The Care2Report System: Automated Medical Reporting as an Integrated Solution to Reduce Administrative Burden in Healthcare

Lientje Maas
Utrecht University
j.a.m.maas@uu.nl

Mathan Geurtsen
Utrecht University
m.p.geurtsen@uu.nl

Florian Nouwt
Utrecht University
f.r.nouwt@uu.nl

Stefan Schouten
Utrecht University
s.f.schouten@uu.nl

Robin van de Water
Utrecht University
r.p.vandewater@uu.nl

Sandra van Dulmen
Nivel (Netherlands institute
for health services research)
s.vandulmen@nivel.nl

Fabiano Dalpiaz
Utrecht University
f.dalpiaz@uu.nl

Kees van Deemter
Utrecht University
c.j.vandeemter@uu.nl

Sjaak Brinkkemper
Utrecht University
s.brinkkemper@uu.nl

Abstract

Documenting patient medical information in the electronic medical record is a time-consuming task at the expense of direct patient care. We propose an integrated solution to automate the process of medical reporting. This vision is enabled through the integration of speech and action recognition technology with semantic interpretation based on knowledge graphs. This paper presents our dialogue summarization pipeline that transforms speech into a medical report via transcription and formal representation. We discuss the functional and technical architecture of our Care2Report system along with an initial system evaluation with data of real consultation sessions.

1. Introduction

Electronic medical records (EMRs) are designed to improve communication among providers within and between healthcare organizations regarding the collection, use and storage of patient information. Thereby, the EMR facilitates guideline compliance and decision support [1]; these tasks have become increasingly relevant, as medicine has become more evidence-based and care becomes more standardized for quality improvement. As a result of these developments, adequate recording and documentation of findings and decisions have become more relevant over time.

However, thorough documentation comes with notable negative side effects such as a high administrative burden on care providers, a well-recognized problem in many healthcare disciplines, e.g., general practice, home care, trauma surgery, medical specialities [2, 3, 4, 5]. A recent time-motion study in the United States shows that first year residents spend 43% of their time interacting with the EMR [6], at the expense of time for direct patient care. Further, it is estimated that in the Netherlands alone, administration tasks in long-term care require over 100,000 full-time positions, costing over 5 billion euros per year [7]. In the United States, administrative tasks are estimated to consume 13.5% of physicians' time, valued at 15.5 billion dollars [8]. Clearly, a more efficient way of reporting is required.

The use of speech recognition in medical reporting to reduce documentation time has been studied extensively. Most studies focus on dictation for reporting after a consultation [9], although this is not often used in practice [10]. Furthermore, initial work has been performed to record patient care provider dialogues and automatically extract clinical meaning, e.g., [11, 12].

Our Care2Report (C2R) research program (`www.care2report.eu`) aims at automated medical reporting. The enabler is a novel integration of state-of-the-art speech and action recognition technology with knowledge-based summarization of the care provider-to-patient interaction. The aim is to automatically generate a consultation report that the care provider checks before uploading to the EMR.

Our computational approach to summarization separates linguistic interpretation from sentence generation. *Interpretation* encodes the core information content of the dialogue as a set of semantic triples following the Resource Description Framework (RDF) data model [13, 14]. *Reporting* uses natural language generation [15] techniques to convert a selected subset

URI: https://hdl.handle.net/10125/64184
978-0-9981331-3-3

of triples into natural language text. In the long run, this two-stage approach allows us to obtain an accurate representation of the consultation by combining the information in the dialogue (represented as RDF triples) with the objects and actions recognized by other system components (RDF triples, too). Practically, we build on powerful existing language technology software, both for triple extraction from text (e.g., [16]) and for text generation from triples (e.g., [17]). The use of these rule-based techniques is especially appropriate in a medical (and hence fault-critical) context because it maximizes the designer's control over the summaries produced, avoiding the opacity that still tends to characterize most machine learning approaches.

Paper organization. Sec. 2 discusses related work. Sec. 3 describes the C2R system and its design principles, while its architecture is presented in Sec. 4. The recording, interpretation and report generation stages of the process are discussed in Sec. 5–7. Sec. 8 provides a preliminary evaluation of our prototype. We draw conclusions and discuss future work in Sec. 9.

2. Related work

We review existing systems that assist medical reporting. Due to space limitations, we do not present research on each step of automated medical reporting.

Speech recognition in medical reporting has been studied for a long time with a focus on dictation for reporting after a consultation [9]. Despite its potential, dictation is seldom used [10] as it requires care providers to adjust their way of working, evoking resistance. As an alternative, we aim to automate reporting without affecting the regular working procedures.

Automated medical reporting was attempted in a project at MIT, aiming to capture patient-care provider interactions in one modality through text, speech, and dialogue processing [11]. We go further and integrate multimodal input, providing more complete, reliable and unambiguous information about the consultation. Chiu *et al.* developed a speech recognition system for transcribing medical conversations, reaching a word accuracy of 81% [18]. Most errors are conversational and unrelated to medical terms; they perform well on crucial medical utterances (92% precision, 86% recall). This system was utilized for classification of symptoms from transcripts of medical conversations, based on a semi-supervised learning approach [19, 12].

The OR Black Box is a multi-channel data recording system for prospective analysis of audiovisual and patient-related data in a real-life operating room (OR) [20]. This system captures structured and unstructured data to identify event patterns and team awareness in the OR [21]. The outputs of the OR Black Box could be used as a rich input source when applying automated medical reporting to the surgical domain.

Another related approach is the generation of textual summaries of temporal clinical data from, e.g., physiological signals [22]. Results show that it is possible to generate effective textual summaries of complex continuous and discrete temporal clinical data. These systems are developed for informative purposes; presenting the data effectively supports clinical decision making. Decision support is not the envisioned goal of our work; instead, we focus on accurate reporting of consultations for administrative purposes.

3. The Care2Report system

We are developing an integrated hardware and software platform for automated medical reporting. In this platform a non-intrusive device with camera, microphone and sensor technology is combined with state-of-the-art speech and video analysis and advanced semantic interpretation through knowledge graphs. Multimodal input is provided: audio, video, and sensor modalities from medical measuring instruments (also: healthcare domotics). Speech recognition allows transformation of medical dialogues to text, action recognition captures examinations and treatments, and sensor data provide results of medical measurements. Background information from medical guidelines and patient history can be employed for a more accurate interpretation of the data. Combining and interpreting all information enables automatic preparation of a medical report that is checked by the care provider before uploading to the EMR. Summarizing, our proposed process consists of three stages (Figure 1):

1. *Recording.* Preprocessing and transformation of audio, video and sensor input data from medical consultations into text using speech and action recognition technology and a domotics analyzer.
2. *Interpretation.* Formal representation of conversations, measurements and treatments based on multimodal inputs combined with semantic technology.
3. *Report generation.* Generation of medical reports based on medical domain practices, followed by report completion, checking by the care provider, and uploading through a generic EMR-interface.

Our main challenge is to integrate state-of-the-art multimodal recognition technology with knowledge representation and reasoning into one software platform. To achieve this, we designed a dialogue summarization pipeline connecting several linguistic tools to go from

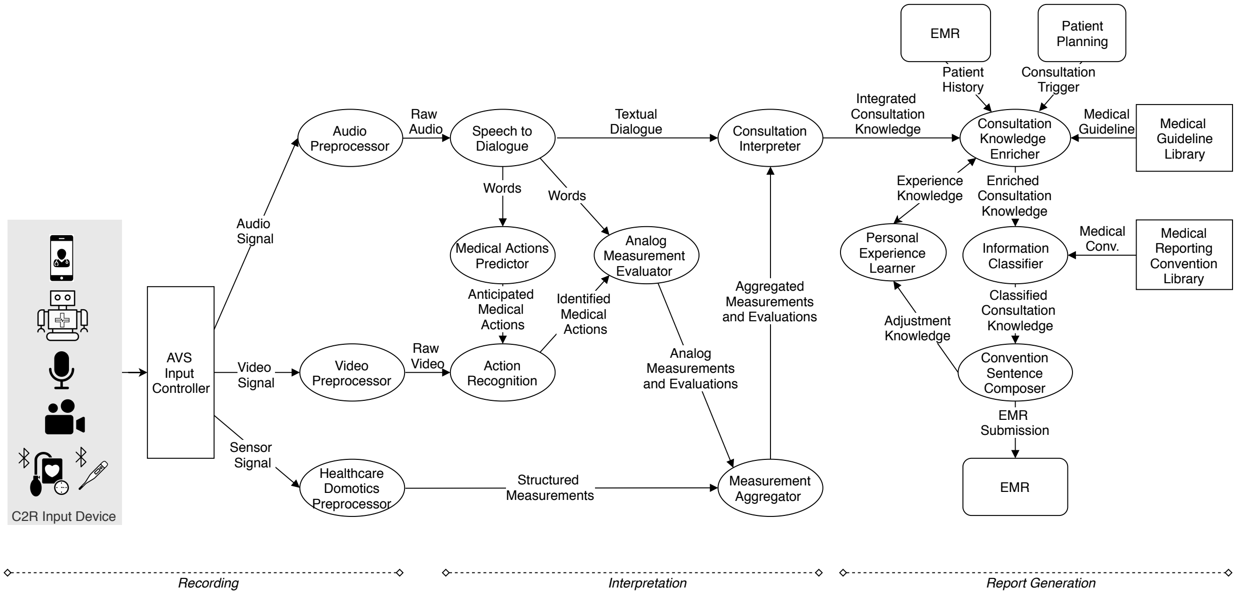


Figure 1. Functional architecture of the Care2Report system with components based on microservices.

speech via transcript and formal representation to a medical report. We will discuss design principles in Sec. 3.1 and then describe the pipeline in Sec. 3.2.

3.1. Design principles

In cooperation with care providers in various disciplines, we identified design principles for creating a system with a high level of usability and extensibility in the context of modern healthcare practice. These are presented in Table 1 and will be described now. First and foremost, the system must be simple in use and non-intrusive. This means it must not interfere with the current working procedures of care providers (P_1). In principle, the care provider must be able to press a record button, get an indication per input modality to check whether recording is enabled (P_2), and proceed with the consultation as usual. Meanwhile, the medical report is to be generated in real time (P_3). Note that this is not a transcript, but a concise and complete summary of the consultation (P_4). The care provider remains responsible for the final content of the report. He or she must be able to edit the report if necessary before approving to upload it to the EMR (P_5). Of course, the aim is to keep the necessity for adaptation by the care provider to a minimum. To facilitate this, the system will include a learning component to learn from personal procedures and adaptations by care providers (P_6).

Further, the system must be widely usable as a solution for the widely recognized problem of administrative burden (P_7). Therefore it must be designed to support multiple languages, but also multiple healthcare disciplines (e.g., general practice, home care, specialists in hospitals). These disciplines may hold specific terminology, medical guidelines and reporting conventions that vary over the disciplines.

Privacy is key for automated medical reporting, and any solution has to be aligned with the rights and obligations laid out in the General Data Protection Regulation (GDPR) [23]. We aim for integrated security and privacy by architecture that is compliant by design through a modular architecture (P_8). Data recorded during consultations is not stored longer than necessary to generate the medical report.

Table 1. Design principles of the C2R system

P_1	No interference with current working procedures.
P_2	Simple input control of all modalities.
P_3	Report generation in real time.
P_4	Complete and concise summaries of consultations.
P_5	Care provider must check and possibly edit reports.
P_6	System learns from corrections by care provider.
P_7	Applicable for multiple healthcare disciplines and languages.
P_8	Compliant with privacy regulations.

3.2. Dialogue summarization pipeline

For the complete process of automated reporting, we designed a dialogue summarization pipeline connecting six components. This pipeline goes from speech to transcript via formal representation to report text.

1. *Transcription of speech.* Speech recognition technology is currently flourishing. Our open architecture interfaces with an off-the-shelf package, which is currently handled by Google Cloud Speech-to-Text service.

2. *Recognition of concepts and relations.* Python-Frog [16] is used for linguistic annotation of Dutch text to extract concepts and relations from the textual dialogue. In addition, the tool FRED [24] is used for relation extraction.
3. *Storing and manipulating triples.* As will be discussed in Sec. 6.2, semantic triples (⟨subject, predicate, object⟩) are extracted from the free text to populate the knowledge graph [14]. Triples are stored and managed with Stardog [25].
4. *Building ontologies.* Protégé [26] facilitates ontology development as the starting point for the knowledge graph that represents the medical consultation. Note that these are prebuilt ontologies as input to the system, thus not as a sequential step in the pipeline (see Sec. 6.2).
5. *Populating ontologies.* The ontology is populated with the extracted triples to form the desired knowledge graph. We are developing an (rule-based) algorithm to match triples to the ontology, as will be discussed in Sec. 6.2.
6. *Generating natural language text.* From the knowledge graph, report text is generated by means of sentence plans in a natural language generation component of our system based on the NaturalOWL system [17]. This component is described in Sec. 7.

Most components rely on off-the-shelf technology, either used as-is or with our extensions. The pipeline is work-in-progress and some changes, especially technological, may be applied in future research.

4. System architecture

Modularity and openness are the key architectural dimensions for the C2R platform, both from a technical perspective as well as from the medical application perspective. We aim at a robust, plug-and-play architecture that is able to deal with the interfacing intricacies of the libraries of medical ontologies and guidelines. In the following, we provide two complementary views on our architecture. The *functional architecture* structures the individual features in a modular fashion according to the functional usage. The *technical architecture* is meant for the technological structures, formats, applications, interactions (networking), and data storage.

4.1. Functional architecture

The functional components of the C2R system are based on a microservice architecture [27], represented in Figure 1. Splitting large components (e.g., audio, video, and domotics analyzer) into smaller ones solves interdependency complications while maintaining a loosely coupled system. Each functional component has a predefined input and output set, which allows for simple configurability and future extensibility.

The recording stage concerns the input setting and control of the modalities with subsequent preprocessing for error and noise correction. The interpretation stage transforms speech to a textual dialogue, recognizes medical objects from video, and transforms sensor signals to measurement data. The Action Recognition Analyzer obtains words that describe medical actions in order to execute their identification in the video stream. The Analog Measurement Evaluator and the Measurement Aggregator combine the data from the modalities in order to determine the execution of a medical treatment and the determinaton of a health status. For example, the doctor measures blood pressure and says “Your blood pressure is very good.” without mentioning the actual blood pressure values of 120 over 80. The latter are automatically transferred into the system. All this information is combined through triple extraction by the Consultation Interpreter and ontology population. During report generation, the consultation knowledge is enriched by EMR patient history data, and the consultation trigger from the patient planning. The medical guideline enables to populate the ontology by matching medical guideline concepts to consultation data (see Sec. 6.2). The Information Classifier selects relevant triples for the report based on the conventions of the medical domain available in a library. The Convention Sentence Composer generates the sentences for the report, which can be improved by the Personal Experience Learner.

More details on the feature realization of the architecture are described in Sec. 5-7. Figure 1 shows the envisioned architecture and that not all microanalyzers are developed yet. Interpretation and report generation are implemented for a selected domain (ear infections, see Sec. 6.2). In future versions of the prototype, we will implement advanced action recognition, such as the identification of emotions from facial expressions (video) and intonation of voice (audio), utilization of patient history information and medical guidelines, and a learning component.

4.2. Technical architecture

The technical architecture of the C2R system contains a collection of linguistic software components (see Figure 2). The Web-app Client interacts with the Server Controller at the Server Cluster, where the

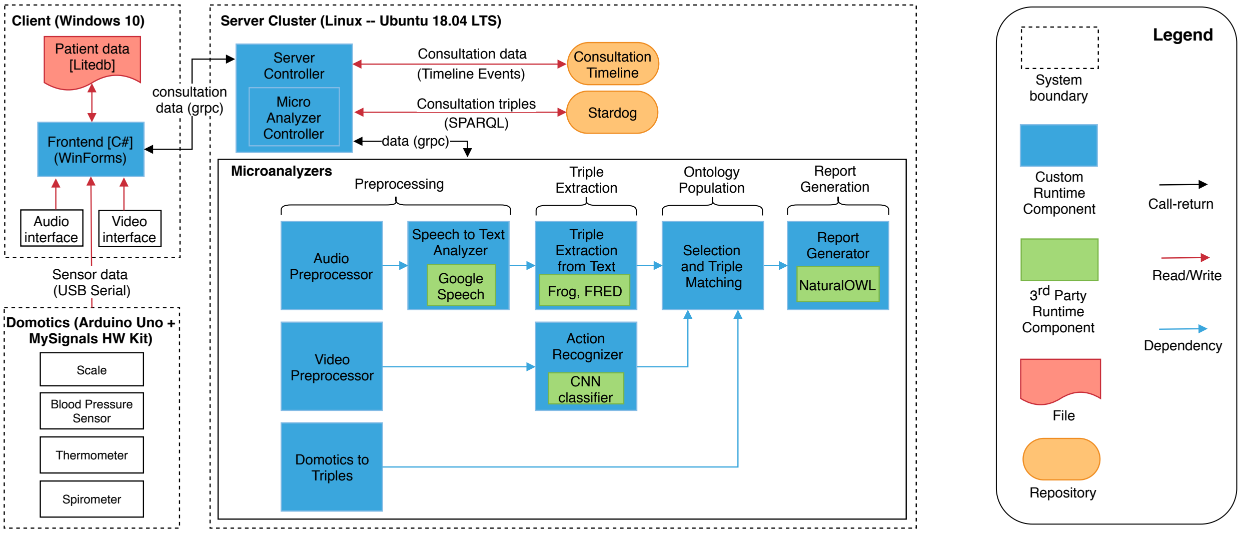


Figure 2. Technical architecture of the Care2Report system.

Microanalyzer Controller controls all analysis processes and ensures that all execution constraints are satisfied. Naturally, the Client manages the audio and video interface. The Domotics section supports multiple medical measurement sensors. We have chosen the MySignals kit [28] for this C2R implementation (see Sec. 5). The report generation operates via Windows Forms for portability to EMRs on Windows, Linux and MacOS platforms. Patient data is temporally stored locally on the client, which enables that reporting sessions can always be read back.

The Server Controller manages connectivity with the Client and receives audio, video and domotics data. These initial data are put on the consultation timeline (see Sec. 6.1) and the Microanalyzer Controller invokes the needed microanalyzer for further processing. Microanalyzers are then triggered on the basis of when the input and output are needed; these dependencies are shown in Figure 2 by blue arrows.

The audio data follow the path through the preprocessing. After transcription by the Speech to Text Analyzer, linguistic open source software components are utilized for annotated triple extraction by the Triple Extraction Analyzer. The video data, after preprocessing, are analyzed by the Action Recognizer based on transfer learning applied to a convolutional neural network (CNN). The audio and video streams join at the Selection and Triple Matching Analyzer, which is based on language independent matching of ontology concepts. By selecting the appropriate language lexicon the Report Generator can produce the sentences in either Dutch or English (see Sec. 7).

Domotics data is fed into the Domotics to Triples Analyzer to be put into the Selection and Triple Matching Analyzer, where it is immediately selected (because of the objectivity of the data) to be put in designated data sections of the report, and possibly as support data in the report sentences. The Report Generator analyzer then finalizes the report, which the Server Controller sends back to the Client.

5. The recording stage

In the recording stage, the medical consultation is recorded in multiple modalities: audio, video, and sensor modalities. The aim is to develop an integrated device for high-quality multimodal recognition, but the architecture will be designed such that it is independent of input technology. Input devices are monitored by the audio-video-sensor input controller (AVS Input Controller) for modality control and quality check.

Off-the-shelf software is used for speech and action recognition; although imperfect, this was the best trade-off for a first prototype. Preprocessing arrangements will be studied to enhance the recognition, e.g., by optimizing speech rate and silence length [29].

In the current prototype, medical sensor input is implemented via the MySignals Hardware kit, a shield (expansion board) for the Arduino platform [28]. Using this kit allows the development of a proof-of-concept with medical instruments without concerns about specific rules and protocols that would rise with medical devices that are currently used in practice.

6. The interpretation stage

The observational signals from the multimodal input need to be transformed into meaningful information.

Utilizing multimodal input enables enhanced event recognition in one modality by using information from another modality. This is realized by a data structure in the form of a timeline, a key artifact in our approach (Sec. 6.1). To interpret the raw data recorded in a consultation, we map it to a knowledge graph to enhance semantic reasoning and querying [13]. As mentioned, RDF triples are extracted and mapped to a prebuilt ontology. The task of extracting information and storing it as part of an ontology is best described as ‘ontology population’ [30]. Our approach to formal knowledge representation is described in more detail in Sec. 6.2.

6.1. Medical consultation timeline

To integrate the information from the multimodal sources, a so-called *medical consultation timeline* is generated to log a medical consultation (e.g. measurements, treatments, diagnosis). This is illustrated in Figure 3. The situations that stem from the occurrence of events are stored along with their time range, enabling enhanced event recognition utilizing multimodal inputs. The integration of knowledge will lead to the complete *modeled consultation knowledge*, from which a report is generated based on reporting conventions in the specific medical domain. Figure 3 illustrates the SOEP convention (often used by general practitioners (GPs) to record medical input): Subjective, Objective, Evaluation and Plan (also indicated as SOAP: Subjective, Objective, Assessment and Plan) [31].

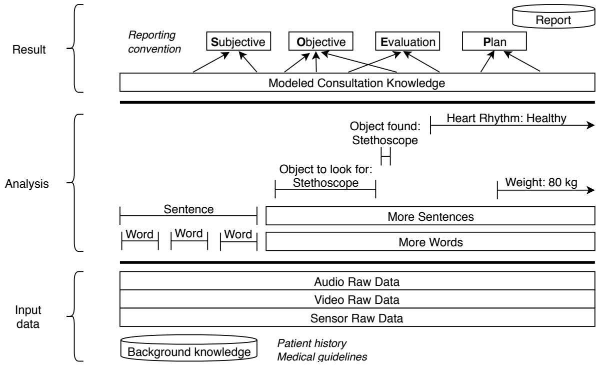


Figure 3. Medical consultation timeline (to be viewed from bottom to top).

6.2. The patient medical graph

In our previous work, we introduced the Patient Medical Graph (PMG) as a formal representation of medical consultations [32]. Figure 4 presents the reference graph that is the starting point for the PMG. This knowledge graph represents human anatomical entities (black), signs and symptoms (blue), medical observations (green), diagnoses (red) and treatment plans (orange). The human anatomy is the kernel structure for the knowledge graph, built based on existing ontologies such as the Foundational Model of Anatomy [33]. Signs and symptoms associated with specific anatomical entities follow from medical guidelines. Interpretation of observations during a consultation assigns values and other characteristics to these signs and symptoms (e.g., presence, duration). Diagnosis and treatment plan are determined by the care provider and also follow from interpretation of the data that is recorded during the consultation. Note that the PMG serves as an internal representation of the consultation knowledge; we do not intend to build an ontology containing complete medical knowledge.

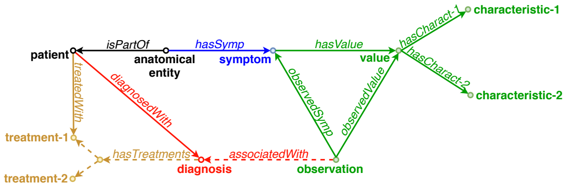


Figure 4. Reference graph for the PMG.

We are currently building an ontology comprising the human anatomy and medical signs and symptoms. This is completed for a selected domain: medical problems related to the ear. Starting with a small domain provides the opportunity to study and test our methods by specification of the ontology and it enhances data interpretation due to specific background knowledge. To populate this ontology with the observations, diagnosis and treatment plans, interpretation of the consultation is required. As mentioned, semantic triples (⟨subject, predicate, object⟩) are extracted from the textual dialogue. Grammatical annotation of dialogue sentences is used to extract concepts and relations for triple creation. Triples are then matched to the ontology by the Consultation Knowledge Enricher (Figure 1).

Matching arbitrary triples to an ontology is generally not an easy task, however given the currently limited scope of the ontology, a breadth-first-search (BFS) based algorithm has been developed, shown in Algorithm 1. The ontology contains distinct classes of entities, for example anatomy and symptom, which always relate in the same way (predicate p would be “hasSymptom” in this case). The algorithm takes a set of unmatched triples T (bare triples without URIs; words or word groups extracted from natural language) and creates matched triples based on a set of subjects S, a fixed predicate p and a set of objects O. Both subjects and objects contain tuples consisting of an entity URI and a set of natural language nouns, the latter of which is used to match unmatched entities to the correct URI. In order to do this, T is regarded as an undirected graph where each triple forms an edge between subject and object. Each BFS search starts at an object entity and attempts to find a connection with a subject entity. During this search, only edges with predicates in V are used to prevent irrelevant relations from being considered. Once a pair is found, a triple with predicate p is added to the output set T' and the search is ended. A drawback of this method, however, is that if multiple subjects relate to the same object, only one of these will be found. The search could be continued by replacing the `break` on line 15 with a `continue`, but this increases the risk of finding irrelevant pairs. This may be mitigated by limiting the depth of the search, but this has not been studied yet.

Algorithm 1: BFS Triple Matching

```
input : $T$ = set of input triples $(sub, pred, obj)$
        $S$ = set of possible subjects, which are tuples of the form $(u, N)$
          where $u$ is the uri and $N$ a set of natural language nouns
        $p$ = output triple predicate uri
        $O$ = set of possible objects, same tuple format as for $S$
        $V$ = set of allowed search predicates
output: A new set of triples $T'$
1  $T' \leftarrow \emptyset$;
2  for $(obu, obN) \in O$ do                    // Search from each object
3      $visited \leftarrow \emptyset$;
4      $queue \leftarrow$ makeQueue($obN$);       // New queue with initial
         content
5      while not empty($queue$) do
6          $ent \leftarrow$ dequeue($queue$);
7          if $ent \in visited$ then   // Do not visit entity twice
8              continue
9          add $ent$ to $visited$;
10         $found \leftarrow false$;
11         for $(sbu, sbN) \in S$ do // Check if entity is subject
12             if $ent \in sbN$ then
13                 add triple $(sbu, p, obu)$ to $T'$; // Triple found
14                 $found \leftarrow true$;
15                 break
16         if $found$ then                   // Do not explore further
17             break
18         for $(sub, pred, obj) \in T$ do          // Find connected
             entities
19             if not $pred \in V$ then // Check allowed relations
20                 continue
21             if $ent = sub$ then
22                 enqueue($queue$, $obj$);
23             else if $ent = obj$ then
24                 enqueue($queue$, $sub$);
```

6.3. Structure of medical consultations

The interpretation of unconstrained dialogue text can be problematic, but fortunately detailed knowledge about the context of the utterances is available. Medical consultations follow a general structure: opening, history taking, physical examination, evaluation, treatment recommendations and closing [34]. During history taking and physical examination the presence of signs and symptoms is determined, which are evaluated to determine a diagnosis and treatment plan. To support this process, medical guidelines are designed that improve the structure of care [35]. These guidelines provide valuable information that can be utilized in the C2R system to enhance interpretation. The structured, computer-interpretable model of the relevant medical guideline is retrieved and the text can be interpreted in the context of the appropriate segment of this guideline [36]. This helps to resolve ambiguity, detect missing values, and filter noisy input.

Selecting information In generating medical reports, it is crucial to identify which information from the consultation is relevant to report. To gain some initial insights in this reporting relevance, seven transcripts of video recordings from medical consultations concerning otitis externa and otitis media acuta (external and middle ear infection) are analyzed. These recordings are part of a study by Nivel (Netherlands institute for health services research) and Radboudumc to improve GP communication [37, 38]. The transcripts were presented to a GP, who manually wrote a report of each consultation as it would be written for the EMR, following the SOEP convention (Sec. 6.1).

Each transcript is compared to both the corresponding medical guideline and the SOEP report. The transcripts are split in utterances and each utterance is classified twice: in a guideline category and a reporting category. We define an utterance as a speech segment to which precisely one classification category can be assigned. This segment may vary in length from a single word to a lengthy sentence. A sentence is considered one segment if it conveys only one thought or relates to one category of interest. The *guideline categories* are presented in Table 2. The *reporting categories* are Subjective, Objective, Evaluation, and Plan. An utterance is classified in a reporting category if it contains information that contributes to the information that is (manually) reported in that category; otherwise it is classified as 'Not reported'.

Table 2. Guideline classification categories.

Diagnosis	Policy
- history taking	- counseling
- physical examination	- advice
- evaluation	- non-pharmacological treatment
	- pharmacological treatment
	- schedule follow-up
	- referral

Note. If an utterance does not belong to any category, it is classified as 'Other'.

Based on these categories, we analyzed the course of medical consultations in relation to the medical guideline and report (see Figure 5 for an illustrative example). For brevity, we provide a general description of our findings. In general, there is a one-to-one mapping from the guideline categories to the reporting categories, where history takes maps to Subjective, physical examination to Objective, evaluation to

Evaluation, and all subcategories of policy in Table 2 to Plan. Therefore, the guidelines can support both interpretation and automated classification of information in one of the SOEP categories. However, not every utterance that is classified in a reporting category also has a guideline classification. This mainly concerns parts of the history taking that are not explicitly stated in the guideline but are reported in the Subjective category (e.g., discussion of unrelated symptoms, effects of previous treatment plans). Further, if the medical professional gives counseling and advice, this is reported in the Plan category as 'Counseling and advice' without further substantive information. This implies that in the process of automated reporting, it is only required to recognize the *occurrence* of counseling and advice and not the content of it.

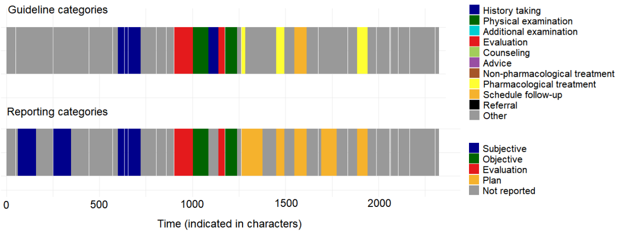


Figure 5. Example of classification of utterances during a medical consultation.

For each consultation, we calculated the proportion of content that included information to be included in the report, see Table 3. Around 60% to 85% of the dialogue does not contain information that requires reporting, and this percentage is higher for longer consultations. This part of the dialogue may nevertheless contain crucial information on the outcome of the consultation, e.g., in personalized care. Although discussion of feelings, relationship building, patient participation and shared decision making are not necessarily reported in the EMR, these factors are important to align the provided care with personal needs and possibilities.

In future work, we will study methods to filter out relevant information based on a.o. knowledge from the structure of medical consultations discussed above.

Table 3. Proportion of the medical consultation that contains information for the reporting categories.

	Length	S	O	E	P	Not rep.
1	4478	8.8%	2.0%	0.8%	15.8%	72.3%
2	2324	13.6%	6.3%	5.9%	15.7%	58.5%
3	4932	10.9%	1.5%	0.0%	2.1%	85.5%
4	3344	20.1%	6.5%	1.4%	11.4%	59.9%
5	5847	7.4%	2.5%	1.8%	12.8%	75.4%
6	4500	7.7%	1.5%	4.3%	5.5%	81.0%
7	7449	8.2%	2.7%	0.0%	6.1%	83.0%
avg.	4696	11.0%	3.3%	2.0%	9.9%	73.7%

Note. Length is indicated in number of characters. Proportions are calculated as the summed length of the utterances classified in each category divided by the total length of the transcript.

Findings will enable efficiency gains of the algorithm that is used for the ontology population.

7. The report generation stage

In the report generation stage, natural language report text is composed. Natural language generation is a fast-growing area of research and an emerging technology in many domains, including healthcare [39, 15, 22]. The report generation stage in the C2R system is based on the open source NaturalOWL software [17]. NaturalOWL is developed to generate fluent textual descriptions of individuals or classes in an OWL ontology. For this end, the ontology must be annotated with linguistic resources: natural language names and sentence plans. Natural language names describe the representation of OWL individuals, i.e., subjects and objects. Sentence plans describe word orders to create a sentence from a single fact (triple).

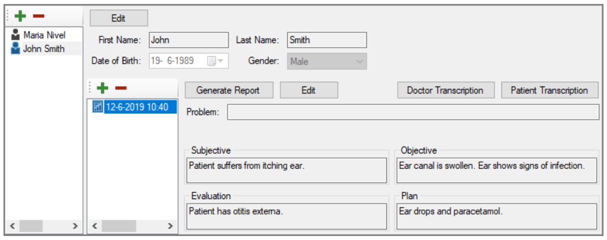
The inputs in the processing stages are the ontology, its linguistic resources, and the individual to describe. First NaturalOWL determines which facts in the ontology are relevant (Content Selection) and then it orders these facts (Text Planning). Next, each fact is converted into a short standalone sentence (Lexicalisation), where some of these sentences are aggregated into longer sentences (Aggregation). The remaining sentences are connected by replacing some of the nouns in the later sentences by references to their mention in the earlier sentences (Referring Expression Generation). Finally, NaturalOWL converts its internal representation of the sentences to text and adds punctuation and capitalization (Surface Realization).

Extending NaturalOWL to support Dutch medical report text generation required the following changes: adding static Dutch resources (articles, pronouns, prepositions, etc.); adding internal representations for dynamic Dutch resources (nouns, adjectives, verbs); extending the Surface Realization for Dutch, e.g., if an adjective is preceded by an indefinite article in Dutch, always use the base form instead of the inflected form. In addition to these linguistic changes, the structure of the code has been refactored to allow the future addition of new languages.

8. Preliminary system evaluation

We are building a large corpus of recordings of simulated and real medical consultations. Corresponding reports will be written by medical professionals and constitute our gold standard. These can be split into a training set and a test set, enabling training and evaluation of the C2R system using text comparison metrics such as BLEU (bilingual evaluation understudy) [40]. Also, we plan to evaluate, based on human judgments, the correctness, completeness, and fluency of the generated reports.

We present some preliminary reflections on the performance of the system, although a detailed evaluation is left to future work. As input for this early evaluation, we used 8 (flawless) transcripts of medical consultations regarding ear problems. We did not use audio files due to the unavailability of high-quality recordings from our medical partners. Test data included transcripts of the real consultations of Sec. 6.3 and of simulated consultations. An example result from processing a simulated consultation transcript is shown in Figure 6. It shows the feasibility of classifying the generated sentences according to the SOEP categories.



First Name: John
Last Name: Smith
Date of Birth: 19- 6-1989
Gender: Male

Problem:

Subjective
Patient suffers from itching ear.

Objective
Ear canal is swollen. Ear shows signs of infection.

Evaluation
Patient has otitis externa.

Plan
Ear drops and paracetamol.

Figure 6. Example of a report generated by the C2R prototype based on a consultation transcript.

Considering the difficulty of the task, the generated reports showed promising results. While 2 transcripts were summarized correctly (i.e., containing all and only relevant facts from the input transcript), other 6 transcripts showed mixed results, for important facts were missing from the report. The root of this problem lies in the intrinsic difficulty of triple matching (where the words of the transcript are matched with concepts in the ontology). In many cases, medical concepts that are important for the subdomain at hand, and all the words that can be used to express these concepts, had to be added manually; in some cases this also necessitated adding new sentence plans to the report generator. We plan to address this by harvesting synonyms from existing medical ontologies such as SNOMED [41], avoiding the need to proceed by trial and error.

Moreover, due to the limitations in the state of the art of natural language interpretation, negative statements can sometimes appear positive (e.g., “I don’t suffer from headaches” may accidentally be parsed into a triple that asserts that the patient does suffer from headaches). In the future, we aim to address this problem by performing consistency checks on the triples extracted, trying to detect triples that are inconsistent with other information about the patient (e.g., other triples, patient history, etc.)

9. Conclusion and future work

We presented our vision for automated medical reporting and an overview of the C2R system that is under development. The stages in the summarization process are discussed on the basis of the dialogue summarization pipeline that we designed. In principle, this pipeline is domain-independent, enabling implementation in domains other than healthcare, e.g., for police reports, customer services.

In future work, we will study methods to extract information from a medical consultation based on its structure (Sec. 6.3) and on methods to populate the *PMG* with this information (Sec. 6.2). Further, we will focus on architecting a privacy warranting modular integration for high-quality multimodal recognition. Although more research is required to realize fully automated reporting, our first results indicate that this ambitious vision is achievable.

Acknowledgements This work has been developed in collaboration with the Software Project of the department of Information and Computing Sciences of Utrecht University. We thank the student teams Kiéli and KettleHawks, and we thank Milou Burgers for her support.

References

[1] P. Campanella, E. Lovato, C. Marone, L. Fallacara, A. Mancuso, W. Ricciardi, and M. L. Specchia, “The impact of electronic health records on healthcare quality: a systematic review and meta-analysis,” *The European Journal of Public Health*, vol. 26, no. 1, pp. 60–64, 2015.

[2] S. Woolhandler and D. U. Himmelstein, “Administrative work consumes one-sixth of US physicians’ working hours and lowers their career satisfaction,” *International Journal of Health Services*, vol. 44, no. 4, pp. 635–642, 2014.

[3] J. F. Golob Jr, J. J. Como, and J. A. Claridge, “The painful truth: The documentation burden of a trauma surgeon,” *Journal of Trauma and Acute Care Surgery*, vol. 80, no. 5, pp. 742–747, 2016.

[4] A. J. E. de Veer, K. de Groot, M. Brinkman, and A. L. Francke, *Administratieve druk: méér dan een kwestie van tijd (Administrative burden: more than a matter of time)*. Utrecht: Nivel, 2017. Original doc. in Dutch.

[5] P. Mishra, J. C. Kiang, and R. W. Grant, “Association of medical scribes in primary care with physician workflow and patient experience,” *JAMA internal medicine*, vol. 178, no. 11, pp. 1467–1472, 2018.

[6] K. H. Chaiyachati, J. A. Shea, D. A. Asch, M. Liu, L. M. Bellini, C. J. Dine, A. L. Sternberg, Y. Gitelman, A. M. Yeager, J. M. Asch, *et al.*, “Assessment of inpatient time allocation among first-year internal medicine residents using time-motion observations,” *JAMA internal medicine*, 2019.

[7] M. Hanekamp, *Administratieve taken langdurige zorg kosten jaarlijks € 5 miljard (Administrative tasks in long-term care cost € 5 billion yearly)*. Utrecht: Berenschot, 2016. Original doc. in Dutch.

[8] S. Woolhandler, T. Campbell, and D. U. Himmelstein, “Costs of health care administration in the United States and Canada,” *New England Journal of Medicine*, vol. 349, no. 8, pp. 768–775, 2003.

[9] S. Ajami, “Use of speech-to-text technology for documentation by healthcare providers,” *The National medical journal of India*, vol. 29, no. 3, p. 148, 2016.

[10] E. Luchies, M. Spruit, and M. Askari, “Speech technology in Dutch health care: A qualitative study,” in *BIOSTEC*, vol. 5, pp. 339–348, 2018.

[11] J. G. Klann and P. Szolovits, “An intelligent listening framework for capturing encounter notes from a doctor-patient dialog,” *BMC medical informatics and decision making*, vol. 9, no. 1, p. S3, 2009.

[12] A. Rajkomar, A. Kannan, K. Chen, L. Vardoulakis, K. Chou, C. Cui, and J. Dean, “Automatically charting symptoms from patient-physician conversations using machine learning,” *JAMA internal medicine*, vol. 179, no. 6, pp. 836–838, 2019.

[13] G. Antoniou, P. Groth, F. van Harmelen, and R. Hoekstra, *A Semantic Web Primer*. Cambridge, Massachusetts: The MIT Press, 3 ed., 2012.

[14] K. Rohloff, M. Dean, I. Emmons, D. Ryder, and J. Sumner, “An evaluation of triple-store technologies for large data stores,” in *OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”*, pp. 1105–1114, Springer, 2007.

[15] A. Gatt and E. Krahmer, “Survey of the state of the art in natural language generation: Core tasks, applications and evaluation,” *Journal of Artificial Intelligence Research*, vol. 61, pp. 65–170, 2018.

[16] A. v. d. Bosch, B. Busser, S. Canisius, and W. Daelemans, “An efficient memory-based morphosyntactic tagger and parser for Dutch,” *LOT Occasional Series*, vol. 7, pp. 191–206, 2007.

[17] I. Androutsopoulos, G. Lampouras, and D. Galanis, “Generating natural language descriptions from OWL ontologies: the NaturalOWL system,” *JAIR*, vol. 48, pp. 671–715, 2013.

[18] C.-C. Chiu, A. Tripathi, K. Chou, C. Co, N. Jaitly, D. Jaunzeikare, A. Kannan, P. Nguyen, H. Sak, A. Sankar, *et al.*, “Speech recognition for medical conversations,” *arXiv preprint arXiv:1711.07274*, 2017.

[19] A. Kannan, K. Chen, D. Jaunzeikare, and A. Rajkomar, “Semi-supervised learning for information extraction from dialogue.,” in *Interspeech*, pp. 2077–2081, 2018.

[20] M. G. Goldenberg, J. Jung, and T. P. Grantcharov, “Using data to enhance performance and improve quality and safety in surgery,” *JAMA surgery*, vol. 152, no. 10, pp. 972–973, 2017.

[21] J. J. Jung, P. Jüni, G. Lebovic, and T. Grantcharov, “First-year analysis of the Operating Room Black Box study.,” *Annals of Surgery*, 2018.

[22] F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes, “Automatic generation of textual summaries from neonatal intensive care data,” *AI*, vol. 173, no. 7-8, pp. 789–816, 2009.

[23] S. Shastri, M. Wasserman, and V. Chidambaram, “The seven sins of personal-data processing systems under GDPR,” *USENIX HotCloud*, 2019.

[24] A. Gangemi, V. Presutti, D. R. Recupero, A. G. Nuzzolese, F. Draicchio, and M. Mongiov, “Semantic Web Machine Reading with FRED,” *Semantic Web*, vol. 8, no. 6, pp. 873–893, 2017.

[25] Stardog Union, “Stardog 6 Manual: The knowledge graph platform for the enterprise,” 2019. https://www.stardog.com/.

[26] H. Knublauch, R. W. Fergerson, N. F. Noy, and M. A. Musen, “The Protégé OWL plugin: An open development environment for semantic web applications,” in *ISWC*, pp. 229–243, Springer, 2004.

[27] S. Klock, J. M. van der Werf, J. P. Guelen, and S. Jansen, “Workload-based clustering of coherent feature sets in microservice architectures,” in *ICSA*, pp. 11–20, 2017.

[28] Libelium Comunicaciones Distribuidas S.L., Zaragoza, Spain, *MySignals SW eHealth and Medical IoT Development Platform Technical Guide*, 2019.

[29] S. Kwon, S.-J. Kim, and J. Y. Choeh, “Preprocessing for elderly speech recognition of smart devices,” *Computer Speech & Language*, vol. 36, pp. 110–121, 2016.

[30] P. Buitelaar and P. Cimiano, *Ontology learning and population: bridging the gap between text and knowledge*, vol. 167. Ios Press, 2008.

[31] S. Cameron and I. Turtle-Song, “Learning to write case notes using the SOAP format,” *Journal of Counseling & Development*, vol. 80, no. 3, pp. 286–292, 2002.

[32] L. Maas, A. Kisjes, I. Hashemi, F. Heijmans, F. Dalpiaz, S. van Dulmen, and S. Brinkkemper, “Automated medical reporting: From multimodal inputs to medical reports through knowledge graphs,” *Paper presented at the KR4HC workshop, Conference on Artificial Intelligence in Medicine, Poznań, Poland*, 2019, June.

[33] C. Rosse and J. L. Mejino Jr, “A reference ontology for biomedical informatics: the Foundational Model of Anatomy,” *Journal of Biomedical Informatics*, vol. 36, no. 6, pp. 478–500, 2003.

[34] D. W. Maynard and J. Heritage, “Conversation analysis, doctor–patient interaction and medical communication,” *Medical education*, vol. 39, no. 4, pp. 428–435, 2005.

[35] M. Lugtenberg, J. Burgers, and G. Westert, “Effects of evidence-based clinical practice guidelines on quality of care: a systematic review,” *BMJ Quality & Safety*, vol. 18, no. 5, pp. 385–392, 2009.

[36] M. Peleg, “Computer-interpretable clinical guidelines: a methodological review,” *Journal of Biomedical Informatics*, vol. 46, no. 4, pp. 744–763, 2013.

[37] J. Houwen, P. L. Lucassen, H. W. Stappers, W. J. Assendelft, S. van Dulmen, and T. C. olde Hartman, “Improving GP communication in consultations on medically unexplained symptoms: a qualitative interview study with patients in primary care,” *Br J Gen Pract*, vol. 67, no. 663, pp. e716–e723, 2017.

[38] M. C. Meijers, J. Noordman, P. Spreeuwenberg, S. van Dulmen, *et al.*, “Shared decision-making in general practice: an observational study comparing 2007 with 2015,” *Family practice*, 2018.

[39] E. Reiter and R. Dale, *Building natural language generation systems*. Cambridge university press, 2000.

[40] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th annual meeting on ACL*, pp. 311–318, ACL, 2002.

[41] K. A. Spackman, K. E. Campbell, and R. A. Côté, “SNOMED RT: a reference terminology for health care.,” in *Proceedings of the AMIA annual fall symposium*, p. 640, AMIA, 1997.