|  |  |  |  |
| --- | --- | --- | --- |
| Title | HQA-Data: A historical question answer generation dataset from previous multi perspective conversation | | |
| Author(s) Name | Sabbir Hosen, Jannatul Ferdous Eva, Ayman Hasib, Aloke Kumar Saha, MF Mridha, Anwar Hussen Wadud | | |
| Contact Email(s) | firoz.mridha@aiub.edu | | |
| Published Journal Name | Data in Brief | | |
| Type of Publication | Journal | | |
| Volume | 48 | Issue |  |
| Publisher | Elsevier | | |
| Publication Date | 2023/6/1 | | |
| ISSN |  | | |
| DOI | <https://doi.org/10.1016/j.dib.2023.109245> | | |
| URL | <https://www.sciencedirect.com/science/article/pii/S2352340923003645> | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| This data article contains a quality assurance dataset for training the chatbot and chat analysis model. This dataset focuses on NLP tasks, as a model that serves and delivers a satisfactory response to a user's query. We obtained data from a well- known dataset known as “The Ubuntu Dialogue Corpus” for the purpose of constructing our dataset. Which consists of about one million multi-turn conversations containing around seven million utterances and one hundred million words. We derived a context for each dialogueID from these lengthy Ubuntu Dialogue Corpus conversations. We have generated a number of questions and answers based on these contexts. All of these questions and answers are contained within the context. This dataset includes 9364 contexts, 36,438 question-answer pairs. In addition to academic research, the dataset may be used for activities such as constructing this QA for another language, deep learning, language interpretation, reading comprehension, and open-domain question answering. We present the data in raw format; it has been open sourced and publicly available at https://data.mendeley.com/datasets/p85z3v45xk. | |