33rd Annual Conference of the Australasian Association for Engineering Education (AAEE 2022)
4 - 7 December 2022 – Sydney, NSW

Finding Answers to Student Questions in Engineering Textbooks through Machine Learning

Jake Van Heekeren[a]; Samuel Cunningham[a,b], and Wageeh Boles[a].
School of Electrical Engineering and Robotics, Queensland University of Technology[a], Learning and Teaching Unit, Queensland University of Technology[b]
Corresponding Author Email: sam.cunningham@qut.edu.au

ABSTRACT

CONTEXT

Many tertiary education courses prescribe or recommend textbooks to provide context and more elaborate content to enhance student learning. However, these textbooks contain a large amount of information for students to digest. For students who could be new to the topic, finding a single concept or answering a question they have might prove difficult or daunting.

PURPOSE

Machine learning techniques have been used for text analysis and can extract useful insights from textual data, and therefore have the potential to assist students in utilising their textbooks. These techniques can be used to extract key information from substantial amounts of text. This leads to the question; can machine learning models extract useful answers to engineering content questions, from a range of textbooks?

METHODOLOGY

This research is focused on using techniques based on *Dense Passage Retrieval for Open-Domain Question Answering* (Karpukhin, et al., 2020). These models have been used to extract information from chosen textbooks and thus, given a question, the models will produce a text answer to the question. The chosen textbooks broadly cover three key common first year engineering subjects, mechanical engineering, electrical engineering, and interpersonal and management skills essential for engineers.

OUTCOMES

Our approach has been tested by posing questions and having the model extract possible answers from the selected textbooks. For example, from the Hambley Electrical Engineering textbook, we used open ended questions such as “What is a Wheatstone bridge?” and “How are series resistances calculated?” The model currently provides answers such as “A circuit used to measure unknown resistances” and “By their sum”, respectively.

CONCLUSIONS

Students could greatly benefit from these models as an additional resource, in the form of an interactive textbook or other implementation, to support their learning. This may help students find answers to conceptual questions without waiting for answers from educators regarding a specific content related question, especially for students who may not have many tutors available. The results show the ability of these models to recognise intent in each question and provide answers for students. These models do have some difficulties in questions that are more complex, with fine-tuning, further optimisation could potentially yield better results in the future.

KEYWORDS

Textbooks, Text Analysis, Natural Language Processing, Chatbot

Introduction

The internet provides students with access to a vast amount of information, which can be invaluable for their learning journey. Although this information is valuable for students, it can be difficult to navigate. Students are required to first find the relevant information online and then make the determination if the source is trustworthy or accurate. This is a difficult task for a student who is learning new material. For example, previous research shows that Wikipedia articles on various topics within the medical field contained information of very different qualities (Azer et al., 2015). One benefit of these online sources however is that they are easy to search and find.

In their studies, students are often prescribed recommended textbooks related to the subject of study. These textbooks are used for a variety of purposes but have been written by an expert in the field and selected by an academic for use in the subject who is also knowledgeable in the field. When compared with the online sources, the textbooks are often seen as an authority. It would be beneficial for students for students to be able to benefit from the authority of the textbook but benefit from the searchable nature of online sources. This led to the research question: Can machine learning models extract information, given a question, from a range of engineering textbooks?

Background and Literature Review

Student Use of Textbooks

Textbooks form a key resource to support student learning, and often act in a complementary way to key learning resources directly provided by an educator. They form a valuable part of the learning experience, studies showing that reading excerpts from textbooks before solving problems can reduce the time taken to solve them (Atman and Bursic, 1996).

Many textbooks are now accessible in both hard copy, and digital, with some exclusively online. Research shows that students are utilising electronic textbooks more often, and that there are no significant differences between learning in groups of students that utilise paper or electronic textbooks (Weisberg, 2011). One potential benefit of online textbooks is the ability to create interactive activities to support active learning, overcoming a limitation of hard copy textbooks. These interactive textbooks can improve student grades and experience (Edgcomb et al., 2015). Technology and automated processes such as machine learning provide an opportunity to create these interactive experiences.

Machine Learning and Natural Language Processing

Machine Learning is currently being used in many applications, including information extraction from text. Machine Learning involves applying techniques of pattern recognition and classification. Specific algorithms and processes attempt to model human learning (Langley, 1996). Machine learning also provides the opportunity for certain tasks to be performed a scale by a computer, that would not be possible for humans to complete.

Natural language processing, or text analysis allows documents or pieces of text to be grouped and classified to answer key questions. Examples of applications of natural language processing include thematic analysis (Odden et al., 2020), sentiment analysis (Tang, Qin and Liu, 2015) and summarisation of large pieces of text (Denil et al., 2014). These applications all allow for large amounts of text to be analysed.

One key element of text analysis involves identifying the features of text. Bag of Words features have been traditionally used, mapping individual words in a document to a ‘1’ if they occur, or a 0 if they don’t occur (Zhang, Jin and Zhou, 2010). These 1s and 0s are mapped to a large dictionary, containing a list of each word contained within the document. Term Frequency - Inverse Document Frequency (TF-IDF) improves on Bag of Words by placing weightings on each word, dependant on how frequently it appears, following a probabilistic model of information retrieval (Robertson, 2004). Word embeddings provide an opportunity to improve on this even further by modelling words in a vector space (Mikolov et al., 2013). This allows for the word features to contain levels of semantic information, increasing the information available for machine learning models.

Bidirectional Encoder Representations from Transformers (BERT) machine learning models are commonly used for these natural language processing tasks (Devlin, Chang, Lee, & Toutanova, 2018; Liu, et al., 2019). These BERT models use word embeddings with semantic representations to make predictions. Specific to this research, Dense Passage Retrieval (DPR) processes utilise word embeddings, and a BERT model to identify key parts of documents (Karphukin, et al., 2020). DPR is used in this research to support automated question and answering from textbooks. These models have been applied in this education space, in extracting information from textbooks.

Methodology

Question and Answer Algorithm

BERT-based machine learning models can perform extractive question and answers (Q&A) on short passages effectively by being able to return the location of a likely answer to a question within that context. RoBERTa acts as a *reader* for the Q&A task. RoBERTa is a pretrained and optimised variant of a Q&A BERT model. It takes a context and question and is able to comprehend a given passage and question, and then outputs an answer. However, the size of the context that can be *read* is limited. To process substantial amounts of text another component is needed. This is where the DPR model functions as a *retriever* to reduce the context size for the *reader* to process (Karpukhin, et al., 2020). The Dense Passage Retriever model uses 2 BERT models, one encodes the question, and the other model encodes the provided source passages into passage representations and are indexed and stored for later *retrieval* using a similarity-based document store (Johnson, 2019).

The process of the extractive Q&A model can be seen below in Figure 1. The question and passage encodings are compared for similarity. This returns the most relevant passages to the question based on semantic similarity, rather than just lexical similarity, which is where DPR can outperform other methods such as TF-IDF and BM25 (abbreviation of Best Matching). BM25 is an improved version of TF-IDF, but suffers the same drawback mentioned; TF-IDF and BM25 only match lexical representations. The DPR model, using dense vectors and similarity functions, can compare the meaning behind words and as such is able to *retrieve* passages that may not use the exact same words, but has content relevant to the question nonetheless (Karphukin, et al., 2020). This is due to the words being encoded as dense vectors which represent a projection in a multidimensional continuous vector space (Johnson, 2019). Words with similar meanings may have a similar direction or be closer to each other in this space than words that are inherently different, this enables many more similarity operations than just the comparison of whether one word in a question in exactly equal, letter for letter, to another word in a passage.

The relevant passages that are extracted are input into the pre-trained RoBERTa model which is available online (Pietsch, et al. 2022). The pre-trained model allows a user to take advantage of an initial semantic understanding. This model then extracts the location of the likely answer within the text, which can then be decoded to *retrieve* the answer to the question. A RoBERTa model forms the *reader* component of the pipeline. The RoBERTa and DPR-Base model are stored and accessed locally.

An existing process follows the method mentioned above and shown in Figure 1 by using DPR-base as the *retriever*, and RoBERTa as the *reader,* linking them together (Pietsch, et al. 2022). A set amount of the top relevant passages to the question are *retrieved* using the DPR model, and then a top set number of relevant answers are retrieved, with their relevant passages also displayed in the command interface. For the purposes of this explorative work, the top ten passages are *retrieved*, with the top three answers *read* by RoBERTa returned.

Examples of the top three answers for each question are presented in the results and discussion section. These answers are qualitatively analysed, and the resulting performance is determined based on the models' ability to provide relevant information to the user. The time taken by the developed Q&A model to respond, is also considered and presented in the results and discussion section.

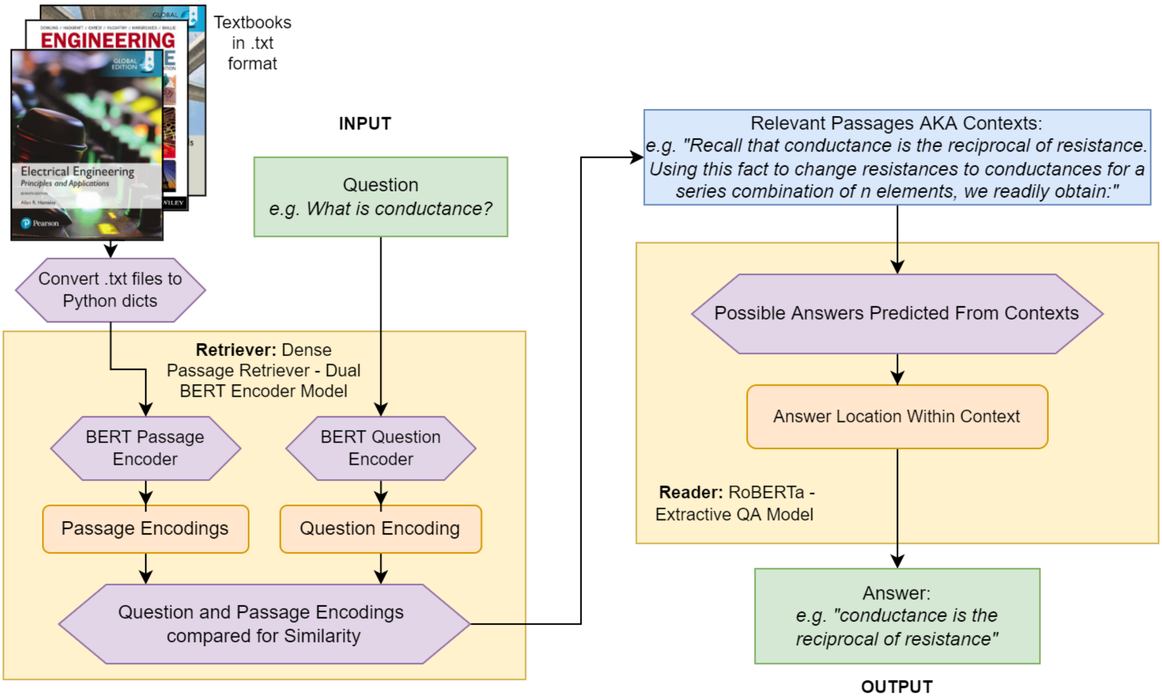

Figure 1 – Input to Output Pipeline for Extractive Q&A process

Selected Textbooks

For this research, three textbooks that cover a wide range of first year engineering topics were chosen. These were: Engineering your Future: An Australasian Guide by David Dowling, Roger Hadgraft , Anna Carew, Tim McCarthy, Doug Hargreaves, and Caroline Baillie (2015); Electrical Engineering Principles and Applications by Allan Hambley (Hambley, 2019), and; Statics and Mechanics of Materials by Russell Hibbeler (Hibbeler, 2018).

These books provide a broad range of knowledge across mechanical and electrical engineering along with important professional skills required for engineers. To comply with fair use of these texts, only portions of each textbook are used by the model. From the electrical engineering text by Hambley, chapter 2 on resistive circuits is selected. From the mechanical engineering text by Hibbeler, chapter 8 on material properties is selected. Finally, from the text on engineering practices by Dowling et al., part 3 is selected, which primarily discusses collaboration and self-management.

To prepare these chapters, the texts are converted to text files and stored in a single folder. It was determined that the order of the text file had a small effect on the *retrieved* answers, occasionally modifying the 4th of 5th top answer. However, this did not appear to significantly affect the model's ability to return a relevant answer, and often the 4th and 5th returned answers were not relevant, that is why the top 3 responses are shown in the results section.

During the conversion to text files, some irrelevant information for the Q&A model may remain, so the text files should be cleaned before processing. This irrelevant information needs to be removed from the text file before processing. Since the model returns text answers based on engineering concepts, numerical problems and solutions are also removed from these text files. The problems and solutions are not removed from the text file as these can provide valuable conceptual answers to the user. Other anomalies, resulting from the conversion to text format, are removed such as character representations of figures and table lines.

Algorithm Performance and Hardware

The hardware used to develop and train the models is GPU (RTX–2060), CPU (Intel i7-1070) and RAM (32 GB). As can be seen in the results section below, the slowest time was approximately 32 seconds, with the fastest response being obtained in approximately 13.5 seconds. Although this is not ideal when considering large volumes of questions and answers from large volumes students in an academic environment, it is noted that these results have been obtained with consumer grade hardware. Through further optimisation, and additional hardware this process could run significantly faster. Adding additional text did not appear to dramatically increase the response time, (response times increased by an additional 1 or 2 seconds only). Another factor to consider is that it may be possible for the passages to be encoded prior to runtime with the question (Karpukhin, et al. 2020). This would also save a significant amount of run time, thus, further enabling the models to process Q&A tasks faster.

Results and Discussion

Four questions are selected for each textbook to assess the performance of the model. This allows for questions with similar content to be asked and analysed, whilst also allowing for other questions to be asked in the same field. Similar questions were often asked to ensure that the model could differentiate between the two. For example, in electrical engineering, asking how series resistances are calculated, and then asking how parallel resistances are calculated.

Electrical Engineering: Principles & Applications (Chapter 2, Resistive Circuits)

The questions asked in Table 1, are electrical engineering based, specifically regarding resistive circuits, as the textbook material provided to the model is chapter 2, “Resistive Circuits”, from Electrical Engineering Principles and Applications. (Hambley, 2018). Since RoBERTa is extractive, any answers *read* from the textbook will be direct quotes, as such it is largely their relevance to their question that will be evaluated. Cells in the table are highlighted with blue if they are considered to be the most appropriate answer.

Question 1 presents a series resistance question about how the total resistance is calculated and the model responded with the correct answer in response 2: ‘their sum’ (p72). Although this is concise and correct, it would be preferable to have some more detail. An alternative answer can be seen on page 68 “a series combination of resistances has an equivalent resistance equal to the sum of the original resistances.” The model may have chosen to take the answer from a solution may be due to the use of the word “calculated” in the question. Notably the third response is taken from the mechanical engineering textbook, perhaps because resistance can be used in both contexts for different meanings.

Question 2 asks about parallel resistances had a detailed and correct answer in response 2, (from p77) which answers the question quite well. Interestingly, as opposed to the series resistance question in question 1, it has stated the current-division principle rather than referencing steps taken in a solution. This may be that its harder to answer the question with a simple “by their sum” as was done in question 1.

Question 3, regarding the voltage-division principle, is met with the correct answer in response 1 correctly identifying how voltage-division is calculated (from p76), however the use of the word ‘Circuit’ does make it more evident that the model has extracted a partial sentence and it would have been preferable for it to extract more of the sentence “Of the total voltage, the fraction that appears across a given resistance in a series circuit is the ratio of the given resistance to the total series resistance. This is known as the voltage-division principle ". The model may have cut the answer short due to the way the passages have been organised. Responses 2 and 3 are also reasonable given the question is asking “what” which could be open to interpretation, providing information on when it applies in both responses 2 and 3.

#	Question asked	Response 1	Response 2	Response 3	Time (s)
1	How are series resistances calculated?	A fraction of the voltage appears across each of the resistances	By their sum	Normal stress on the vertical axis and normal strain on the horizontal axis	13.51
2	How are parallel resistances calculated?	By their equivalent resistances	The fraction of the total current showing in a resistance is the ratio of the other resistance to the sum of the two resistances	A fraction of the total current shows through each resistance	13.97
3	What is the voltage-division principle?	Circuit is the ratio of the given resistance to the total series resistance	Applies only for resistances in series	When a voltage is applied to several resistances in series	14.49
4	What is a Wheatstone bridge?	A circuit used to measure unknown resistances	A circuit used to measure unknown resistances	balanced	19.61

Table 1 – Electrical Engineering-based Questions and Responses

Finally, **Question 4** asks the model, “What is a Wheatstone bridge?”. It was met with the correct and concise response of “A circuit used to measure unknown resistances” (p127). This was an easy question for the model to answer, compared to the previous questions about calculating resistances, as it is a definition, which is written in the text clearly by the author. This is also likely why question 3 was a reasonably easy question to answer, with the top responses for questions 3 and 4 being correct.

Statics and Mechanics of Materials (Chapter 8, Mechanical Properties of Materials)

The questions asked in Table 2 are in the mechanical engineering context, specifically related to material properties as the textbook section used is Chapter 8, Mechanical Properties of Materials from Statics and Mechanics of Materials. (Hibbeler, 2019).

In **Question 1**, the model is asked about material toughness. This resulted in the correct answer given in response 1, stating that toughness is from ‘the area under the a-e diagram’ (p408). It should be noted that omega and epsilon have been converted in the ‘a’ and ‘E’ in the process of converting to a text file for the model. As this response is referring to the stress-strain diagram, this is the correct answer. Being a shorter response, the answer could be more detailed, however given that this context provided the practical mathematical way to determine the toughness it was one of the best responses from Chapter 8. The other responses stated ‘experiment’ which while true, is not necessarily going to assist the user, nor does it provide details of the experiment.

The model found it more difficult to identify material resilience in **Question 2**. Response 3 was the closest, stating ‘Linear elastic behaviour’ (p414), as resilience can be measured as the area under the linear elastic section of the stress-strain curve. Interestingly, the model chose this answer of the context of a question, and did not choose a definition, however this may be due to Chapter 8 of the book defining the *modulus of resilience* rather than defining *resilience* by itself as a term. In fact, when asked the same question using *the modulus of resilience*, the model returns a more appropriate answer in response 2: “strain energy density” which is correct.

#	Question asked	Response 1	Response 2	Response 3	Time (s)
1	How do I find the toughness of a material	From the area under the a-E diagram	Experiment	By Experiment	14.00
2	What is material resilience	Toughness	Perfectly plastic	Linear elastic behaviour	14.49
3	What is the definition of material strength?	Ultimate Stress	Ability to sustain a load without undue deformation of failure	About 12.5 times greater than its tensile strength	14.14
4	What does the stress-strain diagram show?	How its resilience and toughness can be changed	How stress can be related to strain	Breakdown of the material and cause it to deform permanently	13.88

Table 2 – Mechanical Engineering-based Questions and Responses

When asked about the definition of material strength in **Question 3**, the model was able to return a reasonably accurate answer in response 2, ‘ability to sustain a load without undue deformation of failure’ (p397). Note the use of ‘of’ is a text processing error when changing the content to a format suitable for the model. This was a promising response as the strength of materials is a broad field involving many material characteristics discussed in Chapter 8 of the textbook (Hibbeler, 2019). As such extracting this answer from the introduction was appropriate.

When asked what the stress-strain diagram shows in **Question 4**, the model answers the question best with response 2: “How stress can be related to strain” (p397). This is what the stress-strain diagram shows, however responses 1 and 3 also mention other properties that the diagram can present. Response 2 is taken from the front page of Chapter 8, however 8.2 discusses the stress-strain diagram in detail.

Engineering Your Future (Part 3, Self-Management and Teamwork)

Table 3 focuses on questions and answers regarding teamwork and self-management practices, based on content from Part 3, Self-Management and Teamwork, from Engineering Your Future (Dowling, et al., 2019). These answers can be more difficult to summarise in one answer, as professional skills such as teamwork and self-management often depend on the situation, environment and current state. As such these responses will be discussed based on their appropriateness as an answer to a question.

In **Question 1**, when asked what the best way to work in a team is, the model responds with reasonable responses. Response 2 is perhaps too specific and does not really answer the question. Responses 1 (p300) and 3 (p287) show different approaches to leadership and group structure and are listed as different approaches to working in team in the textbook. The model lends a stronger focus towards leadership regarding working well in a team, this may be due to focus shown in the textbook towards this.

In **Question 2**, the machine model was asked about how to collaborate with uncooperative team members and returned a detailed response in number 3 stating ‘taking a positive approach and encouraging other team members to adopt a similar attitude’ (p306). Responses 1 and 2 were vague and less detailed.

#	Question asked	Response 1	Response 2	Response 3	Time (s)
1	What is the best way to work in a team?	Shared leadership	Collaborative writing assignment	Single-Leader Discipline	31.93
2	How do I collaborate with uncooperative team members?	Motivation	Teamwork and cooperation Teamwork	Taking a positive approach and encouraging other team members to adopt a similar attitude	17.06
3	How do you manage your own time?	efficiently	Isolated areas	By using a timesheet	16.47
4	How do I manage my own time?	Prioritising competing demands	Prioritising competing demands to achieve	demands	17.51

Table 3 – Teamwork and Self-Management Focused Questions and Responses (Blue Shading Identifying Answers that Appropriately Answer the Question)

Questions 3 and 4 then ask the model how to effectively manage time. The responses changed depending on if the subject of the question was 'you' or 'I'. In Question 3 the most reasonable response was 3; 'By using a timesheet' (p275). This isn't wrong but is perhaps too specific. In Question 4 when asking 'How do *I* manage *my* own time?' the model returned a conclusive answer, which it repeated in the top responses. This being 'prioritising competing demands' the top 2 coming from page 222. Considering how broad the question is, this could be good advice.

Concluding Remarks

The results of this research show the potential as well as the difficulty machine learning models can have, when used to assist students' learning from textbooks. The relevance of the answers shows promise for these models to analyse broad amounts of information without specific finetuning. This has potential to not only be useful for students, but to any task that requires the analysation of large amounts of text.

Regarding limitations, the model did appear to struggle with some questions only showing the correct answer on second or third result. It also in one case extracted an answer from a different engineering field, however these effects could be reduced by finetuning the models to specific textbooks. Work could be done in the future to decide how much finetuning improves the relevancy of returned answers, and how much improvement is achieved. Through further experiments, and provision of large sets of established questions and answers, the performance of the model can be finetuned to maximise the amount of useful answers that are given.

For students and the education sector, this presents a potential avenue for interactive textbooks, providing a more active learning experience. The model presented provides an opportunity to utilise machine learning to support the development of interactive experiences. One potential implementation could be the presentation of these answers as well as linking to the areas in the textbook where these concepts are discussed enabling further reading. The ability of this model to provide several answers to one question will provide access to the relevant context, and thus enriching the learning experience.

References

Atman, C. J., & Bursic, K. M. (1996). Teaching engineering design: Can reading a textbook make a difference?. Research in Engineering Design, 8(4), 240-250.

Azer, S. A., AlSwaidan, N. M., Alshwairikh, L. A., & AlShammari, J. M. (2015). Accuracy and readability of cardiovascular entries on Wikipedia: are they reliable learning resources for medical students?. BMJ open, 5(10), e008187.

Denil, M., Demiraj, A., Kalchbrenner, N., Blunsom, P., & de Freitas, N. (2014). Modelling, visualising and summarising documents with a single convolutional neural network. arXiv preprint arXiv:1406.3830.

Devlin, J., Chang, M., Lee, K. And Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2 [cs.CL] https://doi.org/10.48550/arXiv.1810.04805

Dowling, D, Hadgraft, R, Carew, A, McCarthy, T, Hargreaves, D, & Baillie, C (2015). Engineering your future: An Australasian guide (3rd ed.). John Wiley & Sons, Melbourne.

Edgcomb, A. D., Vahid, F., Lysecky, R., Knoesen, A., Amirtharajah, R., & Dorf, M. L. (2015, June). Student performance improvement using interactive textbooks: A three-university cross-semester analysis. In 2015 ASEE Annual Conference & Exposition (pp. 26-1423).

Hambley, A. R. (2019). Electrical Engineering, Principles and Applications. (7th ed). Pearson Education, Limited.

Hibbeler, R. (2018) Statics and Mechanics of Materials. (SI 5th ed.). Pearson Education, Limited, Harlow.

Johnson, J., Douze, M. and Jégou, H., (2019) Billion-Scale Similarity Search with GPUs. IEEE Transactions on Big Data, vol. 7, no. 3, pp. 535-547, https://doi.org/10.1109/TBDATA.2019.2921572

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Ledell, W., Edunov, S., Chen, D and Yih, W. (2020). Dense Passage Retrieval for Open-Domain Question Answering. arXiv:2004.04906v3 [cs.CL] https://doi.org/10.48550/arXiv.2004.04906

Langley, P. (1996). Elements of machine learning. Morgan Kaufmann.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692v1 [cs.CL] https://doi.org/10.48550/arXiv.1907.11692

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Odden, T. O. B., Marin, A., & Caballero, M. D. (2020). Thematic analysis of 18 years of physics education research conference proceedings using natural language processing. Physical Review Physics Education Research, 16(1), 010142.

Pietsch, M., Soni, T., Chan, B., Möller, T., & Kostić, B. (2022). Haystack (Version 1.1.0). GitHub. https://github.com/deepset-ai/haystack/

Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. Journal of documentation.

Tang, D., Qin, B., & Liu, T. (2015, September). Document modeling with gated recurrent neural network for sentiment classification. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432).

Weisberg, M. (2011). Student attitudes and behaviors towards digital textbooks. Publishing research quarterly, 27(2), 188-196.

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. International journal of machine learning and cybernetics, 1(1), 43-52.

Acknowledgements

Conference travel support for Sam Cunningham was provided by Queensland University of Technology’s Centre for Data Science.

Copyright statement