

Towards a Knowledge Graph of Health Evolution

Alba Catalina Morales Tirado¹[0000-0001-6984-5122], Enrico Daga¹[0000-0002-3184-5407], and Enrico Motta¹[0000-0003-0015-1952]

KMi, The Open University, Milton Keynes, United Kingdom
{alba.morales-tirado, enrico.daga, enrico.motta}@open.ac.uk

Abstract. Electronic Health Records (EHR) contain detailed data of a person’s health conditions and could provide emergency first responders with useful information. In previous works, we envisaged an intelligent system able to inspect health records and identify people in need of special assistance, by reasoning on the evolution of conditions over time. Unfortunately, there is a lack of resources regarding health condition evolution and recovery time. However, information available on the web could help in supporting domain experts for building a database of Health Condition Evolution Statements (HES).

This paper addresses this knowledge gap and proposes a four-step methodology based on knowledge acquisition (KA) techniques that support the extraction of HES from public sources. The approach uses text classification algorithms and exploits SNOMED CT taxonomy to build a database of HES. More importantly, the proposed KA pipeline includes a human-in-the-loop model that captures knowledge from experts and ensures the construction of high-quality Knowledge Graphs (KG) to support the task at hand. We evaluate the approach with domain experts’ help and discuss the user study results. Finally, we contribute the first curated Knowledge Graph of HES.

Keywords: Knowledge Graph · Health evolution · Condition evolution · Dataset · Knowledge Acquisition · Knowledge extraction · SNOMED CT.

1 Introduction

Healthcare data use, particularly Electronic Health Records (EHR), has received increasing attention in recent years. EHR constitutes a valuable information asset for emergency support systems, they include extensive and fine-grained details of a person’s medical issues. This information could provide a snapshot of people’s health status. For instance, identifying vulnerable people or people otherwise requiring special assistance in the context of an emergency [7,8]. However, assessing ongoing health issues represents a challenge to first responders. EHR contain an overwhelming amount of information that emergency services cannot process effectively, for both its size and specificity; as a result, crucial data might be overlooked.

In previous work [8], we describe two knowledge components an intelligent system requires to reason on EHR automatically. The first component is the HECON Ontology (Health Condition Evolution Ontology) [9], a model for representing and reasoning on condition evolution over time. HECON defines the recovery process as a set of features called Health Evolution Statement (HES); whose components are *type*, *pace* and *duration*. Type refers to how the health condition evolves (e.g., improvement, decline, permanent, unaffected). Pace indicates the speed at which it changes (slow, moderate, fast), and Duration is an estimation of span (expressed as minimum and maximum range). The second element is knowledge about health condition evolution, specifically structured data that will support the annotation of conditions according to HECON. However, to the best of our knowledge, no existing structured data about condition evolution is available for reuse.

In this paper, we address this gap by designing a Knowledge Acquisition (KA) pipeline focused on extracting Health Evolution Statements (HES) from unstructured data sources. We expand the initial work presented in [8] by including a Human-in-the-loop (HITL) module. The HITL step uses the recommendations generated in previous steps to facilitate domain experts' tasks and annotate conditions with HES statements. By capturing domain experts' knowledge, we accelerate the annotation of health conditions and ensure the construction of a high-quality Knowledge Graph (KG). Contributions¹ of this paper are:

- A methodology that implements a Knowledge Acquisition pipeline for building a Knowledge Graph of Health Evolution.
- A tool that instantiates the HITL module.
- A user study with domain experts to (a) refine the HES KG and (b) evaluate the overall approach.
- The first database of health evolution information published as KG.

The rest of the paper is organised as follows. Section 2, describes the background and related work. Section 3 gives an overview of the proposed KA pipeline and Section 4 describes its implementation. In Section 5 we give details of the Knowledge Graph construction. In Section 6 we present the KA pipeline's evaluation and the user study's results. Finally, in Section 7, we summarise the conclusions.

2 Background and Related Work

In the Smart City environment, data is gathered by different means and from different sources to improve city services. In recent years, attention has focused on using Electronic Health Records to assist emergency services [1,8,7].

Data held in Electronic health records could reveal ongoing health issues and help us identify vulnerable people in need of assistance during an emergency. To illustrate this, consider a fire emergency in a large building. Among the people in the building, two require special assistance: one is a wheelchair user, and the

¹ Repository: <https://github.com/albamoralest/Health-Condition-Evolution-database>

other suffers from lung disease. Evidently, the wheelchair user needs support to evacuate, but such information is not typically known to emergency services. An intelligent system with access to the person’s health records could identify a diagnosis of a ‘Fracture of the spine’ (a permanent condition) and understand that such a condition does not improve over time. However, the second case is less obvious. A person suffering from a lung disease like ‘Obstructive bronchitis’ might suffer acute symptoms (difficulty in breathing and walking) only when the diagnosis is recent. In fact, these symptoms may disappear with appropriate treatment over time. An intelligent system with enough information about health evolution could automatically evaluate this person’s EHR, provide crucial information about their recent health issues and evaluate how severe they are at the time of the emergency.

Concerning EHR analysis, much of the literature is oriented to facilitate the visual representation of historical clinical data by reasoning on past health events [11,2]. However, existing literature [4] does not address the problem of identifying ongoing health issues by reasoning on health condition evolution from EHR.

In [8,9] we presented the work focusing on providing the knowledge components required by an intelligent system to represent and reason on health evolution. The HECON Ontology (Health Condition Evolution Ontology) [9] is the first model created with this objective. As described previously, HECON represents the recovery process as a set of features called Health Evolution Statement (HES). However, we identified a lack of structured data available for reuse.

As a result, in [8] we presented initial work on the automatic extraction of Health Evolution Statements (HES) from unstructured data sources. We collected text from public websites, such as NHS England² and MAYO Clinic³ and linked this information to the SNOMED CT taxonomy. We used knowledge classification techniques such as Machine Learning to classify the collected text according to HES features. We used the identified HES and took advantage of SNOMED CT semantic features to propagate health evolution statements to a more significant number of SNOMED CT concepts. However, this approach had some limitations:

- The generation of HES was narrowed to the number of health conditions collected from public sources, typically the most common diseases. Instead, domain experts could provide knowledge of a larger number of health conditions.
- The methodology presented in [8] demonstrated that health condition evolution data supported the identification of vulnerable people effectively; nevertheless, it is imperative to evaluate the overall approach and the accuracy of the automatically generated HES.
- Although the recommended HES are generated using reliable and authoritative sources, the resulting data was not validated by domain experts.

This paper describes the implementation of a four-phase Knowledge Acquisition (KA) pipeline. The pipeline uses health conditions text descriptions as

² <https://www.nhs.uk/conditions/>

³ <https://www.mayoclinic.org/diseases-conditions>

input and returns a list of recommended HES. We complete the KA pipeline by including a Human-in-the-loop (HITL) step, enhancing the construction of the KG of Health Condition Evolution. We can infer from the limitations listed that integrating experts in the approach can benefit the overall process of automatically generating HES and that capturing their knowledge is a crucial step for ensuring the accuracy and quality of the information.

In order to validate the overall approach, we carried on a user study involving domain experts. The user study’s objective is to evaluate the accuracy of the recommendations generated as part of the KA pipeline. Furthermore, we aim to assess the viability of annotating health conditions by incorporating a HITL step. Consequently, we sought to answer the following research questions:

- RQ1: Can health condition evolution statements be extracted automatically from descriptions in natural language?
- RQ2: Can we use ontological knowledge to derive new health condition evolution statements automatically?
- RQ3: How sustainable is the proposed methodology to populate a database of health evolution?

In what follows, we present the application of the proposed methodology for building a structured health condition evolution database and the results obtained from the user study that will answer the proposed research questions.

3 Methodology overview

This section describes briefly the knowledge acquisition pipeline proposed to build a database of health evolution information (see Figure 1). First, we collect the data that describes health condition evolution. Next, we apply text classification techniques and knowledge completion methods to extract recommended health condition evolution statements from natural language. Lastly, we use the recommendations and incorporate domain experts’ knowledge to accelerate the annotation of health conditions and ensure a high-quality HES generation. In what follows, we summarise the steps of our proposed approach.

1. Corpus preparation. The first step of the pipeline is dedicated to identifying data sources that describe health evolution and preparing the corpus to be used in the next step. The sources should comply with characteristics such as: being an authoritative source, publicly available, extensive and including a description of health evolution. The aim is to collect text describing diseases, procedures, and conditions (e.g., asthma, appendicitis, bronchitis) and link them to the corresponding concept in SNOMED CT taxonomy. The final output is a corpus of health conditions organised by sentences.

2. Knowledge components extraction. The output from the previous step is a large corpus of sentences; however, only a few sentences provide information on the evolution of health conditions. Therefore, the next task is to identify and classify the sentences according to the HES components defined by HECON Ontology. We rely on Machine Learning techniques and develop a

pipeline that includes the training and testing of a set of models [5] for each feature of the HES statement. Next, the best-performing models are used to predict a HES for each sentence in the corpus. Since a condition can have one or more recommended HES, the next task is to clean inconsistent and repeated HES. Lastly, we apply an algorithm that uses support and confidence as metrics to rank the most frequent combination of annotations. The output of this step is a collection of SNOMED CT concepts linked to one or more recommended HES.

3. Knowledge completion. The recommended annotations generated in the previous step have limited coverage of SNOMED CT concepts; therefore, we exploit the semantic structure of SNOMED CT taxonomy to find similar concepts that could share the same HES. Specifically, we use the SNOMED CT concepts' features to identify patterns and derive propagation rules. The rules expand the coverage of the HES to other concepts in SNOMED CT and make it possible to elicit a large dataset of SNOMED CT concept annotations.

4. Human-in-the-loop. Until now, the proposed methodology generated one or more recommended HES for each condition. Selecting the more accurate HES requires additional knowledge. Therefore, in this step, the objective is to capture domain experts' knowledge and build a more accurate database of health condition information. Domain experts contribute in two ways: (a) by assessing the recommended annotations and (b) by creating new ones. The final output is a curated database of health evolution statements.

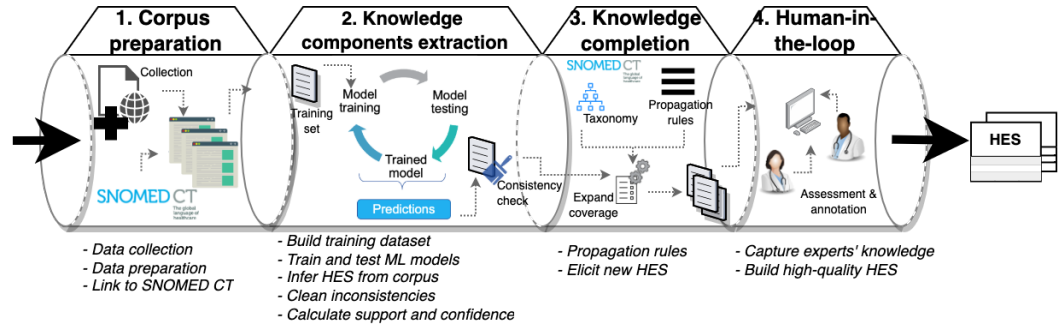


Fig. 1: Knowledge Acquisition pipeline

4 Knowledge Acquisition Pipeline

This section describes how we developed the proposed knowledge acquisition pipeline, as shown in Figure 1.

4.1 Corpus preparation

The first step is dedicated to identifying data sources that describe health condition evolution. The sources should come from (a) an authoritative organisation

and (b) publicly available. Also, sources should be (c) extensive and (d) contain descriptions of condition evolution.

We identify two health websites: NHS England and MAYO Clinic. NHS England is the largest health website in the UK, providing straightforward access to content about symptoms, conditions, and treatments. The MAYO Clinic is a non-profit organisation; its website provides comprehensive and easy access to condition descriptions. NHS England website displays information of 972 health conditions and MAYO Clinic, 1170.

From both websites, we collect the HTML files that contain conditions descriptions. We clean the text by removing HTML tags, line breaks, special characters and empty spaces. After reviewing the text, we found out that condition evolution is usually described in one sentence. For instance, the evolution of ‘Broken ankle’ is described as ‘*A broken ankle usually takes 6 to 8 weeks to heal, but it can take longer.*’. Other conditions such as ‘Cataract surgery’ has more than one description: ‘*It can take 2 to 6 weeks to fully recover from cataract surgery.*’ and ‘*These side effects usually improve within a few days, but it can take 4 to 6 weeks to recover fully.*’. Therefore, we organise the corpus in sentences. The dataset contains 208,838 sentences in total, grouped by health conditions.

Typically, EHR uses SNOMED CT as a standard to describe clinical conditions [10]; therefore, we need to align the conditions’ names from the web sources to SNOMED CT. These alignments facilitate the link of the HES to the EHR. We use Levenshtein distance to perform this alignment and find matching conditions’ names. We run a manual review of the results in randomly selected conditions. The final corpus is a collection of sentences grouped by health conditions, where each health condition is linked to its corresponding SNOMED CT identifier.

4.2 Knowledge components extraction

Here, the focus is on extracting Health Evolution Statements candidate recommendations from the corpus (see Figure 1, step 2). In what follows, we describe each task of the knowledge component extraction process in detail.

Building a gold standard dataset of HES. In order to create this dataset, we examine the corpus. Health condition descriptions are extensive and have an average of 180 sentences. However, only a few of them describe health condition evolution; therefore, we use a distance supervision approach to identify these sentences. First, we select a sample of text snippets (expressions such as *last between, lifelong condition, no specific cure*, among others) that refer to condition recovery. Next, we use this sample and apply cosine similarity to automatically find a larger sample of sentences in the corpus. Then, we manually annotate each sentence with its corresponding HES components: type of condition (improve, decline, permanent), pace (fast, moderate, slow) and duration (maximum and minimum duration).

Finally, in preparation for the ML classification task, we add negative annotations to the training set. As mentioned earlier, the corpus contains sentences

that do not express health evolution. We emulate this by adding sentences without this information and annotating them as ‘NONE’. The output is a manually curated gold standard of 1987 sentences and their corresponding HES. Table 1 summarises the total number of sentences grouped by HES.

Table 1: Number of sentences per HES in the training dataset

Health Condition Evolution Statement (HES)			Total
Type	Pace	Duration	
NONE			1437
PERMANENT			141
IMPROVEMENT	MODERATELY	8 DAYS TO 2 MONTHS	106
IMPROVEMENT	FAST	5 MINUTES TO 1 DAY	74
DECLINE	SLOWLY	1 YEAR TO MORE YEARS	56
IMPROVEMENT	MODERATELY	2 MONTHS TO 6 MONTHS	53
IMPROVEMENT	FAST	1 DAYS TO 1 WEEKS	37
IMPROVEMENT	SLOWLY	1 YEAR TO MORE YEARS	37
IMPROVEMENT	SLOWLY	6 MONTHS TO 1 YEAR	30
DECLINE	FAST	1 DAY TO 1 WEEK	6
DECLINE	SLOWLY	6 MONTHS TO 1 YEAR	4
DECLINE	MODERATELY	8 DAYS TO 2 MONTHS	4
DECLINE	MODERATELY	2 MONTHS TO 6 MONTHS	2
TOTAL sentences			1987

Training and testing Machine Learning (ML) algorithms for the classification task. This task focuses on classifying sentences according to the dimensions used in HECON Ontology [9]. To perform this classification, we use the gold standard built previously as input for training and testing different ML algorithms [5].

We randomly divide the gold standard into training (70%) and test (30%) datasets, both with the same proportion of class labels as the gold standard. Table 2 shows the list of the ML algorithms we trained and the accuracy of each model grouped by HES features. Highlighted in bold are the models with the best performance. Also, hyper-parameters configuration for each algorithm can be found in the repository ⁴.

The training and test task is divided into two parts. For the first part, we perform a preliminary classification (C0). We train a boolean classifier to discriminate sentences that describe health condition evolution from those that do not contain such description. For instance, text such as ‘*Landing awkwardly from a jump*’ is classified as NO (it does not describe health evolution) and ‘*There’s currently no cure for chronic obstructive pulmonary disease (COPD)*’ is classified as a description of health evolution. In this way, we aim to identify sentences containing condition evolution information and increase the number of true positives.

For the second part, we take into account the definition of the Health Evolution Statement. As described in [9], the HES comprises three features: type of health condition, pace and duration. Therefore, we use the sentences to train our classifiers across the three dimensions of the HES. For instance, the sentence ‘*There’s currently no cure for chronic obstructive pulmonary disease (COPD)*’

⁴ KG of Health Condition Evolution

is used to train the different algorithms, first to obtain the type of condition, then the pace and the duration.

Table 2: ML training results: Accuracy per algorithm & HES features

ML Algorithms	C0	Direction	Pace	Duration
<i>Logistic Regression</i>	0.8816	0.9727	0.8148	0.8114
<i>Decision Tree</i>	0.8337	0.9272	0.8934	0.8606
Linear SVC	0.8789	0.9545	0.8271	0.8360
MLP Classifier	0.8337	0.9545	0.7530	0.6803
Naïve Bayes	0.4181	-	-	-
Multinomial NB	0.8136	0.8636	0.6790	0.5737
Random Forest Classifier	-	0.8818	0.7901	0.8442

Application of the machine learning approach. Once we are satisfied with the results obtained in the previous task, we use the best-performing models for each feature of the HES (see Table 2) and make predictions on the entire corpus.

First, we run predictions using the best ‘C0’ classification model. A total of 5,174 out of 208,838 sentences were classified as providing information about condition evolution. Table 3, column ‘Sentence’, shows examples of sentences identified as positives. Second, we take this reduced dataset and run an independent classification process for each dimension of the HES. The first dimension is the type of health condition: improvement, decline or permanent. For instance, in Table 3, the sentence in the first row is classified as ‘decline’ and the one in the second row as ‘permanent’. As described in HECON, only values ‘improvement’ and ‘decline’ have *progress* dimension. Thus, only 4,306 sentences, annotated as improvement or decline, were selected to complete the following two classification tasks (Pace and Duration). The example in Table 3 illustrates these cases. The sentence in the first row has a value for pace (‘slowly’) and duration (‘1 year to more years’), unlike the sentence in the second row that is classified as ‘permanent’. The output is a dataset of 5,174 sentences annotated according to the different features of HES model.

Consistency check. In this task, our objectives are: (a) clean any inconsistencies that may arise from the classification and (b) produce metrics that allow the selection of the best HES among the recommended annotations. First, as some sentences were annotated with the same HES, we deleted repeated combinations of ‘condition + sentence + HES’, leaving us with a total of 3,635 sentences. Next, we proceed to verify that the combination of features forms a coherent HES. For example, inconsistent combinations may have a pace annotation such as ‘fast’ while duration indicates a long recovery ‘from 6 months TO 1 year’. We rely on the pace and duration features to remove incoherent HES combinations.

The classification task generates one or more recommended HES; therefore, we provide metrics to select the best statement. We use an association rule learning method to identify how likely it is for a combination of HES features to represent a health condition. Firstly, we calculate how frequently the combination of health conditions and HES features appear in the dataset, and its support value (See Equation 1). Then we calculate how often the combination of *health condition and HES is valid*, its confidence value as shown in Equation 2. Table 3

shows the list of recommended HES for ‘Chronic obstructive lung disease’ ranked by confidence. The ultimate output is a set of recommended Health Condition Evolution Statements (HES) linked to their corresponding SNOMED CT concept. A total of 1,324 SNOMED CT concepts have one or more recommended HES.

$$\begin{aligned} \{HES\} &\Rightarrow Condition \\ \{Type, Pace, Duration\} &\Rightarrow Condition \\ support = P(Type \cap Pace \cap Duration) &= \frac{\text{number of predictions containing T, P and D}}{\text{total number of predictions}} \end{aligned} \tag{1}$$

$$\begin{aligned} confidence(HES \Rightarrow condition) &= \frac{supp(HES \cap condition)}{supp(HES)} \\ confidence &= \frac{\text{number of inferences containing HES and condition}}{supp(\text{number of inferences containing HES})} \end{aligned} \tag{2}$$

Table 3: HES best confidence value

SNOMED cept	Con- SNOMED Identifier	HES	Conf.	Sentence	Source
Chronic obstructive lung disease (disorder)	13645005	DECLINE SLOWLY FROM 1 YEAR TO MORE YEARS	0.0036	Although COPD is a progressive disease that gets worse over time, MAYO COPD is treatable.	
Chronic obstructive lung disease (disorder)	13645005	PERMANENT	0.0034	There’s currently no cure for chronic obstructive pulmonary disease (COPD), but treatment can help slow the progression of the condition and control the symptoms.	NHS

4.3 Knowledge completion

The data collected in Step 4.1 has limited coverage of SNOMED CT concepts. Therefore, in this task, we take advantage of SNOMED CT taxonomy and analyse the relationships and attributes of a given concept with the aim of finding similar concepts that could share the same HES. The objective is to identify patterns and create propagation rules, thus guiding an automatic HES expansion from SNOMED CT concepts with HES to other SNOMED CT concepts without HES, as illustrated in Figure 1.

The logic model of SNOMED CT taxonomy includes components that represent two types of relationships [12]:

1. **Subtype relationship.** This is the most used relationship and is known as “*is a*” relationship or hierarchical relationship because they form the hierarchies in SNOMED CT. This means that the clinical detail of a concept increases with the depth of the hierarchies. For example, ‘Elbow fracture’ \rightarrow *is a* \rightarrow ‘Fracture of upper limb’.
2. **Attribute relationship.** This relationship contributes to the definition of the source concept by associating it with defining characteristics. The characteristics are called attributes and are specified by (a) the *relationship type* and (b) the *value* provided by the destination of the relationship. For example, ‘Diabetes mellitus’ attribute \rightarrow *is* ‘Finding site (attribute)’ and its value is \rightarrow ‘Structure of endocrine system (body structure)’.

Using these SNOMED CT taxonomy definitions, we follow a structured process to find patterns and derive generalised rules of propagation; in what follows, we enumerate the steps taken:

1. Manually select a source concept and analyse its features: the number of parents, attributes, and values.
2. Analyse if the features of the source concept (with HES) are shared by other concepts (without HES).
3. If identical or similar relationships (subtype or attributes) are found, then build a general query using SNOMED CT Expression Constraint Language (ECL) [13] and retrieve all concepts sharing the identified relationships.
4. Select a number of concepts from the results in the previous step and manually verify that the results share the same HES.

In what follows, we describe each of the rules created using the SNOMED CT features⁵. Table 4 presents an overview of the rules and examples.

Table 4: Propagation rules details.

Rules	General Rule	Example (ECL syntax)
Rule 1	All descendants of administrative related concept	<<120646007 Antibody screen (procedure)
Rule 2	All immediate descendants of a source concept	<! 23406007 Fracture of upper limb
Rule 3	All target concepts that share two or more attributes similar to source concept	(*):([1..1]363698007 Finding (attribute)) =<<955009 Bronchial structure (body structure) ,116676008=4532008 AND ...
Rule 4	All target concepts with one attribute and direct children or the source concept	(102482005 Growing pains (finding)) OR <!102482005 Growing pains (finding)):([1..1]363698007 Finding (attribute)) =<<66019005 Limb structure (body structure))
Rule 5	All target concepts with same source parents OR source is parent AND similar attributes	((<<197480006 Anxiety disorder (disorder)) OR <<21897009 Generalized anxiety disorder (disorder)):([1..1]363714003 Interprets (attribute))=285854004 Emotion (observable entity))
Rule 6	All target concepts with two or more similar parents	(<!111273006 Acute respiratory disease (disorder) AND <!32398004 Bronchitis (disorder) AND <!128482007 Acute inflammatory disease (disorder))

- **Rule 1.** The hypothesis is that concepts describing administrative procedures do not affect people’s health. For example, ‘Antibody screen’ is an administrative procedure, thus, we annotated it as ‘UNAFFECTED’. The same applies to its descendants.
- **Rule 2.** The hypothesis is that target concepts with a direct ‘is a’ relationship inherit the source’s HES. For example, ‘Elbow fracture’ *is a* ‘Fracture of upper limb’ and therefore inherits the source’s HES.
- **Rule 3.** In this case, the target concept with two or more attributes similar to the source concept inherits the HES. For example, ‘Acute bronchitis (disorder)’ shares its HES with concepts with similar attributes (e.g. Finding site, Associated morphology and Clinical course).

⁵ The term ‘source concept’ is used to refer to a SNOMED CT concept that already has a HES annotation, and ‘target concept’ to refer to a SNOMED CT concept that has no HES assigned.

- **Rule 4.** Same as Rule 3, but restricted to one attribute descendants. For instance, ‘Growing pains’ has one attribute ‘Finding site’.
- **Rule 5.** In this case, a target concept with the same number of parents or similar attributes inherits the source concept HES.
- **Rule 6.** Here, a target concept with two or more similar parents as the source concept inherits the HES. This rule does not take into account concepts with one parent because the retrieved concepts are general.

The recommendations generated as a result of the classification and the knowledge completion task can be presented to experts to support the construction of the KG.

4.4 Human-in-the-loop

In order to scale up the construction of the health evolution KG and build high-quality data, it is imperative to include domain experts in the loop; therefore, the last step of the pipeline focuses on capturing human knowledge (see Figure 1).

This knowledge can be captured in three ways, by providing experts with (a) a list of recommended HES for each condition or (b) with a list of recommended target concepts that can share the same HES as the source concept; thus, they can assess the most accurate option swiftly. Also, experts can (c) build a new HES according to their best judgement.

We provide experts with a tool that reflects the options described above. The first interface displays the name of a condition and the list of candidate statements obtained in the knowledge components extraction step 4.2. Experts’ task is to select the ‘Correct’ HES according to their best judgement. The second interface uses the responses generated in the previous interface and the output from the knowledge completion step 4.3. The tool displays a condition (source), the HES that was selected as ‘Correct’ in the first interface and the recommended conditions that could share the given source HES condition. When there is no recommendation available, experts can use a third interface and input a new HES using the different elements of the statement (type, pace, duration).

The final Knowledge Acquisition pipeline’s output is a curated Health Evolution Statement (HES) database linked to its corresponding SNOMED CT concept. The KA pipeline is reproducible, and all the resources are available in the repository⁶.

5 Knowledge Graph

In order to make the newly created database available in a structured and machine-readable format, we built a Knowledge Graph following the HECON Ontology model [9]. The Health Condition Evolution Ontology is a formal model representing the evolution of health events over time. Each HES in the curated database is linked to a SNOMED CT concept identifier; likewise, each SNOMED CT concept could be linked to one or more HES. The KG also stores data that represents the relationships between the data sources (MAYO Clinic and NHS

⁶ Repository: <https://github.com/albamoralest/Health-Condition-Evolution-database>

England) and the process used to generate the annotation (knowledge component extraction, knowledge completion or HITL). This information supports the reasoning on the evolution of health conditions over time and the identification of ongoing health issues from EHR. The KG was built using SPARQL Anything [3], and it can be queried using SPARQL. Extended documentation, sample queries and the KG are available in repository⁷.

6 Evaluation

In Section 4, we presented a complete knowledge acquisition pipeline to build a database of health evolution information. This pipeline included components that extract knowledge automatically from web sources and capture domain experts’ knowledge. In what follows, we present the results of the user study carried out to evaluate the overall approach and answer the research questions stated in Section 2.

Evaluation settings. To conduct the user study, we used the tool described in the Human-in-the-loop step 4.4. We invited medical students, interns, nurses, general practitioners, paramedics and first responders who are knowledgeable on how health events (medical procedures, health conditions, diseases) evolve. Seven people agreed to participate, each participant with a different level of expertise, as shown in Table 5. We divided the user study into two parts and adapted the HITL tool accordingly. For the first part, participants annotated the same randomly selected list of SNOMED CT concepts taken from the Knowledge components identification output. In the second part, participants annotated whether a target SNOMED CT concept shares the same HES as the source concept. For both parts, participants indicate whether the HES is correct or not using a five-category Likert scale: Incorrect, Partially correct, Neither correct nor incorrect, Partially correct and Correct.

Table 5: Participants by level of expertise

Total	Expertise	Current role	Specialisation
1	Research	EU project Officer	Palliative Care for Cancer Patients
1	Doctor	Trainee doctor	Psychiatry
2	Nurse	Nurse	Respiratory
		Nurse practitioner	Minor illnesses in a GP surgery
3	Student	Intern	Gynecology
		3rd year undergraduate student	Medicine
		3rd year PhD student	Clinical medicine research

6.1 Results

In what follows, we present the evaluation of the research questions formulated in Section 2.

RQ1. Can health condition evolution statements be extracted automatically from descriptions in natural language?

To answer RQ1, we (a) evaluate the feasibility of the task and (b) measure the accuracy of the recommendations. We present participants with the first interface of our HITL tool and ask them to annotate as many concepts as possible in 30 minutes. Participants had to indicate their level of familiarity with a

⁷ <https://github.com/albamoralest/Health-Condition-Evolution-database>

given concept (familiar, partially familiar or unfamiliar) and whether the HES is correct or not using the five-category Likert scale. Also, they could input a new HES according to their best judgment.

Table 6, columns ‘Part 1’, display the total number of annotations by participant. The results show that participants were able to use the recommendations and the HITL tool to annotate an average of 47 conditions in 30 minutes. Also, they generated, on average, seven new HES.

Turning now to the accuracy of the recommendations, the classification process generates one or more recommended HES per condition; therefore, to measure the number of relevant HES, we calculate Precision@k. It can be seen from the data in Table 7 that the system was able to provide useful recommendations in more than half of the cases (median precision@8 of 0.56). These results show that the extraction of HES is an achievable task.

Table 6: RQ1 - Number of annotated SNOMED concepts per participant

	Part 1		Part 2	
	SNOMED concepts annotated	New HES generated	Source concepts	Target concepts annotated
P1	60	25	18	70
P2	47	0	32	143
P3	31	6	29	126
P4	67	0	62	284
P5	30	3	27	117
P6	53	6	37	162
P7	26	6	29	126
<i>Aug</i>	<i>45</i>	<i>7</i>	<i>33</i>	<i>147</i>

Table 7: RQ1 - HES Precision@k per participant

Precision @ k	1	2	3	4	5	6	7	8
P1	0.20	0.28	0.31	0.36	0.38	0.39	0.39	
P2	0.37	0.49	0.57	0.62	0.65	0.65	0.66	
P3	0.31	0.41	0.49	0.53	0.57	0.57		
P4	0.31	0.42	0.46	0.50	0.53	0.55	0.56	0.56
P5	0.26	0.30	0.37	0.42	0.47	0.49		
P6	0.28	0.34	0.39	0.42	0.42			
P7	0.23	0.30	0.38	0.38	0.38	0.38		
<i>Median</i>	<i>0.28</i>	<i>0.34</i>	<i>0.39</i>	<i>0.42</i>	<i>0.47</i>	<i>0.52</i>	<i>0.56</i>	<i>0.56</i>

RQ2. Can we use ontological knowledge to derive new health condition evolution statements automatically?

To answer RQ2, we use the recommendations generated by the knowledge completion step (see Figure 1, step 3) and measure the number of annotations participants produce in 30 minutes. We provided participants with the second interface of our HITL tool and asked them to indicate whether a target SNOMED CT concept shares the same HES as the source concept. Similar to the evaluation in RQ1, participants should answer using a five-category Linkert scale. The source concept sample is constituted from the SNOMED CT concepts annotated as ‘Correct’ and ‘Partially correct’ in the first part of the study and the concepts for which the participants provided a new HES.

As shown in Table 6, columns ‘Part 2’, each participant was able to review an average of 33 source concepts and annotate 147 target concepts. In comparison with results in RQ1, where participants annotated an average of 45 HES, what stands out is that the exploitation of SNOMED CT taxonomy *produces three times (312%) more recommended HES*. Further analysis of the results shows that participants reviewed a total of 1,028 recommendations; these HES were annotated as ‘Correct’ in half of the cases (501 conditions in total, see Table 8). These results demonstrate that the recommendations generated by the knowl-

edge completion method are useful in half of the cases to swiftly populate the part of SNOMED that was not originally covered by the web sources.

Table 8: RQ2 - Total Correct annotations per participant

	Rule1	Rule2	Rule3	Rule4	Rule5	Rule6	Total
P1	13	0	4	4	6	0	27
P2	13	0	7	13	19	0	52
P3	9	0	19	11	28	3	70
P4	15	0	61	20	73	0	169
P5	19	0	16	14	32	4	85
P6	11	0	25	8	20	0	64
P7	1	0	10	4	18	1	34
<i>Total</i>	81	0	142	74	196	8	501
<i>Proportion</i>	0.60	-	0.51	0.52	0.44	0.36	0.49

Table 9: RQ1 - Total annotations by familiarity

	Familiar	P. familiar	Unfamiliar
P1	27	10	23
P2	9	19	19
P3	17	4	10
P4	33	22	12
P5	13	6	11
P6	14	17	22
P7	10	13	3
<i>Total</i>	123	91	100
<i>Proportion</i>	0.39	0.29	0.32

RQ3. How sustainable is the proposed methodology to populate a database of health evolution?

To answer RQ3, we analyse the results obtained in RQ1 and RQ2 and give an account of the effort (expressed in ‘person-month’) required to populate the KG. We take as a reference the last edition of SNOMED CT, which included 353,567 concepts (published on January 31, 2020). On the one hand, in RQ1, one participant annotated an average of 50 correct concepts per hour (350 a day). If we only use recommendations generated by the knowledge component extraction step, it will take approximately four years and a half (55.20 person-months) to populate SNOMED CT. On the other hand, in RQ2, one participant annotated an average of 144 correct annotations per hour (1008 a day); it will take approximately a year and a half (19.16 person-months) to complete the task. We calculated only one person’s effort, yet experts could perform the task simultaneously. For instance, with seven experts (emulating our user study) and the effort required in RQ1, the task will take approximately eight months (7.88 months). Likewise, the task is reduced to approximately three months (2.73 months) if considering the effort in RQ2 and seven participants. From these results, we can conclude that the approach is sustainable.

In addition, we evaluate the inter-rater reliability. We use Krippendorff’s alpha coefficient [6] (applicable to missing data, various samples and different measures) and obtain an agreement of 0.4685. Although data in Table 9 indicate that participants were somehow familiar with 7 out of 10 concepts, the agreement result reflects the difficulty of finding participants with shared specialised expertise (as shown in Table 5).

7 Conclusions

In this paper, we presented a knowledge acquisition methodology to build a database of health evolution information. The pipeline implementation included automatic knowledge components such as text classification and completion. It also includes a Human-in-the-loop step to complete the methodology and obtain knowledge from domain experts.

The main goal of this paper was to fill the knowledge gap of resources regarding health condition evolution and recovery time. This study has found that

extracting health evolution statements (HES) from natural language is possible. The results confirm that the recommendations facilitate the capture of knowledge from experts. Furthermore, exploiting SNOMED CT features accelerates the production of recommendations, hence the coverage of SNOMED CT.

A key strength of this research was the inclusion of the Human-in-the-loop module. The results of our user study show that including domain experts as part of the methodology accelerates the construction of the KG. More importantly, it ensures the capture of their valuable and accurate knowledge. With these results, we fill a gap in the literature and provide structured data on health evolution.

References

1. Abu-Elkheir, M., Hassanein, H.S., Oteafy, S.M.: Enhancing emergency response systems through leveraging crowdsensing and heterogeneous data. In: 2016 Int Wireless Communications and Mobile Computing Conference (IWCMC). IEEE (9 2016), <http://ieeexplore.ieee.org/document/7577055/>
2. Alfattni, G., Peek, N., Nenadic, G.: Extraction of temporal relations from clinical free text: A systematic review of current approaches. *Journal of Biomedical Informatics* (8 2020). <https://doi.org/10.1016/j.jbi.2020.103488>
3. Daga, E., Asprino, L., Mulholland, P., Gangemi, A.: Facade-X: An Opinionated Approach to SPARQL Anything. In: Volume 53: Further with Knowledge Graphs. IOS Press (2021), <http://oro.open.ac.uk/78973/>
4. Glicksberg, B.S., Oskotsky, B., Thangaraj, P.M., Giangreco, N., Badgeley, M.A., Johnson, K.W., Datta, D., Rudrapatna, V.A., Rappoport, N., Shervey, M.M., Miotto, R., Goldstein, T.C., Rutenberg, E., Frazier, R., Lee, N., Israni, S., Larsen, R., Percha, B., Li, L., Dudley, J.T., Tatonetti, N.P., Butte, A.J.: PatientExploreR: an extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model. *Bioinformatics* (11 2019). <https://doi.org/10.1093/bioinformatics/btz409>
5. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* p. 20 (2007)
6. Krippendorff, K.: *Computing krippendorff's alpha-reliability* (2011)
7. Morales Tirado, A.C., Daga, E., Motta, E.: Effective use of personal health records to support emergency services. In: *EKAW 2022* (9 2020)
8. Morales Tirado, A.C., Daga, E., Motta, E.: Reasoning on health condition evolution for enhanced detection of vulnerable people in emergency settings. In: *Proceedings of the 11th KCAP Conference*. p. 9–16. ACM (12 2021)
9. Morales Tirado, A.C., Daga, E., Motta, E.: HECON Health: Condition Evolution Ontology. In: *5th Workshop on Semantic Web solutions for large-scale biomedical data analytics* (May 2022)
10. NHS Digital services: National requirements for SNOMED CT, <https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct>
11. Olex, A.L., McInnes, B.T.: Review of Temporal Reasoning in the Clinical Domain for Timeline Extraction: Where we are and where we need to be. *Journal of Biomedical Informatics* (6 2021). <https://doi.org/10.1016/j.jbi.2021.103784>
12. SNOMED CT International: *SNOMED CT Starter Guide* (2017)
13. SNOMED CT International: *Expression Constraint Language - Specification and Guide* (2022), <http://snomed.org/doc>