Dental TutorBot: Exploitation of Dental Textbooks for Automated Learning* **

Fizza Tauqeer[1] Muhammad Khubaib Raza[1] Myat Su Yin[2] Saeed-Ul Hassan[1]
Peter Haddawy[2] Mihai Pomarlan,[3] and Suppawong Tuarob[2]

[1] Information Technology University, 346-B, Ferozepur Road, Lahore, Pakistan
{msds19034,msds19064,saeed-ul-hassan}@itu.edu.pk
[2] Faculty of ICT, Mahidol University, Nakhon Pathom, Thailand
{myatsu.yin,peter.had}@mahidol.ac.th, suppawong.tua@mahidol.edu
[3] Faculty of Linguistics, University of Bremen, Germany
mihai.pomarlan@uni-bremen.de

Abstract. Active learning has been shown to provide benefits over traditional didactic approaches to teaching and learning. These benefits are particularly important in fields in which students must master large amounts of information and effectively operationalize it, as in medicine and dentistry. While online learning platforms have the potential to provide students with active learning without taxing scarce faculty resources, a recognized challenge in producing such systems is the engineering of the domain knowledge needed for engaging interaction. In this paper we address this problem by developing an open-source chatbot-based tutoring system trained on dental textbooks in the area of endodontics, one of the most challenging areas of dentistry. Dental TutorBot is built using Rasa for modular training purposes. It asks short questions from students and evaluates their answers. If the student cannot answer a question, the system provides a hint, rather than immediately giving the student the answer. In this way, it coaches the student to find the answer and thus helps them to understand the connections between concepts while creating a more intellectually stimulating learning experience.

Keywords: Smart Interactive Learning · QA Pairs · Hint Generation · GPT-2 · Rasa TutorBot

1 Introduction

Active learning has been shown to provide benefits over traditional didactic approaches to teaching and learning, including increased learner engagement, material retention, critical thinking, and problem-solving ability [4]. These benefits are particularly important in fields in which students must master large amounts

* This work was partially supported by a grant from the Mahidol University Office of International Relations to Haddawy in support of the Mahidol-Bremen Medical Informatics Research Unit.

of information and effectively operationalize it, as in medicine and dentistry. Indeed, studies recognize the importance of active learning in medicine and dentistry, but also point to its slow adoption [2]. A major obstacle for wider adoption is the higher demand for teacher-student interaction and the consequent need for lower student-faculty ratios. A promising approach to implementing active learning more widely is to use it as the model for computer-based learning. Such an approach can provide students with increased active learning time without taxing scarce faculty resources. Indeed, this is one of the motivations that has led dental schools to embrace the use of virtual reality simulation for the teaching of psychomotor skills and simple procedural skills [7]. However, the adoption of computer-based systems for the teaching of knowledge and cognitive problem-solving skills remains still a novelty. One reason is that producing qualitative interactive online content is highly labor-intensive, yet the knowledge for such systems already exists in the numerous high-quality textbooks that are available for dentistry. The bottleneck could be more easily addressed if such knowledge could be extracted and structured for use by an automated tutoring system.

A healthcare profession that could benefit positively from such an implementation is that of Dentistry. With the precaution yet necessity of social distancing when dealing with possibly asymptomatic patients, a dentist's job – which is reliant on applicable learning – can be significantly impacted. In comparison to other fields of education, medical areas still lag behind due to the reluctance of learning appropriately and efficiently by virtual education. It has hence become important for aspiring dentistry students to be able to learn the theoretical and practical aspects of their expected job by digital means just as well as traditional ones during such times.

To address this concern, this research aims to develop an open-source chatbot-based tutoring system trained on dental textbooks to teach and test dental students. It is trained on standard Natural Language Processing (NLP) datasets along with academic textbooks pertaining to endodontics, one of the most challenging areas of dentistry. The Dental TutorBot is built using Rasa for modular training purposes (rasa.com). It asks short questions from students and evaluates their answers. If the student cannot answer a question, the system provides a hint, rather than immediately giving the student the answer. In this way, it coaches the students to link their understanding to the answer, thus helping them gauge the connections between concepts better while establishing a more intellectually stimulating exchange.

2 Related Work

Prior work has explored development of intelligent tutoring systems for medical problem-based learning as one approach to active learning. The work of Suebnukarn et al. [6] used a hand-crafted Bayesian network knowledge representation to represent problem solutions and to generate hints. Subsequent work by Kazi et al. [5] explored the use of the UMLS medical ontology to generate hints as a means to avoid having to hand-craft the domain knowledge. While evaluation studies showed these two approaches to be effective, the range of hints they could generate was somewhat limited.

Chan et al. [3] utilized pre-trained BERT models to create realistic questions given any text passage. Based on their investigations, they decided to improve on how the BERT architecture handled the text generation task. The BERT baseline was implemented in order to compare it to their improvements since the naive model performed poorly on the standardized Stanford Question Answering Dataset (SQuAD) due to token generation of the text all at once, resulting in information loss over the context. The improvements made for this task were to apply two BERT models sequentially, in order to capture the encoded results from the first state and carry it over as input to the final state. This resulted in enhanced generative performance, as now no information was being lost due to the transformed nature of sequential tokenizing.

Bocklisch et al. [1], the originators of the Rasa methodology, explored why the framework was outperforming existing systems for tasks involving dialogue flow. The tools that were utilized – Rasa Natural Language Understanding (NLU) and Rasa Core – were based on standard tasks involving NLP using baseline methods of spaCy implementations, such as Part of Speech (POS) tagging with annotations. Uniquely, it brought to the chatbot field features such as adhering to a structure-free intent, storyboard and action sub-module that gave the user freedom to implement the type of dialogue flow they wanted, which certain state-of-the-art systems still have limitations on, such as the IBM Watson Tutor.

The IBM Watson Tutor [8] is a conversational tutoring system with similar objectives to our work. The tutor operates through five core functionalities: 1) A question is asked, 2) A student responds to the question, 3) The tutor provides feedback (correct, incorrect, or clue required), 4) interaction to improve thread performance and 5) the tutor continues the process until it is sure the student has understood the objective (by way of measuring the responses, interaction and the feedback provided). Their user experience studies showed that the degree of control over the conversation is not as robust as it may appear. The tutor cannot always provide the best objective dialogue when there is unanticipated behaviour from the user. Rasa, on the other hand, operates on learnable weights to train itself in identifying the user pattern over time, thus providing better objectives as the conversation continues from session to session.

3 Data and Methodology

In this section we describe data acquisition from dental text books, as well as the methods used to build the Dental TutorBot system.

3.1 Data Acquisition

This work uses the text corpus from the work of Yin et al. [9], where the authors acquired text from fifteen textbooks in the field of endodontic surgery. Since only the body text was of interest, they removed image captions, tables, box captions, footnotes, index, references, page numbers, words separated by a hyphen, headers/footers, citations and other irrelevant characters and objects from the text by applying multiple regular expressions manually. After cleansing, the corpus consisted of 1,012,922 words (tokens), 59,679 unique words (types). Apart from this, the SQuAD dataset is also utilized for validation purposes due to its standardized nature. It provides a baseline model to rely on for more commonly asked questions that overlap with the topics handled by the SQuAD dataset. The data generated from the sub-systems are also tied into the Rasa framework for initiating the dialogue flow process of this study.

3.2 Methodology

Our overall objective is to create a tutoring system in which the TutorBot asks the students questions based on the topic and difficulty level they select, checks whether the answers they provide are a) correct, b) semi-correct or c) incorrect, and provides solution hints to coach them to find the right answer. The specific objectives achieved in pursuit of this are described below.

Question Answer Generation: First, such a TutorBot must be different – in implementation and deployment – than usual chatbots since the user responds to the bot's query rather than vice versa. By this, the user in question – that is the student utilizing the system – would respond to the bot's questioning in a continuous manner. This specification has primary importance in the setting up of the research problem, as the approach varies substantially: we must implement the training and learning phases of each sub-solution with respect to the bot's responses, rather than the user's as the dialogue at their end would be limited to providing correct or incorrect responses to questions, topic selection, the usual salutation/goodbye, and the start/end period commands to signal the initiate/stop of the tutoring session. However, at the bot's end, we must check if relevant questions and their correct answers are being generated and selected given the user's topic selection. After this validation, the dialogue flow would initiate in a loop, until the generated QA pairs for the selected topic or difficulty have been exhausted (in which the tutor would ask to select a new topic or to adjust the difficulty level, or end the session altogether) or the user has specifically asked to stop the session. While within this loop, the underlying bot system keeps checking the similarity of the student input with the answers stored in a database housing pre-generated QA pairs from dental textbooks. A certain level of matching would see the response deemed correct, and if this match fails to meet the set threshold – say 75% – clues will be provided in order to aid recalling or at least attempting to guess the correct answer.

This process of QA generation is carried out by the T5 model through the algorithmic approach of answer aware generation, in which the model is presented with user-selected topic sets of all exhaustive passages extracted from the dental textbooks - for which it has been trained - where each potential sentence is considered an answer to a question. The system generates a question for these 'answers' by considering the passage context. Ingesting the textual data in the form of spans, questions are mapped to each answer outputted by way of the Encoder-Decoder framework and then compared with the generated answer pair of the same model validated on SQuAD dataset, which is used as baseline to check the capability of the questions being generated due to its academically accepted nature. The encoder captures the passage sentences as answers based on each sentence span, while the decoder then breaks down these very sentences into possible pairs of possible question and potential answer. Higher similarity to the ingested passage text amongst these pairs implies that a correct QA pair was generated, and is thus appended to a comma-separated values (csv) file with the relevant metadata read from the dental texts. This approach ensures that relevant questions are generated instead of non-addressing ones.

Hint Generation: The sample questions show a potential set of questions similar to what are asked in physical courses. To generate clues for the student interacting with the TutorBot in case of incorrect responses, we address the "hint" as a semantic chunk in an input passage that will be included (or rephrased) in the target question and the accompanied answer. Based on this definition, we perform syntactic parsing and chunking on input text, and identify those chunks which are the most relevant to the target question as the clue set. By leveraging two sub-systems for the task (with and without the Word Embeddings) - **Model A** for Lexical Simplification using WordNet, BabelNet and Rapid Automatic Keyword Extraction (RAKE) for the processes of Text Summarization, Expansion, Keyword Similarity and Rephrasing - along with **Model B**, which uses a GPT-2-small model trained on Semantic Clues learned from SQuAD using above pipeline and tested with the dental text. GPT-2 was used to exploit its generative ability to produce coherent text from minimal inputs. From our initial test-run, the word embedding module coupled with Model B performed the best in giving syntactically learned hints. A few examples of hints from these models are presented in Table 1. However, due to the lack of exhaustive hint-related data that could be sourced once the system is tested by its targeted audience as influenced through the literature review conducted, the hints still remain quite similar to the actual answer. With more data consumed from its intended employment, the models are expected to comparatively have a better chance to learn and provide more generic and answer-excluded hints.

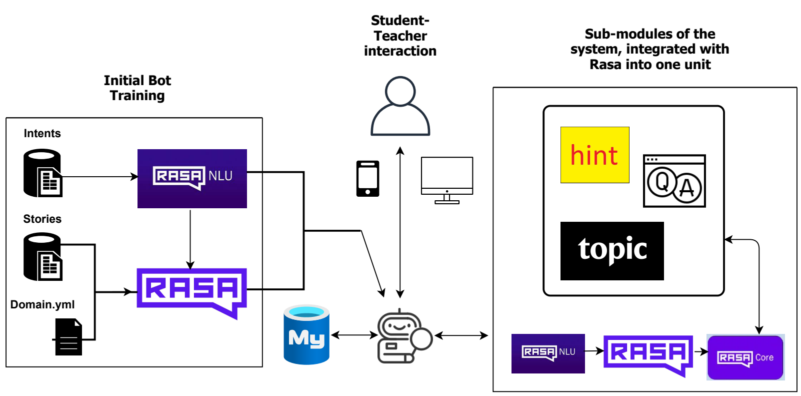

Fig. 1. Overview of TutorBot Framework showing the QA module (for question-answer pair generation), the Hint module (for hint generation) and the Topic module (for topic identification).

Rasa Framework: The modules described above are incorporated into the Rasa bot using the system architecture shown in Fig. 1. The integrated subsystems are able to effectively handle multiple modes of interaction between the student and the TutorBot, based on the user intent and accompanying action trigger. Student answers are matched against model-generated answers using cosine similarly. If the similarity between answer and student response is less than 75% a hint is generated to help the student answer the question. For bot interaction, we use Rasa which is easily integrated with different platforms and supports multiple languages that are available for such NLP exploitation. The students directly interact with the TutorBot in this manner as shown in Fig. 2, which illustrates how different parts of the interaction with the student are implemented on Rasa. After the initial interaction where the tutoring session is initiated, the inherent question-answer flow and hint generation start concurrently as well.

Table 1. Examples of questions and hints prior to word embedding (Model A) and after word embedding (Model B).

QA Pair	Hints	Model
Q: What is the ability of the pulp to do? A: "respond effectively to the insult"	It involves respond effectively to the result.	A
Q: What is the pulp encased in? A: "hard dentine shell"	It involves hard dentine eggshell.	A
Q: What does apical periodontitis lack? A: "sufficient collateral circulation"	It involves efficient collateral publication.	A
Q: What is the ability of the pulp to do? A: "respond effectively to the insult"	It involves robustness.	B
Q: What is the pulp encased in? A: "hard dentine shell"	It involves enamels.	B
Q: What does apical periodontitis lack? A: "sufficient collateral circulation"	It involves movement.	B

4 Results and Discussion

While conducting this study, the primary step was to deploy an essentially working mechanism and then enhance it by extending on the same order of sub-problems. We were able to successfully generate appropriate and extensive Question-Answer pairs, label topics for these pairs by using their associated occurrence in the dental textbooks, enable hints with regards to the user's response and conduct matching to determine correctness of answers. Most importantly, we were able to bring these sub-modules into one customized API and link the Rasa chatbot to it as well for endowing the dialogue flow with instructional content. From this preliminary study, we intend to investigate several ways to further augment the user–bot experience with the roles of a student–teacher relationship.

As seen in Table 1, the resultants of the QA and Hint Module show considerable improvement in the hint generation alone once the dataset's extracted word embeddings are utilized in the train-test phase of the module. With the local information well-captured by the embedding (Model B), the hints are likely to become more relevant and clue-like to the student rather than what a simple rewording of the keywords would ever appear to be (Model A).

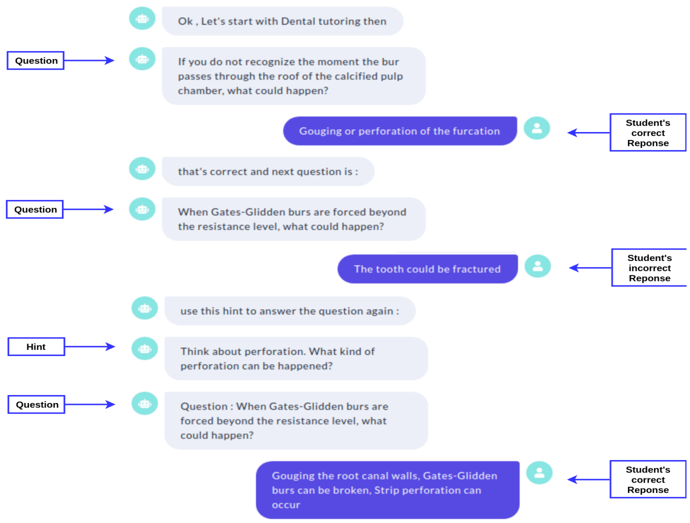

Fig. 2. An example student conversation with Dental TutorBot

5 Limitations and Future Work

Our future work includes analyzing how certain parameters – when fine-tuned – could affect the response of the system. For example, the threshold for similarity matching could be learned by the TutorBot by carrying out trend analysis for each individual question in a tutoring session. Similarly, we aim to include multiple and more profound hints in order to ease the student's experience and provide maximal learning capability. Once the system has been well-used by the stakeholders of the research problem – the students and the dental experts – users can provide rating tags for questions after each session in order for the bot to adapt and provide questions based on the gaps in the student's knowledge, apart from the initial topic selection. This will provide an evaluation metric to gauge the difficulty level of associated topics and their related questions, while capturing the number of attempts the student takes in answering them accordingly. Field testing and evaluation by dentists along with medical students is a necessity even if rating generation was not an immediate concern, as they can provide better elements of critique. The problem upon conclusion aims to create testing and feedback procedures of the implementation in collaboration with the respective dental experts, health-care professionals and students to maximize the influence and outreach of this research.

Up to this point we have evaluated the system as part of the iterative development process, receiving feedback from endodontists at each iteration. We are currently working on hosting the custom API as a mobile application in order to evaluate the system's effectiveness in teaching students. Once deployed, we will be able to gauge the student and bot responses and to fine-tune the system. In addition, we plan to expand to other question types such as multiple-choice questions and descriptive questions.

References

1. Bocklisch, T., Faulkner, J., Pawlowski, N., Nichol, A.: Rasa: Open source language understanding and dialogue management. NIPS 2017 Conversational AI workshop pp. 1–9 (2017)
2. Bucklin, B.A., Asdigian, N.L., Hawkins, J.L., Klein, U.: Making it stick: use of active learning strategies in continuing medical education. BMC medical education **21**(1), 1–9 (2021)
3. Chang, Y.H., Fan, Y.C.: A recurrent bert-based model for question generation. Proceedings of the 2nd Workshop on Machine Reading for Question Answering pp. 154–162 (2019)
4. Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.P.: Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences **111**(23), 8410–8415 (2014)
5. Kazi, H., Haddawy, P., Suebnukarn, S.: Employing umls for generating hints in a tutoring system for medical problem-based learning. Journal of biomedical informatics **45**(3), 557–565 (2012)
6. Suebnukarn, S., Haddawy, P.: A bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system. Artificial intelligence in Medicine **38**(1), 5–24 (2006)
7. Towers, A., Field, J., Stokes, C., Maddock, S., Martin, N.: A scoping review of the use and application of virtual reality in pre-clinical dental education. British dental journal **226**(5), 358–366 (2019)
8. Ventura, M., Chang, M., Foltz, P., Mukhi, N., Yarbro, J., Salverda1, A.P., Behrens, J., Ahn, J.w., Ma, T., Dhamecha, T.I., Marvaniya, S., Watson, P., D'helon, C., Tejwani, R., Afzal, S.: Preliminary evaluations of a dialogue-based digital tutor. 2018 International Conference on Artificial Intelligence in Education pp. 480–483 (2018)
9. Yin, M.S., Pomarlan, M., Haddawy, P., Tabassam, M.R., Chaimanakarn, C., Srimaneekarn, N., Hassan, S.U.: Automated extraction of causal relations from text for teaching surgical concepts. In: 2020 IEEE International Conference on Healthcare Informatics (ICHI). pp. 1–3. IEEE (2020)