RUTA:MED – Dual Workflow Medical Speech Transcription Pipeline and Editor

Arturs Znotins ✉[1], Roberts Dargis[1], Normunds Gruzitis[1], Guntis Barzdins[1], and Didzis Gosko[2]

[1] Institute of Mathematics and Computer Science (IMCS), University of Latvia
Raina bulvaris 29, Riga, LV-1459, Latvia
{arturs.znotins,roberts.dargis,normunds.gruzitis}@lumii.lv
[2] Riga East University Hospital (REUH)
Hipokrata iela 2, Riga, LV-1079, Latvia

Abstract. In the medical domain various approaches are used to produce examination reports and other medical records. Depending on the language-specific technology support, the type of examination, the size of the hospital or clinic, and other aspects, the reporting workflow can range from completely manual to (semi-)automated. A manual workflow may completely depend on the doctor itself or may include a transcriptionist centre in the loop. In an automated workflow, the transcriptionist centre is typically replaced by an automatic speech recognition (ASR) system. While the latter approach is well suited for high resource languages where word error rate (WER) is as low as 5–10%, for less resourced languages a dual approach combining automated transcription with the support from a transcriptionist centre may be more suited. In this paper, we present a platform that supports both workflows simultaneously. The RUTA:MED platform currently includes an ASR pipeline for the less resourced Latvian language, and it is being deployed and tested at several hospitals and clinics in Latvia. The platform can be adopted for any other language, and it emphasizes that WER is only one of the performance indicators in case of medical transcription.

Keywords: speech recognition · medical transcription · post-editing

1 Introduction

We present a software platform for automated transcription of medical dictations, developed by a leading Latvian language technology research group at IMCS in cooperation with the largest hospital in Latvia (REUH). Our initial focus was on digital imaging reports (computed tomography, magnetic resonance, etc.), but the scope has been extended to include other areas (histopathology, gastroenterology, etc.).

In Latvia, medical reports so far were produced completely manually. Several hospitals and clinics employ transcriptionist centres to assist the production of medical reports. Due to a constant growth of diagnostic and laboratory examinations, clinicians and patients often have to wait for the reports up to several

days if they are not considered urgent. The manual transcription of a dictation takes about a day of the total waiting time, and the doctor must verify the final report. Services of transcriptionist centres are expensive, and smaller hospitals and clinics often cannot afford it. Thus, a more efficient approach and technology infrastructure is clearly needed to significantly reduce the dependency on transcriptionist centres. However, many doctors would prefer to keep transcriptionist centres (in a reduced capacity) as an option.

RUTA:MED[3] is a platform for automated medical transcription, which addresses these issues. It is not only a speech transcription pipeline consisting of an ASR system and automatic post-processing modules – it also provides an integrated editor (text-audio alignment) and supports both workflows: it greatly facilitates post-editing by the doctor him/her self, but it also allows to submit the dictation and its draft transcript to a transcriptionist centre any time.

RUTA:MED builds on our previous work on a general-domain speech transcription system for Latvian [5], which we have adapted and extended for the medical domain [1, 2]. The platform is showcased with the less resourced Latvian language and the situation in Latvia regarding the usage of automatic speech recognition (ASR) and text processing technologies in the medical domain, however, the Latvian case is not unique. RUTA:MED can be adopted for other languages by replacing the language-specific ASR and text post-processing components. We argue that word error rate (WER) – the standard metric used to evaluate ASR systems – is only one of the performance indicators in case of medical transcription. The integrated editor and the dual workflow which involves a transcriptionist centre is relevant not only for less resourced languages but also in case of high resource languages (like English, French and German) to reduce the workload of doctors, especially in case of complex examinations.

The Latvian-specific resources and components of RUTA:MED are described in more detail by Dargis et al. [1] and Gruzitis et al. [2]. In this system demo paper, we focus on the whole platform and the dual workflow it supports. While the web-based UI and API of RUTA:MED, as well as its back-end components, are continuously being enhanced and extended, the platform is production-ready and is being used in a trial mode by several healthcare institutions in Latvia.

2 Related Work

Medical ASR products by Nuance, like PowerScribe One[4] for radiologists, are among the most widely used in case of high resource languages. Apart from ASR and text post-processing, they provide many other features and integration options. RUTA:MED has a comparatively narrow focus and is aimed at less resourced languages for which transcriptionist centres are still a relevant option.

Another well-known product – Trint[5] – represents a more general kind of speech-to-text transcription platforms. Like Trint, RUTA:MED editor provides

[3] https://med.ailab.lv/demo
[4] https://nuance.com/healthcare/diagnostics-solutions/
[5] https://trint.com

interactive text and speech alignment which facilitates the post-editing process, while RUTA:MED supports the specialised dual workflow in addition to the specialised ASR and post-processing components.

As for the ASR and post-processing part, we were inspired by the development of the Estonian ASR for Radiology [3], since Estonian is a similar-size language having comparable amount of language resources available, but again – RUTA:MED adds the dual workflow and interactive editing support.

3 RUTA:MED Platform

RUTA:MED is a web application that integrates an ASR and post-processing pipeline with the help of a task queue for scalability (see Figure 1). This allows for simple setup and integration with other web-based medical information systems.

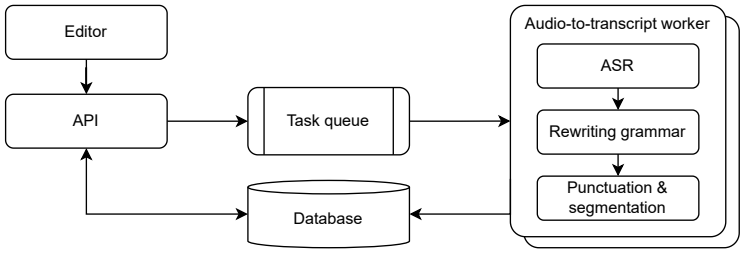

Fig. 1. Overall architecture of the RUTA:MED platform

3.1 Speech Recognition and Post-processing Pipeline

Before a transcription is available for post-editing, the submitted dictation goes through multiple processing stages. First, the audio is processed with an ASR system adapted to the medical domain. The Latvian language model is trained on 1.5GB of plain-text, extracted from the REUH archive of medical reports. The pronunciation lexicon is automatically extracted from the same text corpus, manually extended with pronunciation of abbreviations, Latin terms, drug names, etc. Second, the automatic transcriptions are processed with a rewriting grammar to acquire concise reports, thus, minimise the post-editing work, and to interpret explicit and implicit voice commands. The grammar is implemented as a cascade of finite-state transducers using the OpenFST-based Thrax framework [4]. The final transcription post-processing step is punctuation restoration and text segmentation with a neural network model trained on the text corpus.

3.2 Integrated Editor

The key user-facing RUTA:MED element is the integrated audio-synchronized transcript editor (see Figure 2). The core of this web-based open-source editor

was developed in the H2020 project SELMA for the video subtitling purposes and is adapted in RUTA:MED for the post-editing of medical transcripts.

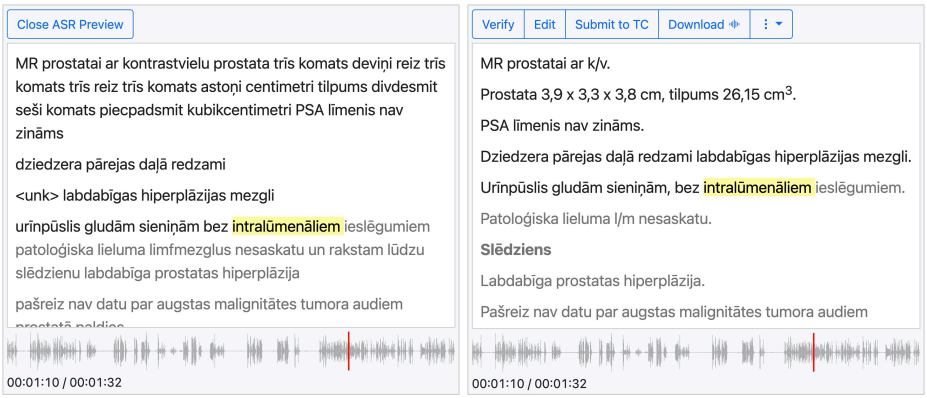

Fig. 2. RUTA:MED editor views: preview of the ASR output (left); the automatically post-processed transcript (right); the time-code synchronized cursor (highlighted)

The text is synchronised with the waveform. Both the text area and the waveform indicate the current progress. The audio timestamp is indicated by a vertical bar in the waveform and by a highlighted background in the the text area. The user can click anywhere in the text to navigate in audio or click anywhere in waveform to highlight the corresponding word/segment.

The editor provides basic text formatting functions, such as bold, italic, underline, sub- and superscript, numbered and bullet lists. Structural formatting (predefined fields of the report) is recognized in the post-processing step. The final text can be copied to a medical information system or a Word document.

3.3 Dual Workflow

The RUTA:MED workflow is designed for two user scenarios that can be switched on the fly. In the primary scenario, we expect that in many cases when a report is fluently dictated and hence has resulted in an overall accurate and clean transcription the doctor who was so far used to the services of a transcriptionist centre will be motivated to use the RUTA:MED platform in a self-service manner, i.e., will prefer to do quick post-editing of the automatic transcription rather than wait for a transcriptionist centre to produce the report.

In the alternative scenario, we anticipate that a transcriptionist centre would still be preferred in the loop by many radiologists in case of more complicated examinations and therefore less fluent dictations. Nevertheless, transcriptionist centres will become more productive, since part of their workload will be overtaken by the self-service user scenario, as well as draft transcripts (generated

by RUTA:MED) will be available to the operators for post-editing instead of transcribing whole reports manually from scratch.

Figure 3 depicts the dual RUTA:MED workflow which involves an optional transition of a report via an associated transcriptionist centre.

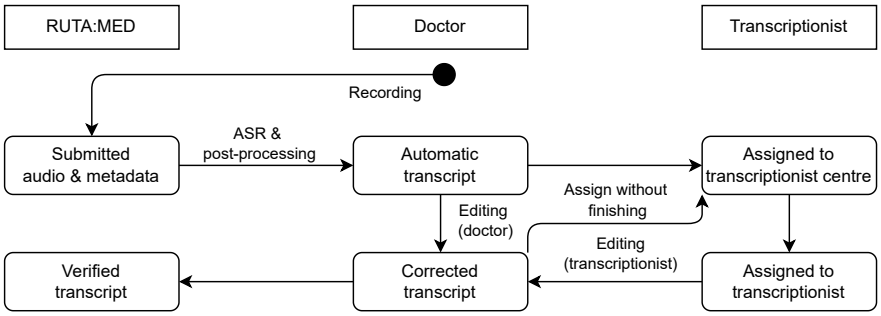

Fig. 3. State diagram of the RUTA:MED dual workflow

Although the RUTA:MED platform currently is not directly integrated with third-party medical information systems, the universal copy-paste integration which has already been exploited in the legacy workflows in Latvia and elsewhere can be used for the time being. Development of a widget and a browser plugin which could be used in combination with any other information system for convenient transfer of the final transcripts is in progress.

4 Initial Evaluation

Word error rate (WER) is the most commonly used metric for performance evaluation of ASR systems. Although WER is an excellent metric to compare the performance of different ASR systems, user experience is the most important criteria that shows how successful the platform actually is.

The user experience (UX) evaluation has to include two aspects – user interviews and feedback (qualitative analysis), and statistical key performance indicators (KPI). The RUTA:MED platform tracks three such KPIs:

- How many of the reports doctors decide to post-edit themselves instead of assigning them to a transcriptionist centre.
- How much time it takes to post-edit a report relatively to the duration of a dictation and the number of characters in the transcription.
- How many words and characters are changed in the final transcript compared to the automatic output of the ASR and post-processing pipeline.

The data collected is also an invaluable resource for further development and improvement. The post-edited texts add to the text corpus used for language modeling. It also helps to extend the lexicon by both reducing the out-of-vocabulary (OOV) rate and adding alternative pronunciations. The aggregated

user edits also provide excellent insight in what needs to be improved in the rewriting grammar as well (e.g. by adding new or alternative voice commands or by extending the set of text shortening rules).

The evaluation phase of the project has just begun and the data is being collected, but the initial user feedback is very positive.

5 Conclusion

We expect that the RUTA:MED platform will improve patient care by reducing the time one has to wait for the medical examination report. So far, doctors operating with the less resourced Latvian language submitted their recordings to a transcriptionist centre or typed them manually. With RUTA:MED, doctors can produce reports themselves spending considerably less time, while keeping the service of a transcriptionist centre as a fall-back option. This in turn reduces the workload of transcriptionist centres, allowing to finish other reports sooner.

Due to the high workload, doctors typically are against anything that requires additional efforts. One of the biggest advantage of RUTA:MED is that its workflow is very simple and close to the current practice. The recording and submission workflow is very similar among hospitals and clinics. The RUTA:MED workflow allows doctors to simply delegate a transcription to an operator if post-editing would be too time consuming at the moment. This is especially crucial for ASR systems with higher error rates.

Acknowledgements

This work was funded by ERDF (grant No. 1.1.1.1/18/A/153). Core development of the editor was done in the H2020 project SELMA (grant No. 957017).

References

1. Dargis, R., Gruzitis, N., Auzina, I., Stepanovs, K.: Creation of Language Resources for the Development of a Medical Speech Recognition System for Latvian. In: Human Language Technologies - The Baltic Perspective, Frontiers in Artificial Intelligence and Applications, vol. 328, pp. 135–141. IOS Press (2020)
2. Gruzitis, N., Dargis, R., Lasmanis, V.J., Garkaje, G., Gosko, D.: Adapting Automatic Speech Recognition to the Radiology Domain for a Less-Resourced Language: The Case of Latvian. In: Intelligent Sustainable Systems, Lecture Notes in Networks and Systems, vol. 333, pp. 267–276. Springer (2022)
3. Paats, A., Alumäe, T., Meister, E., Fridolin, I.: Retrospective Analysis of Clinical Performance of an Estonian Speech Recognition System for Radiology: Effects of Different Acoustic and Language Models. J Digit Imaging **31**(5), 615–621 (2018)
4. Tai, T., Skut, W., Sproat, R.: Thrax: An open source grammar compiler built on OpenFst. In: IEEE ASRU Workshop (2011)
5. Znotins, A., Polis, K., Dargis, R.: Media monitoring system for Latvian radio and TV broadcasts. In: Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH). pp. 732–733 (2015)