Maas, L., Kisjes, A., Hashemi, I., Heijmans, F., Dalpiaz, F., Dulmen, S. van, Brinkkemper, S. Automated Medical Reporting: From Multimodal Inputs to Medical Reports through Knowledge Graphs. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies – HEALTHINF, p. 509-514

Postprint version : 1.0
Journal website : https://www.scitepress.org/HomePage.aspx
Pubmed link :
DOI : 10.5220/0010261605090514

Automated Medical Reporting: From Multimodal Inputs to Medical Reports through Knowledge Graphs

Lientje Maas[1], Adriaan Kisjes[1], Iman Hashemi[1], Floris Heijmans[1], Fabiano Dalpiaz[1], Sandra van Dulmen[2], Sjaak Brinkkemper[1]

[1] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
[2] NIVEL (Netherlands Institute for Health Services Research), Utrecht, The Netherlands

Care providers generally experience a high workload mainly due to the large amount of time required for adequate documentation. This paper presents our visionary idea of real-time automated medical reporting through the integration of speech and action recognition technology with knowledge-based summarization of the interaction between care provider and patient. We introduce the Patient Medical Graph as a formal representation of the dialogue and actions during a medical consultation. This knowledge graph represents human anatomical entities, symptoms, medical observations, diagnoses and treatment plans. The formal representation enables automated preparation of a consultation report by means of sentence plans to generate natural language. The architecture and functionality of the Care2Report prototype illustrate our vision of automated reporting of human communication and activities using knowledge graphs and NLP tools.

1 Introduction

Care providers (CPs) are required to accurately report patient information. As a primary communication tool between CPs, medical records are necessary for good patient care. However, recording and maintaining patient medical information in the electronic medical record (EMR) is time-consuming. A more efficient way of reporting is required to cope with high workload in healthcare while preserving quality of the patient data.

To reduce documentation time, the use of speech recognition in medical reporting has been studied extensively. Recently, Chiu et al. developed a speech recognition system for transcription of medical conversations, reaching a word accuracy of 81.7% (Chiu et al., 2017). Most studies focus on dictation for reporting after a consultation (Ajami, 2016). However, dictation is only used by 1% of medical staff in the Netherlands (Luchies et al., 2018). Klann and Szolovits performed initial work to capture the patient - CP dialogue with speech recognition and automatically extract clinical meaning (Klann and Szolovits, 2009). Further, the project BabyTalk aimed to automatically generate textual summaries of temporal clinical data from physiological signals (Portet et al., 2009). Automated medical reporting is the visionary goal of our Care2Report (C2R) research program (see www.care2report.nl). To achieve this, state-of-the-art speech and action recognition technology are combined with semantic interpretation of data through knowledge graphs. This enables automatic preparation of a consultation report that is checked by the CP (and, if relevant, the patient) before uploading in the EMR. Our solution will substantially reduce administrative load and improve personal engagement in healthcare. Note that we do not provide decision support but solely report consultations.

This paper is organized as follows. The next section describes our approach to enable automated medical reporting. Section 3 provides more in-depth information about the formal representation of events and situations during medical consultations. Section 4 presents the architecture and functionality of the system that is under development. Finally, the status of our research and outlook is described in Section 5.
Kop 2 compact

[Figure 1]

2 Approach

Our main research challenge is to integrate state-of-the-art multimodal recognition technology with knowledge representation and reasoning into one software platform. Globally, the process consists of four stages (illustrated in Fig. 1):

1. Transformation of audio, video and sensor data from medical consultations into text using existing speech and action recognition technology.
2. Formal representation of situations, measurements and treatments based on multimodal input combined with semantic technology.
3. Generation of medical reports using conventions in specific medical domains.
4. Report completion, checking by CP, and uploading through a generic EMR-interface.

We develop a generic hardware and software platform with non-intrusive recording device with microphone, camera and sensor technology that performs optimal recognition of situations and actions. Sensor technology enables wireless connection with healthcare domotics, e.g., a thermometer. Multimodal input is provided: audio, video and sensor modalities. Speech recognition allows to transform medical dialogues to text, action recognition captures examinations and treatments, and sensor data provide results of medical measurements.

2.1 Multimodal Knowledge Integration

To interpret the raw data recorded during the medical consultation, we model it as a knowledge graph to enhance semantic reasoning and querying (Antoniou et al., 2012). We refer to all interpreted information from the consultation as consultation knowledge.

Although the interpretation of unconstrained dialogue text can be problematic, we are in the fortunate circumstance that detailed knowledge about the context of the utterances is available through so-called background knowledge. For most medical consultations, the condition for which the patient is treated is known and the corresponding medical guideline is employed for a more accurate interpretation (Peleg, 2013; Sutton and Fox, 2003). This helps to resolve ambiguity and cope with incomplete or noisy input. Access to the medical record of the patient is of similar use. Additionally, we exploit the large corpus of medical background knowledge that is available. Medical ontologies (SNOMED, ICD-10, LOINC) and large medical knowledge graphs (Drugbank, SIDER, AERS) are utilized to disambiguate the text. This is particularly helpful for cases where knowledge of the patient's condition is partially known or vague.

To integrate information from the multimodal sources, the C2R system constructs a so-called medical consultation timeline to log a medical consultation (e.g., measurements, diagnosis, treatments). The situations that stem from the occurrence of events are stored along with their time range, enabling enhanced event recognition by using multimodal inputs. For example, if a CP verbally announces that he or she is going to listen to a patient's heart (audio input), it can be foreseen that a stethoscope will be used (video input). The integration of inputs will lead to the complete modeled consultation knowledge in a knowledge graph populated by semantic triples ((subject, predicate, object)) (Rohloff et al., 2007), from which a report is generated.

3 Patient medical graph

Medical consultations follow a general structure: opening, history taking, physical examination, evaluation, treatment recommendations and closing (Maynard and Heritage, 2005). 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 formally represent the collected information, we define the *Patient Medical Graph (PMG)* as the knowledge graph of the patient's anatomy complemented with evaluated signs and symptoms with associated diagnosis and treatment plan.[1] The *PMG* serves as an internal representation of the consultation knowledge. An example of the *PMG* for a fictitious external ear infection (otitis externa) consultation is presented in Fig. 2. It consists of five subgraphs: *PMG* = PAG U PSG U POG U PDG U PTG. We will now formally define each subgraph (see also Tables 1 and 2) and illustrate with examples from Fig. 2.

[Figure 2]

The human anatomy is the starting point of the Patient Anatomy Graph (PAG), representing all human anatomical entities. The PAG knowledge graph is universal for each patient, apart from gender differences. Existing ontologies are used as reference, e.g., the Foundational Model of Anatomy (Rosse and Mejino Jr, 2003). The PAG is complemented with the Patient Symptom Graph (PSG), representing signs and symptoms associated with specific anatomical entities. Medical guidelines build the PSG by providing lists of signs and symptoms occurring in specific medical domains (Peleg, 2013). The Patient Observation Graph (POG) assigns values to the signs and symptoms based on observations during the medical consultation. The observations connect the values to a certain sign or symptom (e.g., observation-1 observes symptom pain with value 7/10), appearing as (green) triangles in the POG. Additional characteristics are also in the POG, such as the time of occurrence (e.g.,observation-1 of pain 7/10 has had duration 4 days). Next, the graph is complemented with the diagnosis made by the CP in the Patient Diagnosis Graph (PDG). Based on observations (green), the diagnosis otitis externa is given (red). Finally, we complement the graph with the Patient Treatment Graph (PTG) based on the interpreted treatment plan in the consultation. We consider any treatment in its broadest sense: not only medication, but also referral to a specialist or additional tests.

3.1 Populating the *PMG*

Complementing the PAG[PSG with the POG, PDG and PTG requires interpretation of the consultation. Observations from test scenarios indicate that the key parts in the consultation dialogue are typically uttered in short standard phrases. We aim to capture the medical dialogue through a library of linguistic patterns with placeholders. Medical guidelines are the starting point for identification of these patterns. The placeholders are filled in using part-of-speech tagging and dependency parsing in

[1] The PMG can be seen as the instance level (A-Box) of an ontology. Due to space limitations, we do not discuss the corresponding T-Box that defines the entity and relationship types.

combination with regular expressions, after which semantic triples are deduced. A similar method has been successfully used for automated evaluation of eligibility criteria for clinical trials (Milian et al., 2015).

3.2 Report Generation

From the populated *PMG*, a report of the consultation is generated. Medical reports generally contain short and simple sentences, which enhances automated generation. We are developing a natural language generation component of our system based on the NaturalOWL system (Androutsopoulos et al., 2013), illustrated in Fig. 3. Template sentence plans are specified for the relevant relations in the *PMG*. We will determine the requirements and conventions regarding medical reporting to identify information that is relevant to report and study filtering methods for report texts.

[Table 1] [Table 2]

The sentence plans consist of a sequence of slots along with information on how to fill those in. These plans lead to separate sentences, which are aggregated into longer ones based on rules. In addition, referring expressions are generated to improve readability. After the report is generated, the CP checks it for completeness and correctness.

4 Care2Report prototype

To realize our vision a prototype is under development that takes multimodal input and outputs a draft report. It transforms speech to text, recognizes medical objects from video, and transforms sensor signals to measurement data. Formal knowledge representation based on medical guidelines and sentence composition are implemented 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 e.g. the PAGUPSG and it enhances data interpretation due to specific background knowledge.

4.1 Architecture

The prototype is based on a microservice architecture (Klock et al., 2017). Splitting large unimodal analyzers (e.g., audio analyzer, video analyzer, and domotics analyzer) into smaller microanalyzers solves interdependency complications while maintaining a loosely coupled system. Each microanalyzer has a predefined input and output set, which allows for simple configurability and future extensibility. A microanalyzer controller controls the analysis process and ensures that all execution constraints are satisfied.

4.2 Input Analysis and Report Generation

The system contains a database with data structure in correspondence with the medical consultation timeline described in Section 2. Triples to populate the *PMG* are extracted from dialogue text using linguistic tools. Grammatical annotation of dialogue sentences is used to extract concepts and relations for triple creation. We envision more rigorous methods in the future as described in Section 3. Video analysis is used to identify movement of medical objects (e.g., a stethoscope) to indicate utilization by the CP. Healthcare domotics send data from medical measurements to the system via Bluetooth. The relevant input is added to the (prebuilt) PAG U PSG to form the complete *PMG* comprising the modeled consultation knowledge. A report is then generated based on sentence plans, following the procedure described in Section 3, which is developed for the ear domain. The stages of the process are illustrated in Fig. 3 for the ear infection example.

4.3 Evaluation

We are currently building a large corpus of data including recordings of both simulated and real medical consultations. Corresponding medical reports are written manually by medical professionals to compare with the automatically generated reports. The data can be partitioned into a training set and a test set, enabling training and evaluation of the system.

[Figure 3]

4.4 Technological Platforms

The front end of the system runs on the Windows UWP platform and is mainly written in C#. The back end runs primarily on .NET Core and is written in C#. The analyzers are written in Python, using gRPC for communication between services/modules. Google Cloud Speech-to-Text service transcribes the audio and linguistic annotation is handled by Python-Frog. For video analysis the OpenCV and the YOLO libraries are used. Medical guidelines are modeled in PROforma. Protégé facilitates ontology development and triples are stored and managed with StarDog.

5 Research-Outlook

So far, we presented our grand vision and the implementation of our basic ideas in the first C2R prototype. To reach our proposed objectives, we need to overcome several challenges i.a. in the development of a robust architecture that is independent of input technology, in the semantic interpretation of input that deviates between hospitals on terminology and procedures, and in striking a balance between required expressiveness and computational demands in constructing a formal representation of the transcriptions.

Our future research will focus on device integration for high-quality multimodal recognition (stage 1 in Fig. 1), on methods to build and populate the *PMG* (stage 2 in Fig. 1), and on methods to filter out irrelevant information from medical consultations (stage 3 in Fig. 1). Our preliminary research and results encouraged us that our ambitious goal of fully automated medical reporting is achievable.

Acknowledgements

We thank the students of the software project teams Ki´eli and KettleHawks, Marjan van den Akker, Lennart Herlaar and Sabine Molenaar for their support.

References

Ajami, S. (2016). Use of speech-to-text technology for documentation by healthcare providers. The National Medical Journal of India, 29(3):148–152.

Androutsopoulos, I., Lampouras, G., and Galanis, D. (2013). Generating natural language descriptions from OWL ontologies: the NaturalOWL system. JAIR, 48:671–715.

Antoniou, G., Groth, P., van Harmelen, F., and Hoekstra, R. (2012). A Semantic Web Primer. The MIT Press, Cambridge, Massachusetts, 3 edition.

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

Klann, J. G. and Szolovits, P. (2009). An intelligent listening framework for capturing encounter notes from a doctor-patient dialog. BMC Medical Informatics and Decision Making, 9(1).

Klock, S., van der Werf, J. M., Guelen, J. P., and Jansen, S. (2017). Workload-based clustering of coherent feature sets in microservice architectures. In ICSA, pages 11–20.

Luchies, E., Spruit, M., and Askari, M. (2018). Speech technology in Dutch health care: A qualitative study. In BIOSTEC, volume 5, pages 339–348.

Maynard, D.W. and Heritage, J. (2005). Conversation analysis, doctor–patient interaction and medical communication. Medical Education, 39(4):428–435.

Milian, K., Hoekstra, R., Bucur, A., ten Teije, A., van Harmelen, F., and Paulissen, J. (2015). Enhancing reuse of structured eligibility criteria and supporting their relaxation. Journal of Biomedical Informatics, 56:205–219.

Peleg, M. (2013). Computer-interpretable clinical guidelines: a methodological review. Journal of Biomedical Informatics, 46(4):744–763.

Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., and Sykes, C. (2009). Automatic generation of textual summaries from neonatal intensive care data. AI, 173(7-8):789–816.

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

Rosse, C. and Mejino Jr, J. L. (2003). A reference ontology for biomedical informatics: the Foundational Model of Anatomy. Journal of Biomedical Informatics, 36(6):478–500.

Sutton, D. R. and Fox, J. (2003). The syntax and semantics of the PROforma guideline modeling language. Journal of the American Medical Informatics Association, 10(5):433–443.

Maas, L., Kisjes, A., Hashemi, I., Heijmans, F., Dalpiaz, F., Dulmen, S. van, Brinkkemper, S. Automated Medical Reporting: From Multimodal Inputs to Medical Reports through Knowledge Graphs. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies – HEALTHINF, p. 509-514

Tables and figures

Figure 1: Flowchart of the process of automated medical reporting.

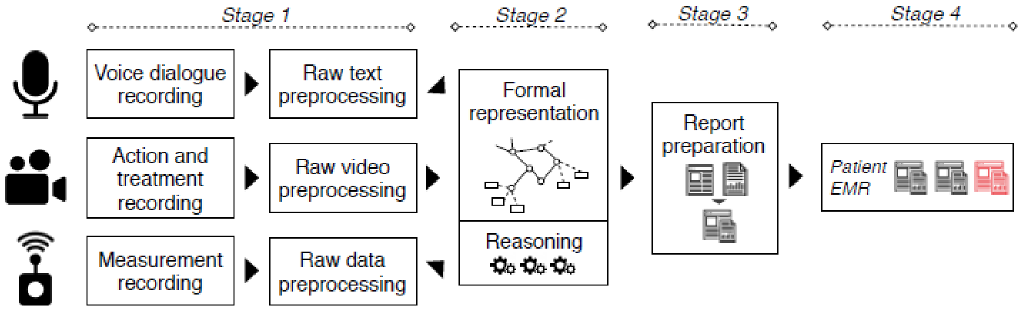

Figure 1: Flowchart of the process of automated medical reporting.

Figure 2: Excerpt of a PMG based on a consultation concerning otitis externa (external ear infection). Note that for explanation reasons the graph is colored for each of the five subgraphs.

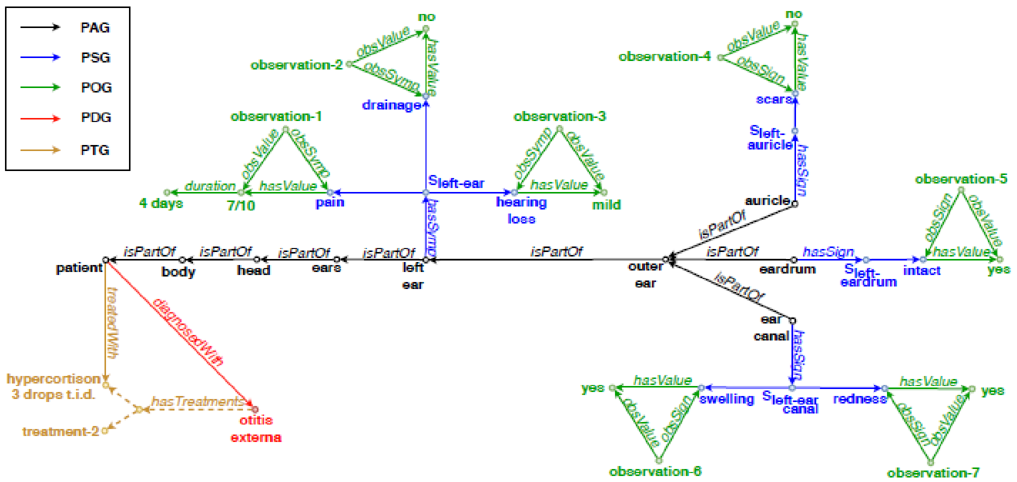

Table 1: Definition of sets required to define the PMG.

Set	Description	Set	Description
P	all patients	O	all medical observations
A	all anatomical entities of the human body	D	all medical diagnoses
S	all medical signs and symptoms	T	all medical treatments
V	all possible values to be assigned to $s \in S$		

Table 2: Formal definitions of the subgraphs comprising the *PMG*.

Graph	Vertices	Typed edges
PAG	A	$\{(a_1,a_2) \mid a_1,a_2 \in A \wedge a_1$ is a direct anatomical subpart of $a_2\}$
PSG	$A \cup S$	$\{(a,s) \mid a \in A \wedge s \in S\}$ i.e., all signs and symptoms of a
POG	$O \cup S \cup V$	$\{(o,s),(o,v),(s,v) \mid o \in O \wedge s \in S \wedge v \in V\}$ i.e., all observations
PDG	$P \cup D$	$\{(p,d) \mid p \in P \wedge d \in D\}$ i.e., all diagnoses for patient p
PTG	$P \cup T$	$\{(p,t) \mid p \in P \wedge t \in T\}$ i.e., all treatment plans for patient p

Maas, L., Kisjes, A., Hashemi, I., Heijmans, F., Dalpiaz, F., Dulmen, S. van, Brinkkemper, S. Automated Medical Reporting: From Multimodal Inputs to Medical Reports through Knowledge Graphs. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies – HEALTHINF, p. 509-514

Figure 3: Example showing part of a transcription of the CP - patient dialogue (left), the resulting PMG (middle) and the sentence plan for report generation (right).