Proceedings of the 9th International Scientific and Practical Conference «International Scientific Discussion: Problems, Tasks and Prospects»

(December 19-20, 2024). Brighton, United Kingdom

No 229

INFORMATION AND WEB TECHNOLOGIES

DOI 10.51582/interconf.19-20.12.2024.054

Development and Evaluation of an intelligent Q&A chatbot system for the medical aesthetics domain

Gamayun Igor[1], Ivashchenko Oksana[2], Gang Zhou[3]

[1] Doctor of Technical Science, Professor;
National Technical University «Kharkiv Polytechnic Institute»; Ukraine

[2] PhD, Senior Lecturer;
National Technical University «Kharkiv Polytechnic Institute»; Ukraine

[3] Master Degree Student;
National Technical University «Kharkiv Polytechnic Institute»; Ukraine

Abstract.
In recent years, the medical aesthetics industry has experienced rapid growth, becoming a focal point for individuals concerned with their appearance and self-image. This surge in interest has led to an increased demand for information and services related to medical aesthetics. The traditional consultation methods, including phone consultations, online chats, and face-to-face meetings, present such challenges as limited availability of experienced professionals, extended waiting times, and failing to meet users' expectations for prompt responses. This article presents development an intelligent Q&A chatbot system tailored to the medical aesthetics domain which should meet users' needs for precise and real-time advice. By employing NLP and deep learning technologies, we seek to create a model to understand user questions and deliver precise answers. The developed system consists of several modules such as User Interface, Intelligent Q&A System, Knowledge Base, Data Processing, and Pre-trained Model for Medical aesthetics.

Keywords:
medical aesthetics
intelligent chatbot
API
natural language processing
BERT

INFORMATION AND WEB TECHNOLOGIES

Introduction. The medical aesthetics industry has witnessed significant growth, this growth is driven by a social shift towards greater emphasis on aesthetics and self-image. This surge in interest has led to a significant increase in the demand for reliable and comprehensive information about medical aesthetics. Individuals are eager to learn about various procedures, post-care methods, products, and the associated risks and outcomes. The traditional methods of obtaining information, such as phone consultations, online chats, and face-to-face meetings, are increasingly inadequate in meeting this growing demand. Current advisory methods in the medical aesthetics domain face several limitations that hinder their effectiveness. The specialized nature of medical aesthetics requires the involvement of experienced professionals who may not always be available to address every inquiry. Traditional methods often provide generic advice that may not fully address the unique concerns of each user. This lack of personalization can lead to dissatisfaction and may even result in users making decisions about their treatments. The accessibility of medical aesthetics consultation service is another significant issue. Traditional methods are often limited by time and location constraints, this makes it hard for users to get the advice when they need it. For instance, scheduling a face-to-face consultation may require users to take time off work or travel long distances, adding to the inconvenience.

The traditional methods often rely on the expertise and availability of individual professionals, that's leading to variability in the quality of advice provided. An intelligent Q&A chatbot system can integrate professional knowledge and experience from the medical aesthetics domain to provide accurate and reliable consultation service consistently.

The Natural Language Processing (NLP) is a crucial component in developing an intelligent Q&A chatbot for the medical aesthetics domain. By implementing text preprocessing, feature extraction [1], [2], [3], named entity recognition [4], sentiment analysis, and contextual understanding [5], [6], the chatbot can accurately interpret user queries and provide relevant and personalized responses.

INFORMATION AND WEB TECHNOLOGIES

The BERT (Bidirectional Encoder Representations from Transformers) model is a state-of-the-art NLP model that has revolutionized the field of natural language understanding. BERT has the ability to capture contextual information from both directions (left-to-right and right-to-left), this makes it particularly well-suited for understanding the nuances of user queries in the medical aesthetics domain [7], [8].

BERT is pre-trained on a large corpus of text data, it uses two unsupervised learning tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). This pre-training enables BERT to learn rich contextual representations of words and sentences. Once pre-trained, BERT can be fine-tuned for specific tasks, such as question classification in the medical aesthetics domain. Fine-tuning involves training the pre-trained BERT model on a labeled dataset of questions and their corresponding categories. Overall, the BERT model is a powerful tool for question classification in the medical aesthetics domain. By leveraging its contextual understanding capabilities, the fine-tuned BERT model can accurately classify user questions, enabling the intelligent Q&A chatbot to provide relevant and personalized responses.

Once the user question has been classified using the BERT model, the next step is to retrieve the appropriate answer from a pre-populated database. Populating the database with accurate and comprehensive answers is crucial for the effectiveness of the intelligent Q&A chatbot. The following points we need to take into consideration for the process of populating the database. The first one is curating answers that means the answers should be curated from reliable and authoritative sources, such as medical aesthetics professionals, research papers, and reputable websites. Each answer should be reviewed for accuracy and relevance. The second one is about categorizing answers, the curated answers should be categorized based on the predefined question types identified by the BERT model. This ensures that the correct answer is retrieved for each classified question. To enhance the accuracy of the retrieved answers, several techniques can be employed, such as contextual matching which indicates the retrieval process can be enhanced by incorporating contextual

INFORMATION AND WEB TECHNOLOGIES

matching techniques that consider the context of the user query. This involves using advanced NLP techniques to match the query context with the stored answers.

Main part. In order to meet the requirements of end user, the architecture design of the intelligent Q&A chatbot system for medical aesthetics is shown in figure 1. The main modules of this system are including:

1 User Interface (UI): the user interface allows users to easily interaction with intelligent chatbot system, such as ask questions, browse answers and related information, and view history. It is developed by using the Vue.js front-end framework.

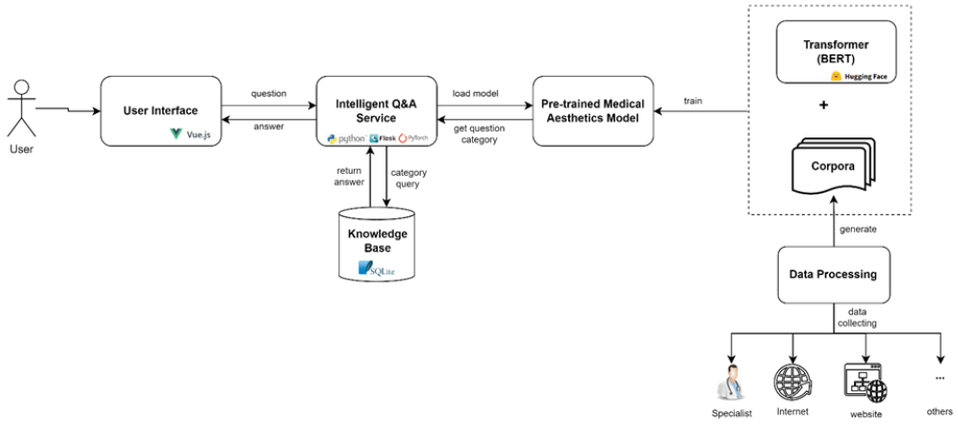

Figure 1
Architecture diagram of Intelligent Q&A Chatbot System

2 Intelligent Q&A System: This core component is mainly responsible for functions related to user services in the medical aesthetics domain, such as user registration, user login, session management, answering questions, and system security. It is developed by using Flask, Python, and Pytorch.

The Q&A service is core component of entire intelligent chatbot system, it's also known as backend service. It exposed APIs to client to ask questions and respond with correct answers. The main features of this component are including user registration, user sign in, user information management, session management, loading pretrained language model, and logout management.

INFORMATION AND WEB TECHNOLOGIES

The session management is also known as conversation management, it's including the creating, modifying and deleting the conversations. Once user logs in successfully, they can manage the conversations, such as creating new consultation or deleting the conversations. The entire process involves the client (UI) requests the APIs of backend service with access token and obtained conversation numbers, content and other data in the API response, thereby managing conversations and performing the corresponding operations. The same user can customize many different conversations (topics) for Q&A.

The question-and-answer function in this service always involves one or multiple sessions where an end-user asks questions and obtains precise answers. In this case, the answer generation involves the service retrieving information from the knowledge base based on the identified question classification, and then generating accurate, clear and professional answers. There are several steps for this function, such as request processing, validity checking, question classification, answer retrieval, and information return. The sequence diagram of function is shown in figure 2.

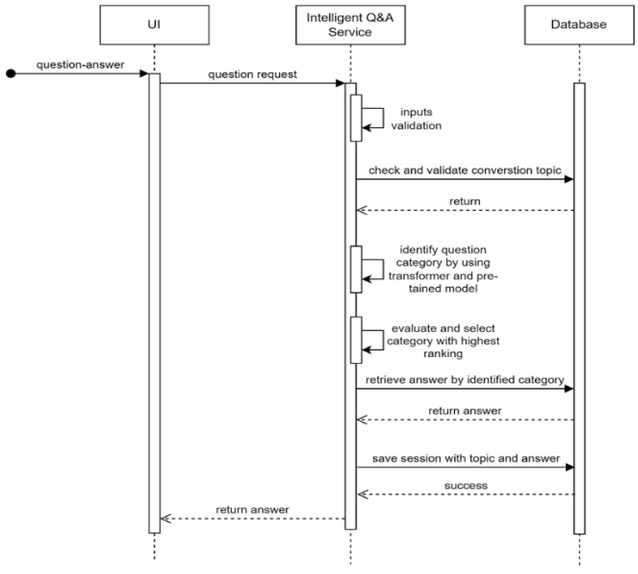

Figure 2

Architecture diagram of Intelligent Q&A Chatbot System

INFORMATION AND WEB TECHNOLOGIES

3 Knowledge Base: This component is mainly to store question categories and answers that related to the medical aesthetics industry, such as surgery periods, recovery periods, side effects, etc., and corresponding answers. It's using SQLite as the data storage.

This intelligent Q&A chatbot system uses the relational database SQLite to store information, which related to the knowledge of medical aesthetics domain, such as product recommendations, treatment plans, frequently asked questions, side effects and safety, beauty salon selection, as well as price consultations. An example of the knowledge base is shown in figure 3.

	category	content	creation_time	creator
1	InjectionPreparation	Before undergoing cosmetic injections, you should consult a professional doctor or aesthetic medical expert. ...	2024-09-22 18:58:00.538762	admin
2	Post-Laser Care	After laser spot removal treatment, avoid exposure to strong sunlight and use sunscreen to protect your skin. ...	2024-09-22 18:58:00.538762	admin
3	DoctorAndInstitutionSelection	Choose a certified and reputable medical institution and doctor. Check the doctor's qualifications and ...	2024-09-22 18:58:00.538762	admin
4	DoctorSelection	When choosing a doctor for cosmetic injections, it is important to work with an experienced physician. They wi...	2024-09-22 18:58:00.538762	admin
5	ThreadLiftDoctorSelection	Choosing the right thread lift doctor is important. Ensure that you select an experienced, professionally trained...	2024-09-22 18:58:00.538762	admin
6	Salon and Clinic Difference	Beauty salons mainly provide basic beauty services, while medical aesthetic clinics offer more professional ...	2024-09-22 18:58:00.538762	admin
7	Salon Services	Different beauty salons offer different services. Some common beauty services include skincare, hairdressing, ...	2024-09-22 18:58:00.538762	admin
8	Style and Positioning	The positioning and style of beauty salons can vary. Some salons may focus on fashion and trends, offering ...	2024-09-22 18:58:00.538762	admin
9	Qualifications and Training	Ensuring that the staff at a beauty salon have professional qualifications and experience is important. You can ...	2024-09-22 18:58:00.538762	admin
10	ProductBrands	Understanding the brands of products used by a beauty salon can help you evaluate their quality and suitabilit...	2024-09-22 18:58:00.538762	admin
11	Hygiene and Cleaning	Hygiene and cleaning are important factors to consider when choosing a beauty salon. You can inquire about ...	2024-09-22 18:58:00.538762	admin
12	Price	Understanding the price of services at beauty salons is important. You can inquire about the salon's pricing ...	2024-09-22 18:58:00.538762	admin
13	Customer Feedback	Understanding the feedback from other customers can provide a reference for the quality of services at a ...	2024-09-22 18:58:00.538762	admin
14	LocationAndTransportation	The location and transportation convenience of beauty salons may influence your choice. You can find out the ...	2024-09-22 18:58:00.538762	admin

Figure 3

The example data of knowledge base

The maintenance and updating of the knowledge base in this intelligent chatbot service system mainly include the following points: Internal and External Data Collection, Knowledge Extraction and Summarization, User Feedback and Evaluation, Regular Maintenance and Updates,

4 Data Processing: This involves collecting medical aesthetics-related data from various data sources (such as medical literature, specialist, drug information, website, etc.), data cleaning and preprocessing of the collected data, which including noise removal and normalize the data format to ensure data quality and consistency. The pre-processed data is unified in format and stored in CSV files.

5 Medical Aesthetics Model: This refers to the pre-trained language model in the medical aesthetics industry. It uses a model based on the BERT language model and utilizing

INFORMATION AND WEB TECHNOLOGIES

the Transformer library to train a medical aesthetics-specific vocabulary on pre-processed corpora, we subsequently obtained a specialized language model for medical aesthetics.

During the pretraining process, this involves pre-collected and processed corpora, such as large-scale medical texts and aesthetics reviews, to pretrain the BERT model. These text data come from medical literature, aesthetics product reviews and other fields, it's ensuring that the model has good generalization capabilities for medical aesthetics tasks. Additionally, the model would get the reasonable hyperparameter settings through continuous testing, such as a 4-layer Transformer encoder and a learning rate of 4e-5. The fine-tuned model would also evaluate its performance on the validation set, including the exploring loss functions and tuning the learning rate. Finally, the fine-tuned BERT model is integrated into our medical aesthetics system to analyze user-submitted questions in real-time.

In the medical aesthetics domain, users may ask questions that cover a wide range of topics, it's including the facelift surgery, surgical procedures, post-operative care, risks, outcomes and so on. To meet user needs and provides precise answers this system employs advanced natural language processing (NLP) techniques for question understanding: part-of-speech tagging and named entity recognition. There are 80 question categories supported by this system, it includes product recommendations, treatment plans, common FAQs, side effects and safety, beauty salon selection, price inquiries and so on which can help determine the category of the user's question, and it allows the system to accurately match the corresponding answers.

The intent recognition module uses a pretrained BERT model, which has been trained on a large amount of labeled data in the medical aesthetics industry, that can help to classify multiple intents of user questions properly. The system primarily enhances intent recognition capabilities through the following aspects: Fine-Tuning BERT, Intent Classification Layer, Improved Clarity, Model Application and Continuous Optimization. The question understanding and intent recognition are challenging tasks, so to enhance system`s accuracy and user experience it has chosen

INFORMATION AND WEB TECHNOLOGIES

appropriate training parameters such as learning rate, Adam with weight decay optimizer, and cross-entropy loss function, it has also combined the knowledge and experience of domain experts to fine-tune and improve the system.

After completing the development of each module, these components should be integrated as a comprehensive intelligent Q&A chatbot system. The goal of system integration is to ensure that all components work together seamlessly, this should provide a smooth and efficient user experience. In this system, it defines the interfaces and communication protocols (HTTP) between various components to ensure smooth data transmission. Additionally, the system has also defined the JSON as the standard data format of the payload, which enables different components to understand and process the same data structure. For example, while the frontend UI requests backend APIs that will use the JSON as payload data structure. The deployment diagram is shown as figure 4.

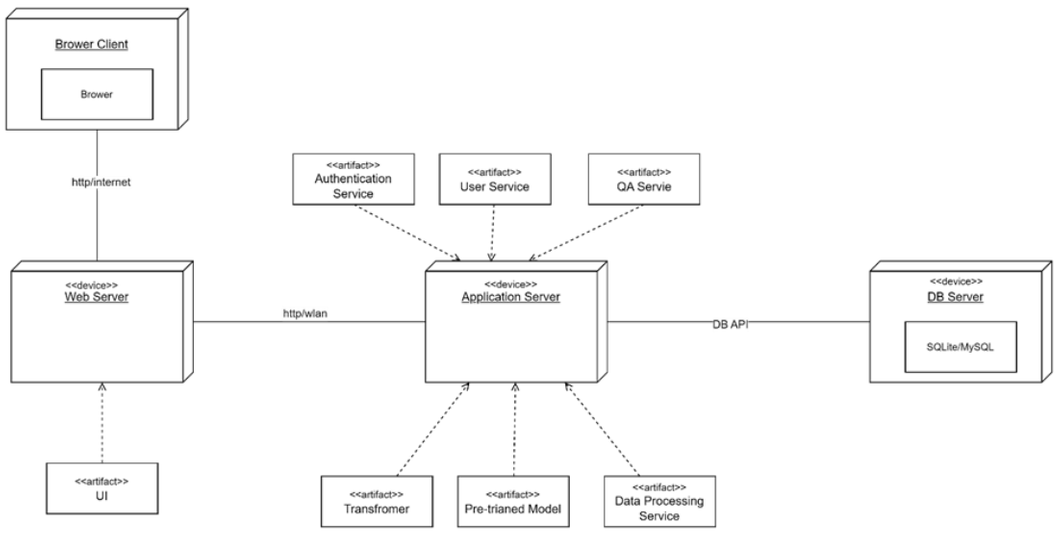

Figure 4

The deployment diagram of Intelligent Q&A Chatbot System

For physical machine deployment, we consider server configuration, database setup, and the installation of necessary dependencies. Cloud deployment can utilize resources from cloud service providers, such as Azure cloud

INFORMATION AND WEB TECHNOLOGIES

and database services, AWS cloud and database services. Afterward, the same commands used for physical machine deployment can be executed on the corresponding cloud servers.

For the testing results of this intelligent Q&A chatbot system, the analysis is conducted through the following aspects: evaluation metrics, system response time, and stability. By comparing the expected and actual returned answers using the test dataset, the accuracy for common questions can reach 90%. The sample of the test accuracy, precision, recall, and F1 score are shown in Table 1.

Table 1

The sample of evaluation metrics for Intelligent Q&A Chatbot System

Q&A	Precision	Recall	F1-score	Support
Clinic Services	0.85	0.93	0.89	30
Facelift Surgery	0.93	0.89	0.91	45
Aesthetic Procedures	0.94	0.90	0.92	50
Surgical Risks	0.85	0.88	0.86	25
accuracy	—	—	0.90	150
macro avg	0.89	0.90	0.89	150
weighted avg	0.90	0.90	0.90	150

For the system response time, through performance testing with Jmeter tool, the response speed of this intelligent Q&A chatbot system is impressive. It can achieve an average response time of 0.55 seconds for answering questions.

Conclusions. The developed intelligent Q&A chatbot system integrates several modules which provide a friendly interactive user experience and targeted, safe, and reliable information related to the medical aesthetics domain. Leveraging NLP techniques and pre-trained language models such as BERT allow the chatbot to understand the specific nuances and terminology of the medical aesthetics domain and accurately classify user questions and provide relevant, personalized responses.

Evaluation metrics demonstrate high accuracy and precision reflecting the system's robustness and reliability. Additionally, response time on user query is impressive ensuring a good experience for users.

INFORMATION AND WEB TECHNOLOGIES

Despite the good results, we may propose a series of further system improvement measures. These include adjustments to algorithms, model optimization and interface design enhancements, they aimed at improving system performance and user experience. One the one hand, the model's hyperparameters for the algorithm can be adjusted, such as learning rate, batch size, and the number of layers, to enhance model performance. Additionally, more suitable algorithms can be selected based on the characteristics of the problems. On the other hand, incorporating deep learning methods can be involved, such as Recurrent Neural Networks (RNNs). Lastly, enhancement in interface design will also be considered to improve the overall user experience.

References:

[1] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
[2] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532-1543.
[3] Peters, M. E., Neumann, M., Iyyer, M., et al. (2018). Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
[4] Liu, Q., Chen, E., Yu, Y., et al. (2021). A Survey on Deep Learning for Named Entity Recognition. IEEE Transactions on Knowledge and Data Engineering, 33(1), 1-21.
[5] Yang, Z., Dai, Z., Yang, Y., et al. (2019). XLNet: Generalized autoregressive pretraining for language understanding. Advances in Neural Information Processing Systems, 32, 5753-5763.
[6] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends in Deep Learning Based Natural Language Processing. IEEE Computational Intelligence Magazine, 13(3), 55-75.
[7] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
[8] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems, 30, 5998-6008.