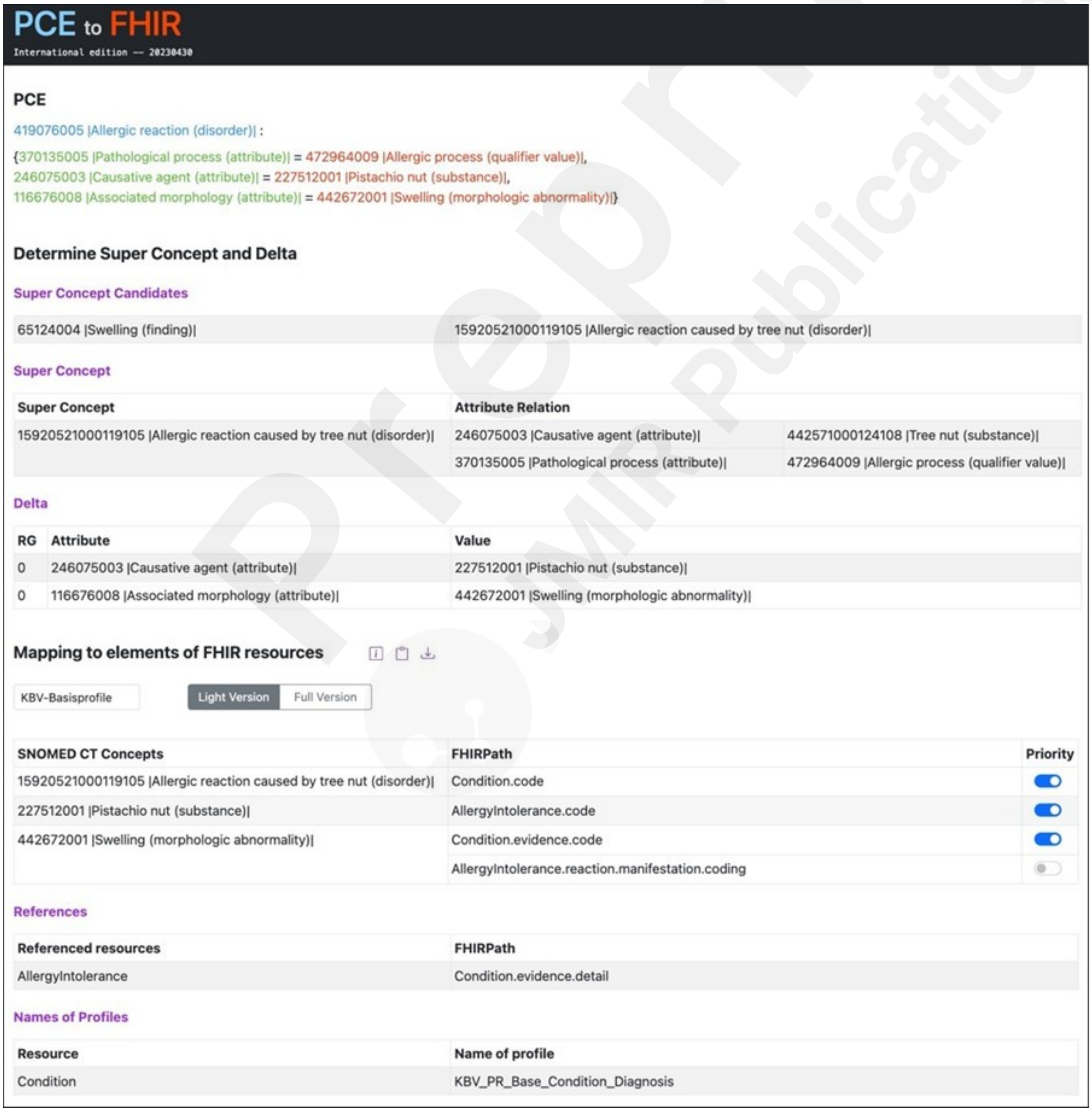


Figure 6. Excerpt of the web application PCEtoFHIR.

An excerpt of the web application is shown in Figure 6. After a syntactically and semantically correct PCE is entered, the superconcept candidates, the superconcept and the delta are determined automatically. This information is displayed to the user in the section “Determine Super Concept and Delta”. The superconcept and delta are mapped to the FHIR elements (FHIR basic resources or FHIR profiles). As already described, FHIR StructureMaps are used for each content category. The appropriate FHIR StructureMaps are automatically determined for the PCE and the desired mapping target can be selected in a combo box. In Figure 6, the “KBV-Basisprofile” (KBV is the German abbreviation for National Association of Statutory Health Insurance Physicians) was used. The FHIRPaths, the corresponding SNOMED CT concepts, the required references, and the names of the profiles can be copied to the clipboard or downloaded in the form of a text file. The properties of the developed web application are shown in Figure 6 in the section “Mapping to elements of FHIR resources“.

Determining the superconcept candidates

In this evaluation, the correct functionality of the algorithm for generating the OWL ClassExpression is analyzed. The OWL ClassExpressions of SNOMED CT concepts, which are published in the monthly release 2022-04-30 [26], serve as reference data.
For this evaluation, 600 randomly selected concept definitions of pre-coordinated SNOMED CT concepts are used. These contain different numbers of ungrouped and/or different numbers of grouped attribute relations as well as SNOMED CT concepts and/or concrete values as attribute values.

The developed algorithm converts the concept definitions into OWL ClassExpressions. These are then compared with the already existing OWL ClassExpressions, which serve as reference data. The comparison involves checking whether both OWL ClassExpressions match syntactically and semantically. This check is not performed for the focus concepts. The reason for this is that the OWL ClassExpressions of the reference data is based on the Stated Concept Definitions, while the Inferred Concept Definitions are used in the concept definitions of the SNOMED CT concepts. Therefore, the validation only considered whether a focus concept was syntactically in the correct position in the OWL ClassExpression. In summary, the validation revealed that the OWL ClassExpressions of all concept definitions and the reference data match.

Choosing the superconcept

As described previously, an information content-based approach by Lin et al. [30] and a path-based approach by Sanches et al. [28] were considered in this work. In this evaluation, both semantic similarity measures are compared. It was investigated which of the two measures is suitable for selecting the superconcept. The two approaches were analyzed in their principal functionality and this is illustrated below using an example:

419076005 |Allergic reaction|:
{ 370135005 |Pathological process| = 472964009 |Allergic process|,
246075003 |Causative agent| = 227512001 |Pistachio nut|,
116676008 |Associated morphology| = 442672001 |Swelling| }.

For the PCE shown, the following two concepts are superconcept candidates:

- *15920521000119105 |Allergic reaction caused by tree nut|*
- *65124004 |Swelling|.*

Table 2 shows the results of the two calculation methods.

The semantic similarity using the measure of Lin et al. [30] is 0.0 for the superconcept candidates.

Based on this result, it is not possible to select the superconcept. The reason for this is the Information Content. When calculating the Information content of the PCE, the subsumers of the PCE are important. In most cases, PCE are very specific and precise. As a result, they usually have no subconcepts. For this reason, the Information content is infinite:

$$IC(pce) = -\log\left(\sum_{n \in subsumer(pce)} \frac{count(n)}{N}\right) = -\log\log\left(\frac{0}{N}\right) = \infty$$

If the calculated value for $IC(pce)$ is used in the equation of Lin et al. for the PCE (pce) and a superconcept candidates (sc), the denominator always delivers infinite as a result:

$$¿(pce, sc) = \frac{2 \times (IC(LCA(pce, sc)))}{IC(pce) + IC(sc)} = \frac{2 \times (IC(LCA(pce, sc)))}{\infty + IC(sc)} = 0.0$$

This leads to a semantic similarity of zero for all candidates. Thus, Information Content-based measures are not useful in this step.

The measure of Sanches et al. [28] provides for the PCE (pce) different results for the two superconcept candidates (sc). As already mentioned, the ratio between the set of non-shared superconcepts and the set of all superconcepts of both concepts is considered. The set of all superconcepts is 37 for the concept *Allergic reaction caused by tree nut* and 33 for the concept *Swelling*.

In SNOMED CT many paths are available due to the many attribute relations, hierarchies and multiple inheritances. Many non-common superconcepts can be reached via these paths. Hence, this approach is an effective method for calculating semantic similarity. For the exemplary PCE, this means the following: With only eight non-common superconcepts, the concept *Allergic reaction caused by tree nut* has a significantly lower number of non-common superconcepts compared to *Swelling* (25). This leads to a significantly higher semantic similarity (2.209 versus 0.387) and therefore *Allergic reaction caused by tree nut* is the superconcept.

As a result, the measure of Sanches et al. was found feasible for superconcept determination and is thus used to calculate semantic similarity in this work.

Table 2. The semantic similarity is based on the two approaches varies. The measure of Sanches et al. provides different semantic similarities for the two superconcept candidates, which allows to determine a superconcept.

Superconcept candidates	**Lin et al. [30]**	**Sanches et al. [28]**
Allergic reaction caused by tree nut	0.0	2.209
Swelling	0.0	0.387

Mapping to FHIR elements

As previously described, the Basisprofile of the NASHIP (version 1.4.0) [4] and the profiles of the Core Data Set of the MII (version 1.0) [5] were used for mapping to the FHIR resources or FHIR profiles. To illustrate the mapping, four categories were considered. For each category and profile type, a FHIR StructureMap was created, resulting in a total of eight FHIR StructureMaps. Table 3 shows the FHIR resources that were used for the mapping.

Table 3. Depending on the category and profile type, different FHIR resources are used, which are listed here.

Category	**Profile type**	**FHIR resource**
Allergies	NASHIP	AllergyIntolerance

	MII	Condition, Observation
Disease due to allergies	NASHIP	AllergyIntolerance, Condition
	MII	Condition, Observation
Allergic reaction	NASHIP	AllergyIntolerance, Condition
	MII	Condition, Observation
Procedure	NASHIP	Procedure
	MII	Procedure

For the categories of the hierarchy *Clinical finding*, the FHIR StructureMaps include entries for the superconcept as well as eight considered SNOMED CT attributes, whereas for hierarchy *Procedure*, the superconcept and 12 additional SNOMED CT attributes were considered. Table 4 illustrates for each FHIR StructureMap how many elements in the profiles could be mapped either directly or through existing extensions, such as those of HL7 International. Overall, between 66.7% and 92.3% can be mapped, depending on the category.

Table 4. SNOMED CT elements can be represented in the profiles either directly or through extensions. Some elements may also not be mappable.

Category	**$\#_{elements}$ (total)**	**Profile type**	**$\#_{elements}$ for mapping**		
			without extension	**with extension**	**not possible**
Allergies	9	NASHIP	5	1	3
		MII	5	1	3
Disease due to allergies	9	NASHIP	7	0	2
		MII	5	1	3
Allergic reaction	9	NASHIP	7	0	2
		MII	5	1	3
Procedure	13	NASHIP	12	0	1
		MII	12	0	1

Overall evaluation with existing PCEs

For this evaluation, 33 PCEs are used, which are provided as part of the publication “Automatic full conversion of clinical terms into SNOMED CT concept” by Kate et al. [36]. It deals with the conversion of clinical texts into post-coordinated SNOMED CT expressions. The 33 PCEs consist of 23 PCEs of the Top-level hierarchy *Clinical finding* and 10 PCEs of the Top-level hierarchy *Procedure*. The following SNOMED CT attributes are used:

- *Clinical finding*: *Associated morphology*, *Finding site*, *Causative agent*, *Clinical course*, *Finding method*, *Pathological process*
- *Procedure*: *Method*, *Procedure-site (direct)*, *Using device*, *Using substance*.

A FHIR StructureMap for procedures, which contains the attributes used, has already been created as described above. This is used for this validation. For the category Clinical finding, a one-time FHIR StructureMap is created manually and saved locally on the HAPI FHIR server [34].

The PCEs were first validated based on the rules of the Compositional Grammar and the Concept Model using a local instance of Ontoserver and the FHIR services *$validate-code*. Here, one PCE was found to violate the cardinality restrictions of the Concept Model and was thus excluded from further processing.

For the remaining 32 PCE, the superconcept, the delta and the FHIRPaths were determined using the developed algorithm and the web application. Based on the FHIRPaths, the FHIR resources are created and stored on the local HAPI FHIR server [34]. Then, the values of the stored FHIR resources are extracted, which can be used to create a new transformed PCE (see Figure 7). This uses the superconcept as the focus concept. The refinement of the transformed PCE consists of attribute relationships that use the extracted SNOMED CT concept from the FHIR resources as attribute values. For each of the extracted SNOMED CT concepts, the corresponding SNOMED CT attribute must be determined. When mapping the attributes to the FHIRPaths, one FHIRPath can also lead to several SNOMED CT attributes. This means that it is not always possible to determine a SNOMED CT attribute directly. In these cases, separate considerations are necessary. Three special considerations were necessary in this validation, which allowed the determination of the attributes:

1. A value range is defined for each SNOMED CT attribute by the Concept Model. This allows a check for each SNOMED CT concept used in the validation to determine whether it is contained in the respective value range of an attribute, which limits the number of possible attributes.
2. Some SNOMED CT attributes are arranged hierarchically (e.g. *Procedure device – Using device – Using access devices*). If several attributes of a hierarchy are available, the most general hierarchical attribute is used in this validation (e.g.: *Procedure device*).
3. The concept definitions of pre-coordinated concepts were analyzed to determine which SNOMED CT attribute is mainly used for SNOMED CT concepts.

The generated PCE and the PCE that serves as reference data are then compared with each other. Ontoserver and the FHIR service *$subsumes* are used for this. If they were identified as being equivalent or if the transformed PCE was found to be a subclass of the original PCE, the overall validation was considered successful. This is possible due to the separate considerations and proves that the developed approach works correctly.

As a result of the validation, 31 of the 32 PCE were evaluated as equivalent. One PCE was found to be a subclass of the transformed PCE. The correct functioning of the approach could therefore be proven.

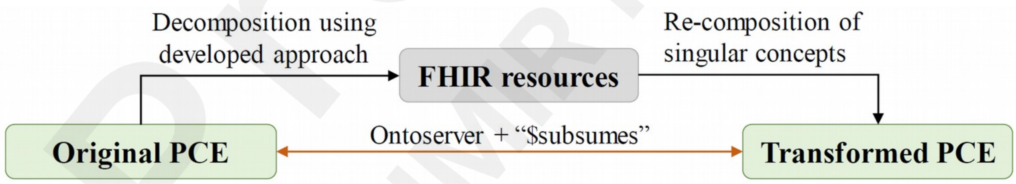


Figure 7. The superconcept and the concepts of the delta are stored on the HAPI FHIR server based on the determined FHIRPaths. Based on the stored values, a transformed PCE is created. The original PCE and the transformed PCE are compared.

Discussion

Principal Results

The aim of this work was to develop an algorithm that enables the efficient storage of post-coordinated SNOMED CT expressions, which is considered problematic. The introduction discussed various reasons for the reluctance to use PCEs. One reason was the familiar approach of editing and saving individual codes. With pre-coordinated concepts, a single code is available for a medical circumstance as well as a human-readable description. These two aspects are seen as advantages when using pre-coordinated concepts but are not given when using post-coordination. However, by identifying a superconcept based on OWL and semantic similarity, it is possible to identify a pre-coordinated SNOMED CT concept that most accurately describes the PCE. The PCE is a subconcept of the determined superconcept that can be displayed and analyzed. In addition, the decomposition of the PCE leads to individual codes that are familiar to the user from practice experience. FHIR is also regarded as one of the central interoperable data standards in the healthcare sector and is also becoming increasingly important in Germany [37] and worldwide [37, 38, 39].

The calculation of semantic similarity is essential for the developed approach. It is necessary that the calculation of semantic similarity truly identifies the concept most similar to a PCE. It became evident that Information Content-based approaches are not suitable for this purpose, as they consider the subnodes of a PCE. PCEs are typically very specific and precise, lacking subnodes, resulting in identical semantic similarity for all concepts. This does not allow a superconcept to be determined. In SNOMED CT many paths are available due to the many attribute relations, SNOMED CT hierarchies and multiple inheritances. These are considered in the approach by Sanches et al. [28]. Hence, the measure by Sanches et al. is an effective method for calculating semantic similarity.

As described above, the superconcept and the delta are stored in the respective FHIR resources or FHIR profiles and there in the matching elements. The basic profiles of the National Association of Statutory Health Insurance Physicians (NASHIP) [4] and the profiles of the core dataset of the Medical Informatics Initiative (MII) [5] were used for this purpose. However, the approach can be applied to all other profiles. As shown in Table 3, depending on the category and profile type, between 66.6% and 92.3% of the elements could be mapped. However, this also shows that not all attribute relations can always be mapped to the FHIR profiles. This illustrates the precision of the attributes of a PCE. It also reflects that through post-coordinated SNOMED CT expressions, medical data can be described in detail.

An analysis of the unmappable attributes of categories of the hierarchy *Clinical finding* for the Basisprofile of the NASHIP revealed that the unmappable attributes (*Clinical course*, *Due to* and *Finding site*) occur in only a very few pre-coordinated concepts. For these three attributes, there are suitable extensions by HL7 International that could be integrated into the Basisprofile. In the core data set, the attributes *Clinical course* and *Due to* cannot be covered, and additionally, the attribute *Causative agent* (cause for an allergy) cannot be mapped. The reason for this is that the core data set currently does not include a profile for the FHIR resource *AllergyIntolerance*.

For the category of the hierarchy *Procedure* based on the Basisprofile of the NASHIP and the profiles of the core data set of the MII, only the SNOMED CT attribute *Method* cannot be covered. For this, HL7 International already provides extensions for the FHIR resource *Procedure*. These could be integrated into the Basisprofile of the National Association of Statutory Health Insurance Physicians as well as into the profiles of the core data set of the Medical Informatics Initiative.

Furthermore, the mapping of SNOMED CT attributes to elements of FHIR resources cannot always be implemented exactly. The difficulty is that the SNOMED CT attributes are mapped to a limited number of FHIR elements, so that in some cases there is no exact match or several attributes have to be mapped to the same element. As a result, some of the original meaning is ultimately lost. An example of this are the SNOMED CT attributes *Using access device* and *Using device*. The SNOMED CT concepts for these two attributes can only be recorded via the FHIRPath *Device.type*, which is then referenced via the FHIRPath *Procedure.usedReference*. Nevertheless, the PCE should be stored in the metadata so that the missing information is not lost.

The FHIR StructureMaps have not yet been reviewed with domain experts for quality assurance purposes, although the authors have some experience with FHIR modeling [40] and a lot of experience with SNOMED CT from previous projects [4, 8, 18]. However, it became apparent that the mapping could be successfully solved for the considered categories and can be extended to other sub-hierarchies of SNOMED CT without any problems. Hence, further possibilities of application may be considered, e.g., regarding the TermInfo problem: As stated above, the decision to choose either terminology or information model as the place to represent medical facts is often variable and depends on the intended use. However, to some extent specific recommendations are established, such as "contextual meaning should rather be represented via the information model". Contrarily, the concepts of the SNOMED CT hierarchy *Situation with explicit context* include such contextual information that go beyond the ordinary scope of a terminology, e. g., suspected diagnoses, procedures not done or family history facts, and therefore put correct logical conclusions at risk (an explanation of the reasons (epistemological versus ontological components of meaning) is beyond the scope of this paper, see [3, 6]). Based on the concepts' logical definitions, which are directly related to PCEs, the approach presented in this paper allows to extract the problematic pieces of information and to store them within separate elements of the information model. As a result, a concept like *165008002 |Allergy testing not done (situation)|* could be represented by separating the epistemological aspect "not done" into the suitable FHIR element *Procedure.status*:

- *Procedure.code: 252512005 |Allergy test (procedure)|*
- *Procedure.status: not-done*.

Like this, the integrity of SNOMED CT's hierarchies may be preserved.

Conclusions

Structural and semantic standards must be adhered to ensure semantic interoperability. This paper has shown how a post-coordinated SNOMED CT expression (semantic standard) can be represented by assigning automatically determined pre-coordinated concepts to suitable FHIR elements (structural standard). For the developed approach, the determination of the OWL ClassExpression and the calculation of semantic similarity are essential. In an evaluation, it was demonstrated that the developed algorithm for creating the OWL ClassExpression provides correct results. Additionally, the approach by Sanches et al. [28] for calculating the semantic similarity was shown to deliver precise results. In summary, in an overall validation with 33 already existing PCEs, the correct functionality of the approach could be proven can be extended to other sub-hierarchies of SNOMED CT without any difficulties.

In this work, the developed approach uses a PCE as input and the mapping to the FHIR resources as output. It would be possible to change the direction of the approach. Based on the values in the elements of the FHIR resources, a PCE would be created in this variant. The created FHIR StructureMaps would provide a basis for this variant. The generated PCEs, which can be stored in extensions, for example, could enable expressive SNOMED CT-based analyses. Furthermore, the new versions of SNOMED CT and FHIR have an impact on the approach presented here. In future, it could be useful for the user to be notified about changes. In addition to mapping to FHIR resources, the approach can also be applied to other information models, such as openEHR or relational databases.

Acknowledgements

This work is funded by the German Federal Ministry of Education and Research (BMBF) as part of the Medical Informatics Initiative Germany, grant 01ZZ2312A.

Conflicts of Interest

The authors declare no conflict of interest.

Abbreviations

CSIRO Commonwealth Scientific and Industrial Research Organization
C-CDA Consolidated Clinical Document Architecture
CDA Clinical Document Architecture
DAG Directed Acyclic Graph

ECL	Expression Constraint Language
FHIR	Fast Healthcare Interoperability Resources
HL7	Health Level 7
IC	Information Content
JSON	JavaScript Object Notation
KBV	Kassenärztliche Vereinigung
LCA	Least Common Ancestor
MII	Medical Informatic Initiative
MRCM	Machine Readable Concept Model
NASHIP	National Association of Statutory Health Insurance Physicians
PCE	Post-coordinated expression
RIM	Reference Information Model
XML	Extensible Markup Language
OWL	Web Ontology Language
WASP	Web application supporting SNOMED CT postcoordination

References

1. Benson T, Grieve G. Principles of Health Interoperability: FHIR, HL7 and SNOMED CT. Fourth edition. Springer; 2021.
2. Braunstein ML. Health Informatics on FHIR: How HL7's API Is Transforming Healthcare. Springer International Publishing; 2022. doi:10.1007/978-3-030-91563-6
3. Ingenerf J, Drenkhahn C. REFERENZTERMINOLOGIE SNOMED CT: Interlingua zur Gewährleistung semantischer Interoperabilität in der Medizin. Springer Vieweg; 2023.
4. Kassenärztliche Bundesvereinigung. KBV-Basis-Profile. Accessed November 29, 2023. https://simplifier.net/organization/kassenrztlichebundesvereinigungkbv
5. Medizininformatik Initiative. Der Kerndatensatz der Medizininformatik-Initiative. Accessed November 29, 2023. https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative
6. Schulz S, Schober D, Daniel C. Bridging the semantics gap between terminologies, ontologies, and information models. Studies in health technology and informatics. 2010;160(2):1000-1004. doi:10.3233/978-1-60750-588-4-1000
7. Markwell D, Sato L, Cheetham E. Representing clinical information using SNOMED Clinical Terms with different structural information models. Representing and sharing knowledge using SNOMED Proceedings of the 3rd international conference on Knowledge Representation in Medicine (KR-MED 2008).2008:72-79.
8. Drenkhahn C, Ohlsen T, Wiedekopf J, Ingenerf J. WASP—A Web Application to Support Syntactically and Semantically Correct SNOMED CT Postcoordination. Applied Sciences. 2023;13(10):6114. doi:10.3390/app13106114
9. Wardle M. Health informatics and information technology; Semantic interoperability: SNOMED CT, post-coordination and the model. Accessed November 29, 2023. https://wardle.org/terminology/2018/10/27/snomed-postcoordination-1.html
10. Metke-Jimenez A, Steel J, Hansen D, Lawley M. Ontoserver: a syndicated terminology server. J Biomed Semant. 2018;9(1):24. doi:10.1186/s13326-018-0191-z
11. CSIRO. Ontoserver: SNOMED CT Post-Coordination support. Accessed November 29, 2323. https://ontoserver.csiro.au/docs/6/postcoordination.html
12. HL7 International. Resource CodeSystem - Content. Accessed November 29, 2023. https://hl7.org/fhir/codesystem.html
13. HL7 International. Resource Condition - Mappings. Accessed November 23, 2023.

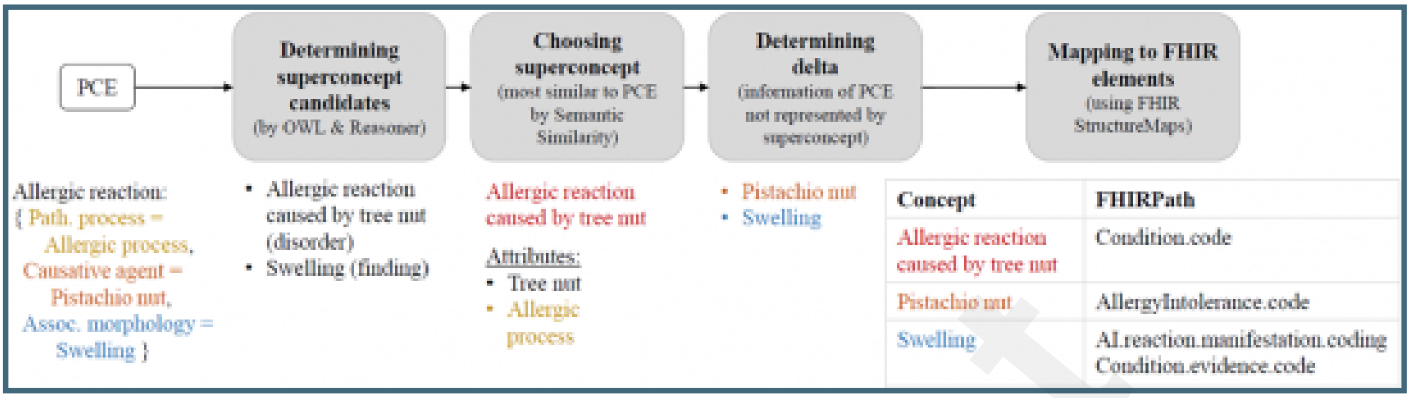
https://build.fhir.org/condition-mappings.html
14. HL7 International. Resource Observation - Mappings. Accessed November 23, 2023. https://build.fhir.org/observation-mappings.html
15. SNOMED International. SNOMED on FHIR. Accessed November 23, 2022. https://confluence.ihtsdotools.org/display/FHIR/
16. Green JM, Wilcke JR, Abbott J, Rees LP. Development and Evaluation of Methods for Structured Recording of Heart Murmur Findings Using SNOMED-CT(R) Post-Coordination. Journal of the American Medical Informatics Association. 2006;13(3):321-333. doi:10.1197/jamia.M1973
17. Miñarro-Giménez JA, Martínez-Costa C, López-García P, Schulz S. Building SNOMED CT Post-Coordinated Expressions from Annotation Groups. Stud Health Technol Inform. 2017;235:446-450.
18. Ohlsen T, Kruse V, Krupar R, Banach A, Ingenerf J, Drenkhahn C. Mapping of ICD-O Tuples to OncoTree Codes Using SNOMED CT Post-Coordination. Studies in Health Technology and Informatics. 2022;294:307-311. doi:10.3233/SHTI220464
19. Castell-Díaz J, Miñarro-Giménez JA, Martínez-Costa C. Supporting SNOMED CT postcoordination with knowledge graph embeddings. Journal of Biomedical Informatics. 2023;139:104297. doi:10.1016/j.jbi.2023.104297
20. Perez-Rey D, Alonso-Calvo R, Paraiso-Medina S, Munteanu CR, Garcia-Remesal M. SNOMED2HL7: A tool to normalize and bind SNOMED CT concepts to the HL7 Reference Information Model. Comput Methods Programs Biomed. 2017;149:1-9. doi:10.1016/j.cmpb.2017.06.020
21. Arguello Casteleiro M, Martinez-Costa C, Des-Diu J, Maroto N, Fernandez-Prieto MJ. From SNOMED CT expressions to an FHIR RDF representation: Exploring the benefits of an ontology-based approach. ODLS 2019: Ontologies and Data in Life Sciences 2019. Published online July 11, 2019:1-13.
22. Hitzler P, Krötzsch M, Rudolph S, Sure Y. Semantic Web. Springer Berlin Heidelberg; 2008. doi:10.1007/978-3-540-33994-6
23. Staab S, Studer R, eds. Handbook on Ontologies. Springer Berlin Heidelberg; 2009. doi:10.1007/978-3-540-92673-3
24. International Health Terminology Standards Development Organisation. SNOMED CT OWL Guide. Published online January 28, 2022. Accessed November 29, 2023. https://confluence.ihtsdotools.org/display/DOCOWL/SNOMED+CT+OWL+Guide?preview=/64265998/142119608/SNOMED%20CT%20OWL%20Guide-v19-20220128.pdf
25. International Health Terminology Standards Development Organisation. Snomed OWL Toolkit. Published online February 2, 2016. https://github.com/IHTSDO/snomed-owl-toolkit
26. International Health Terminology Standards Development Organisation. SNOMED CT International Edition. Accessed January 4, 2024. https://www.nlm.nih.gov/healthit/snomedct/international.html
27. Kazakov Y, Krötzsch M, Simancik F. ELK Reasoner: Architecture and Evaluation. Proceedings of the 1st International Workshop on OWL Reasoner Evaluation (ORE-2012) CEUR Workshop Proceedings. Published online July 2012.
28. Sánchez D, Batet M. Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. Journal of Biomedical Informatics. 2011;44(5):749-759. doi:10.1016/j.jbi.2011.03.013
29. Resnik P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. Published online November 29, 1995. Accessed January 4, 2024. http://arxiv.org/abs/cmp-lg/9511007
30. Lin Dekang. An information-theoretic definition of similarity. Fifteenth International Conference on Machine Learning, ICML 1998, Morgan Kaufmann, Madison (Wisconsin, USA). Published online 1998:296-304.
31. Choi I, Kim M. Topic distillation using hierarchy concept tree. In: Proceedings of the 26th

Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. ACM; 2003:371-372. doi:10.1145/860435.860506
32. HL7 International. Profiling FHIR. Accessed November 29, 2023. https://build.fhir.org/profiling.html
33. HL7 International. Resource StructureMap - Content. Accessed November 23, 2023. https://build.fhir.org/structuremap.html
34. Smile Digital Health. HAPI FHIR. Accessed January 4, 2024. https://hapifhir.io/hapi-fhir
35. International Health Terminology Standards Development Organisation. Template Syntax Specification. 2020;(1.1.1). Accessed November 29, 2023. https://confluence.ihtsdotools.org/display/DOCSTS?preview=/45529301/115875508/doc_TemplateSyntax_v1.1.1-en-US_INT_20201020.pdf
36. Kate RJ. Automatic full conversion of clinical terms into SNOMED CT concepts. Journal of Biomedical Informatics. 2020;111:103585. doi:10.1016/j.jbi.2020.103585
37. Lau, Tobias. Gesundheitsdaten: FHIR wird europaweiter Standard. Ärzteblatt. Accessed January 4, 2024. https://www.aerzteblatt.de/nachrichten/142159/Gesundheitsdaten-FHIR-wird-europaweiter-Standard
38. Vorisek CN, Lehne M, Klopfenstein SAI, et al. Fast Healthcare Interoperability Resources (FHIR) for Interoperability in Health Research: Systematic Review. JMIR Med Inform. 2022;10(7):e35724. doi:10.2196/35724
39. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities. JMIR Med Inform. 2021;9(7):e21929. doi:10.2196/21929
40. Rinaldi E, Drenkhahn C, Gebel B, et al. Towards interoperability in infection control: a standard data model for microbiology. Sci Data. 2023;10(1):654. doi:10.1038/s41597-023-02560-x

Supplementary Files

Figures

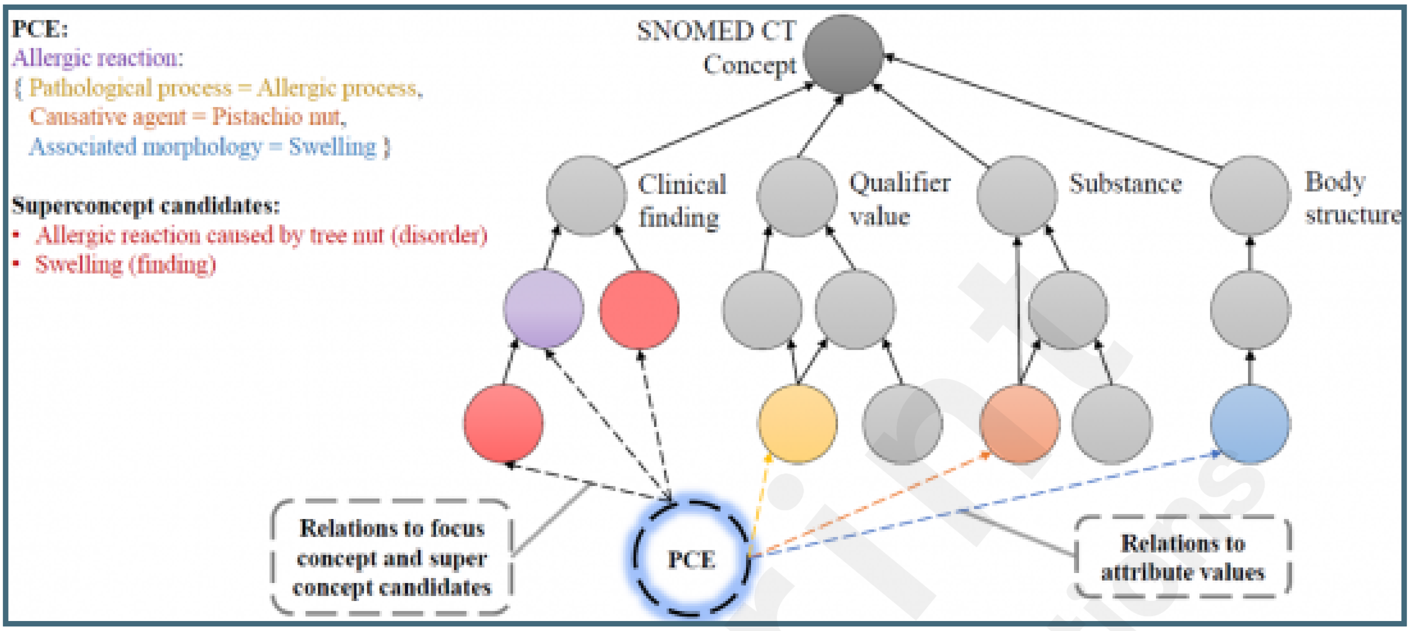
The decomposition of a PCE into elements of (profiled) FHIR resources consists of four steps.



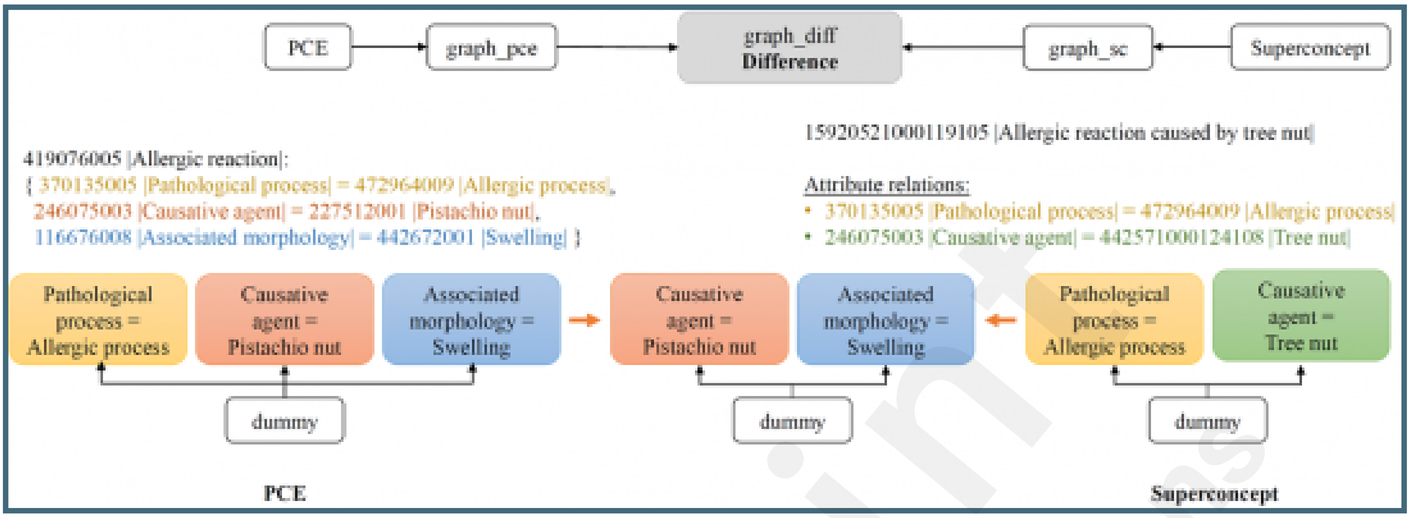
PCE on the left side and the associated OWL ClassExpression based on functional syntax on the right side. The symbol ":" is a placeholder for the defined namespace, here for "http://snomed.info/id/".

PCE	Associated OWL Class Expression
419076005 |Allergic reaction| : { 370135005 |Pathological process| = 472964009 |Allergic process| , 246075003 |Causative agent| = 227512001 |Pistachio nut| , 116676008 |Associated morphology| = 442672001 |Swelling| }	ObjectIntersectionOf(: 419076005 |Allergic reaction| ObjectSomeValuesFrom(: 609096000 |Role group| ObjectIntersectionOf(ObjectSomeValuesFrom(: 370135005 |Pathological process| : 472964009 |Allergic process|) ObjectSomeValuesFrom(: 246075003 |Causative agent| : 44257100012418 |Tree nut|) ObjectSomeValuesFrom(: 116676008 |Associated morphology| : 442672001 |Swelling|))))

The PCE is represented as a sub-concept of its focus concept and the super-concept candidates. It also has edges to its attribute values.



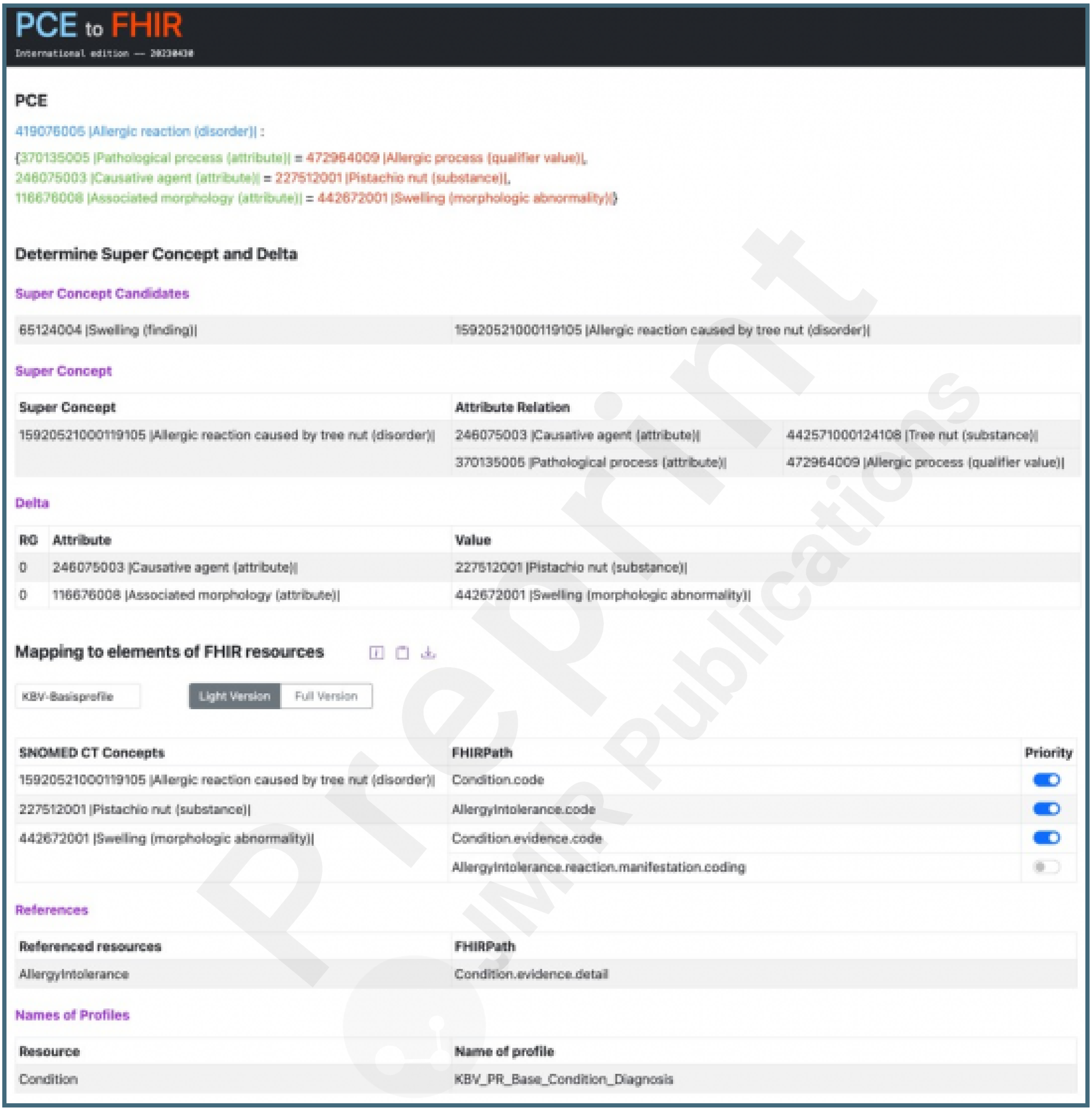
When calculating the delta, the graph of the superconcept is subtracted from that of the PCE. The determined edges contain attribute relations that were not covered by the superconcept but are necessary for the precise representation of the PCE.



Extract of the FHIR StructureMap for allergic reactions in JSON format. It contains an entry for each SNOMED CT attribute and the superconcept as well as references and names of the profiles.

```
{
    "resourceType": "StructureMap",
    "id": "KBV-Mio-AllergicReaction",
    ...
    "status": "active",
    "group": [ {
      "name": "De-Composition-AllergicReaction",
      "typeMode": "none",
      ...
      "rule": [ {
        "name": "CausativeAgent",
        "source": [ {
          "context": "source",
          "element": "246075003"
        } ],
        "target": [ {
          "context": "target",
          "contextType": "variable",
          "element": "AllergyIntolerance.code",
          ...
        } ]
      }, ...
      {
        "name": "References",
        "source": [ {
          "context": "source",
          "element": "references"
        } ],
        "target": [ {
          "context": "target",
          "contextType": "variable",
          "element": "Condition.evidence.detail->AllergyIntolerance",
        } ]
      }, {
        "name": "NamesProfiles",
        "source": [ {
          "context": "source",
          "element": "namesprofiles"
        } ],
        "target": [ {
          "context": "target",
          "contextType": "variable",
          "element": "Condition->KBV_PR_Base_Condition_Diagnosis",
        } ]
      } ]
    } ]
}
```

Excerpt of the web application PCEtoFHIR.



The superconcept and the concepts of the delta are stored on the HAPI FHIR server based on the determined FHIRPaths. Based on the stored values, a transformed PCE is created. The original PCE and the transformed PCE are compared.