Design Engineering

ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

Automatic Question Answering Technology of Attention Mechanism Based on Knowledge Graph

Hongzhi Yu, Fucheng Wan,Dengyun Zhu,Qi Guo, Dongjiao Zhang

Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China

Abstract:

The automatic answer technology of search engine and the automatic answer system based on intelligent machine lack data cleaning steps, and there is a problem that the automatic question answering effect is poor due to the low matching correct answer rate. A research on automatic question answering technology based on Attention mechanism of knowledge graph is proposed. Stores and manages knowledge graphs through Jena. It goes through three processes of ontology data cleaning, entity graphing and linking, and data type labeling, and integrates the knowledge graph oriented to question answering system. The Attention mechanism is used to retrieve automatic question-and-answer keywords, continuously adjust parameters, analyze the relationship between words, and design the automatic question-and-answer process. The experimental analysis is carried out in PyCharm environment, and the matching correct answer rate and the matching result evaluation mechanism are taken as the experimental evaluation indexes. From the experimental results, it can be seen that the matching correct answer rate of this technology can reach up to 93%, which can highlight the user preference and supplement the problem expansion research.

***Keywords**: Knowledge graph, Attention mechanism, Automatic question and answer, Retrieve.*

I. INTRODUCTION

Most of the information on the Internet is expressed in the form of HTML pages, which is very suitable for people to understand. However, with the explosive growth of information, it is difficult for people to find the required information in massive and complex information. Under this background, search engines came into being [1]. The basic principle of this method is to collect information from the network through crawlers and establish an inverted index based on keywords, which provides information retrieval services for users. Users use keywords to describe their query purpose [2]. Based on a sort algorithm, the search engine presents users with information that meets the query criteria [3]. Automated answering system is a new trend

ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

of natural interaction between human beings and machines. They can more accurately understand user questions described in natural language and provide more accurate answers according to users' actual intentions, which will be the next generation of new search engine forms. The answering sources of the traditional data answering system include unstructured text data, such as web documents, search engines, encyclopedia descriptions, and question and answering communities [4].

The appearance of this kind of search engine solves the problem that users obtain information from the Internet to a certain extent. However, because these traditional search engines are based on keywords or strings, their limitations are also very obvious. They do not understand the purpose of the query (usually web pages) and the user's query input. The result is that the search accuracy has obvious defects [5]. In the end, this is because HTML format Web pages' lack semantics, making it difficult for computers to understand. The algorithm using natural language processing is an important aspect of artificial intelligence [6]. Through the use of natural language processing technology, computers can process and analyze natural language, effectively recognize text data, organize and analyze. In the field of natural language processing, named entity is a basic and important concept. Each named entity corresponds to a specific thing or concept, which contains important semantic information. In the field of natural language processing, the research on named entities has always been the foundation and focus. However, in natural language, there are many abbreviations, irregular words and ambiguous words. At the same time, many entities have multiple synonyms and polysemies, which makes the machine unable to determine entities when processing unstructured free text. Accurate reference relations will affect the performance of the whole natural language processing system. In order to solve this problem, the automatic answering technology based on knowledge graph attention mechanism is studied. With the development of big data and knowledge engineering, unstructured data is constructed into a knowledge graph through information extraction, knowledge processing and other technologies, making massive data better close to the human cognitive form, understanding, organization and management of structured and related information. The concept of knowledge graph is thus put forward, and the research and application direction of automatic answering system based on knowledge graph in academia and industry have changed greatly. The automatic answering system based on knowledge graph is an answering method based on knowledge graph. Knowledge graph plays a key role in accurate question-and-answer service.

II. CONSTRUCTION OF KNOWLEDGE GRAPH OF ATTENTION MECHANISM

The key to adopting attention mechanism in the retrieval process is to combine knowledge graph and identification to improve the accuracy and efficiency of retrieval.

2.1 Knowledge Graph Storage and Management

The In essence, knowledge graph is a knowledge base with graph structure, which is called semantic network. Nodes in the chart can represent concepts or entities, and the knowledge

Design Engineering
ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

graph mainly includes two parts: fact statement and ontology. Notional words mainly describe the relationship between entities and entity attribute values, mainly including subject, predicate and object [7-9].

A subject is usually an entity, and a predicate represents an attribute of a subject. It is mainly divided into two categories: data type and object type. Objects vary from predicate to predicate. Objects with data type attributes are specific data, while objects with object type attributes are specific entities [10-12].

Attention mechanism automatic question and answer data needs to be managed and maintained through relevant database management systems. Jena realizes data management and query, and realizes common operations such as data addition, deletion, modification and query [13-14]. The Jena framework structure is shown in Fig 1.

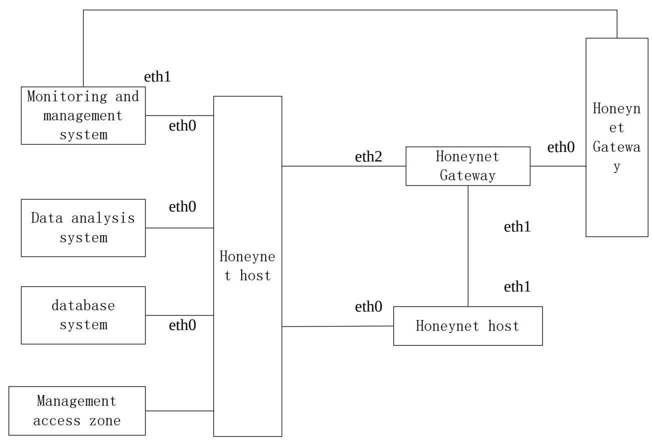


Fig 1: Jena Framework Structure

2.2 Knowledge Graph Fusion for Question Answering System

In this model, you can get a knowledge diagram supporting complex SPARQL queries by annotating and linking objects through object attributes and numerical attributes to object data types. The specific steps of integration include: ontological data cleaning, entity graphing and linking, and data type labeling.

(1) Ontology Data Cleaning

Because the entity classification triple for the attention mechanism to automatically answer questions is extracted from the DBpedia section, the name of the entity includes simplified Chinese and traditional Chinese [19]. Therefore, the first thing to do is to complete the transformation from tradition to simplification. Without distinguishing between traditional and simplified transformations, incorrect entity name transformations will result. For the combination of three elements of entity classification, there are few similar situations. As other elements are manually screened, they are automatically converted through the traditional simplified conversion program [20].

(2) Entity Diagrams and Links

ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

For entity graphing and linking, the core problem is entity linking.

The purpose is to make the object correctly point to semantic entities. Refer to returning a list of related entities and indicate the most commonly used entities. Using API linked entities provided by the knowledge graph as input, and then linking entity synonym to query the most commonly used entities and objects returned by the interface can greatly improve the accuracy of entity linking [21]. In addition to the general entity priority principle, string similarity comparison should also be added [22]. The similarity of strings can be calculated by Levenshtein distance algorithm. As can be seen from Equation 1, L1 and L2 represent the lengths of the source string and the target string respectively, while Levenshtein represents the editing distance of the source and the target string.

$$Similarity = (\mathrm{Max}(\mathrm{L}_1, L_2) - \mathrm{Levenshtein}) / \mathrm{Max}(\mathrm{L}_1, L_2) \quad (1)$$

Since entities with the same name are mainly distinguished by the content after square brackets, and the content after square brackets is most likely to appear in triples, only the most commonly used entity links generate erroneous results. If string similarity matching is added to the square brackets after the fuzzy entity name, the correct entity link result can be obtained [23]. The description content of the entity and the triple of the linked entity are calculated for string similarity. If the similarity is greater than a certain threshold, it is linked to the entity. In addition, the entity synonym graph I of the knowledge graph can be used to link the most commonly used entities.

(3) Data Type Labeling

After the above two steps, a basic knowledge graph is initially obtained, but the Data Properties object in the knowledge graph has no marked data type, so the database treats it as a string type by default. This kind of knowledge graph does not support SPARQL number comparison query, which requires identification and marking of data types of data attribute objects.

A feasible method is as follows: Firstly, most non-string data attributes are sorted and screened according to the occurrence frequency of attributes in the knowledge layer; Then, write regular expressions such as Boolean values, integers, floating-point numbers and date-related expressions, capture data based on the form displayed in the knowledge graph and embed it into the standard data format; The manually selected general attributes are matched to the target, and finally the data type label is obtained according to the matching result.

2.3 Graph Construction

The traditional automatic question answering keyword retrieval method based on attention mechanism usually automatically identifies the relationship between historical data and predicted data. If the correlation between input and output is not considered, the influence weight of each input trajectory on the output result is the same. After many cycles of training, the results are still not accurate enough, not only with low efficiency, but also with low accuracy. It should be noted that the current computer system outputs the hidden layer state value to its own matching degree, i.e. outputs the hidden layer state value h. It has three steps:

ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

(1) The algorithm uses the matching layer to calculate the matching degree; The matching layer module calculates the current state value HT of the hidden layer and its own matching degree, calculates the matching layer for each HT according to the state value output by the hidden layer, and calculates the similarity of the output according to the matching layer.

(2) Because the similarity calculation ut is not normalized, the Softmax function is used for normalization so that the sum of weight weights is 1 at output. Therefore, the weight of the output is.

(3) Therefore, the weighted vector sum of the hidden layer state value ht and the output weight at is calculated to obtain the position of the next time Yt. Fig 2 shows the structure of knowledge graph based on attention mechanism.

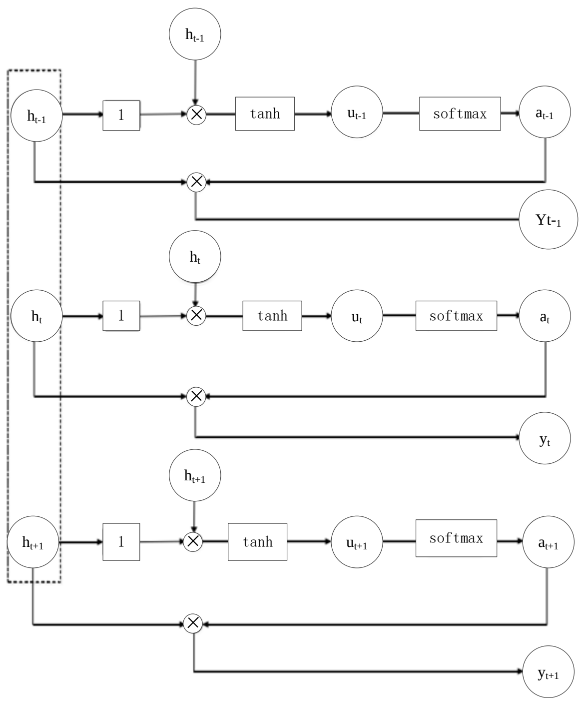


Fig 2: Knowledge graph structure base on Attention mechanism

The state value of the hidden layer at different times is ht, which contains the cell state values of many hidden layers, i.e. $h_t = \left(h_t^1, h_t^2, h_t^3, \cdots h_t^q\right)$, the normalized output weight is $a_t = \left(a_t^1, a_t^2, a_t^3, \cdots a_t^q\right)$

The knowledge graph structure based on Attention mechanism is defined as follows:

ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

$$u_t = \mathrm{Vtanh}\left(W_1^{q\times q}h_t + W_2^{q\times q}h_t\right)$$
$$a_t = \mathrm{softmax}\left(u_t\right) \qquad (2)$$
$$y_t = \sum_1^q a_t^i * h_t^i$$

According to the current hidden layer state values calculated by the pairing calculation equation (2), the pairing degree ut of the current hidden layer state values is respectively obtained. In this way, by using the Softmax function in Equation (2), the weight weight when the output is 1 is added up, and the output weight of each unit can be expressed, and its weight vector and Yt can be obtained. In this way, the attention mechanism can effectively identify the items that play an important role in the input vector, and continuously adjust the parameters through many times of training and learning to achieve higher retrieval accuracy.

III.RELATIONSHIP BETWEEN WORDS AND WORDS

On the basis of attention mechanism, an intelligent answering model is established. It includes input layer, embedded layer, hidden layer, full connection layer and output layer. Fig 3 shows its structure diagram.

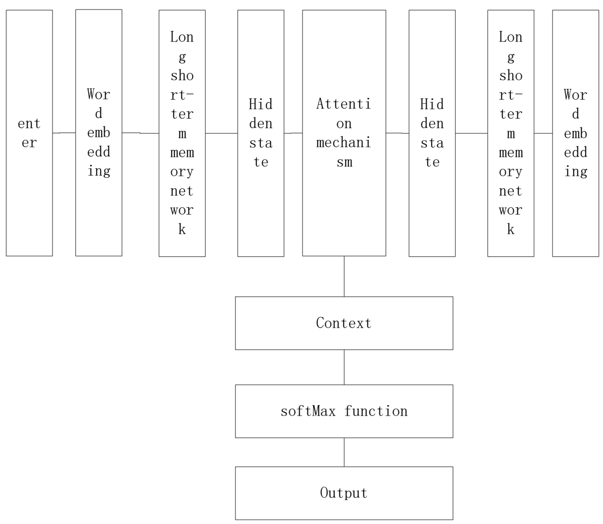


Fig 3:Word and Word Relationship

After research, it is found that the original automatic answering technology is based on the isomorphism of "one-hot", that is, each identity in the text is replaced by a number, but it is not enough to express the relationship between isomorphisms. Therefore, the text is input using Word2VEC coding instead of (one-hot). Word2VEC can encode each of the same vectors with multi-dimensional vectors and use the same embedding as a model to display the relationship between words.

IV.DESIGN OF AUTOMATIC QUESTION ANSWERING PROCESS

Design Engineering

ISSN: 0011-9342

Issue: 11 | Pages: 685 - 695

Under the construction of Attention mechanism knowledge graph, the automatic question answering process is designed, as shown in Fig 4.

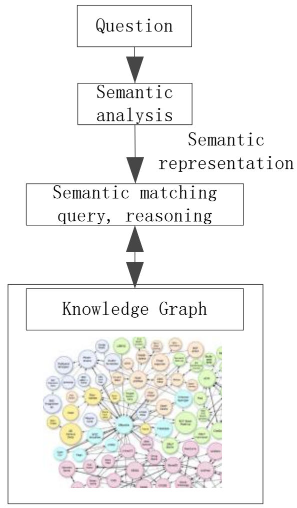


Fig 4:Automatic Question Answering Process Design

Through semantic analysis and understanding of natural language problems raised by users, and then using the existing knowledge graph for query and reasoning, the optimal answer is obtained. In the automatic answering system based on files, there is usually a document library containing a large number of files, each of which corresponds to a topic, and the files contain a certain number of sentences, which is the candidate answer of the answering system. The goal of this system is to answer users' questions and only select the sentences that are considered to match best from the sentences that make up the document.

V.EXPERIMENTS

In order to verify the rationality of the research on automatic question answering technology of Attention mechanism based on knowledge graph, experimental verification analysis is carried out.

5.1 Experimental Preparation

The experimental environment is developed in PyCharm environment. The computer is configured as Inter (R) Core (TM) i5-7400CPU, with 16GB of memory and Win10 operating system.

The data set used in the experiment is Microsoft's developmental answer data set. The data set is mainly divided into two parts, namely, the training set test set, in which the training set consists of 10025 questions and 2215231 answer pairs. The test set consists of 5,000 questions and 100,000 answers are correct.

5.2 Experimental Evaluation Index

The experiment uses two evaluation indexes, namely, the matching correct answer rate and the matching result evaluation mechanism M .

Design Engineering
ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

(1) Correct Matching Answer Rate

The equation for calculating the correct answer rate for matching is as follows:

$$P = \frac{h}{Q} \quad (3)$$

(2) Evaluation Mechanism of Matching Results

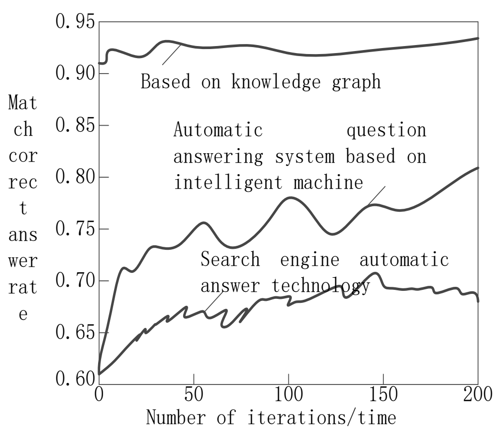
The calculation equation of the matching result evaluation mechanism is as follows:

$$\mathrm{M} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{\mathrm{rank}_t} \quad (4)$$

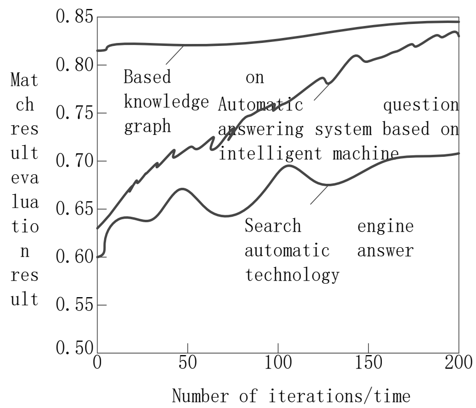
In Equation (4), indicates the ranking position occupied by real values in the answer set for the best answer to the question.

5.3 Experimental Results and Analysis

Taking Attention mechanism as an index, keywords are selected: user attributes, age, occupation, interests and hobbies, and number of people. Five parameters are sampled evenly as inputs, and then the logarithmic loss function is evaluated through multiple classifications. The search engine automatic answer technology, the intelligent machine-based automatic question answering system and the knowledge graph-based analysis matching correct answer rate and matching result evaluation results are respectively used, as shown in Fig 5.



(a) Ratio of correct matching answers



(b) Evaluation results of matching results

Fig 5:Comparison of Effects of Two Evaluation Indexes

As can be seen from the above figure, using the automatic answer technology of search engine respectively, the matching correct answer rate and matching result evaluation result of the automatic question answering system based on intelligent machine are lower than those of Attention mechanism. Among them, the two evaluation indexes of automatic question answering system based on intelligent machine reach 80% at the highest, while the two evaluation indexes of automatic answer technology of search engine are lower than 70%. The Attention mechanism shows good performance and convergence speed, mainly weighting features, so that important feature information plays a greater role in the automatic answer process.

VI.SUMMARY AND PROSPECT

6.1 Summary

The knowledge graph is fused and a knowledge graph supporting complex queries is proposed. This paper studies the current mainstream open Chinese knowledge graphs, integrates and improves the two open download knowledge graphs, and adds entity links and data type labeling attributes. Finally, through fusion and improvement, the obtained knowledge graph supports the query of complex problems, such as multi-entity relations, data comparison, etc. and subsequent Chinese retrieval. In order to solve the fact-oriented natural language problems, a natural language problem classification system is designed. There are many types of natural language questions, many of which are related to the backbone of the respondents, or there are no exact answers, and some people cannot even answer them. For questions with clear experience, clear answers are the basis of research technology.

Attention mechanism is adopted to selectively learn the intermediate state values to improve the accuracy of automatic question and answer. The experimental results show that the matching correct answer rate of the technology can reach 93%, with good performance and convergence speed.

6.2 Prospect

Although the research content of this paper has achieved preliminary results, there are still some problems to be further studied. In the improved fusion knowledge graph, the triples covered by ontology are not comprehensive enough, and many triples do not classify ontology accordingly. In the general knowledge graph, it is a problem worth studying on how to classify entities.

At the same time, the iterative learning of knowledge graph also puts forward the corresponding solution for updating knowledge graph. In the automatic knowledge learning algorithm, the information edited by the user can be further used for feedback to improve the performance of the algorithm through continuous iteration. For conflict resolution, the algorithm is based on rules. Set these rules according to a specific data source. These rules may fail when using a new data source. However, when the knowledge in the knowledge graph is continuously expanded and updated, the knowledge in the knowledge graph itself can be used as an important measurement index of knowledge reliability, and on this basis, corresponding conflict solutions are designed.

REFERENCES

[1] Gao Long, Zhang Hanchu, Yang Liang. (2018) Research on intelligent information search technology based on knowledge graph and semantic computing. Intelligence Theory and Practice 041 (007): 42-47.

[2] Zhang Zhaofeng, Zhang Junsheng, Yao Changqing. (2018) An automatic construction method of technical efficacy graph based on knowledge graph. Intelligence Theory and Practice 041 (003): 149-155.

[3] Qi Bin, Wang Yu, Zou Hongxia, et al. (2018) Technology of network security knowledge graph selection based on information entropy and fuzzy sets. Minicomputer Systems 39 (11): 48-52.

[4] Zhu Guiming, Bin Chenzhong, Gu Tianlong, et al. (2019) Neural modeling framework of user preference based on knowledge graph. Pattern Recognition and Artificial Intelligence 32 (007): 661-668.

[5] Wang Dianli, Wang Zheng. (2019) Knowledge graph, research hotspot and expansion space of foreign public value theory research--visual analysis based on SSCI (1998-2018). China Administration 000 (006): 103-110.

[6] Fang Yang, Zhao Xiang, Tan Zhen, et al. (2018) An improved representation method of knowledge graph based on translation. Computer Research and Development 55 (1): 139-150.

[7] Li Yanyan, Zhang Xiangling, Li Xin, et al. (2019) Construction and innovative application of subject knowledge graph for wisdom education. Research on Audio-visual Education 10 (8): 60-69.

[8] Xie Gang. (2018) Review of the research on the refinement of knowledge graph. Application of Electronic Technology 044 (009): 29-33.

[9] Qi Ershi, Tian Tian, Liu Liang. (2018) Visualization analysis of internet business model research based on knowledge graph. Research on Science and Technology Management 7 (4): 190-196.

[10] Yang Fangquan. (2018) Design of intelligent question answering system based on big data knowledge graph of packaging industry. Modern Electronic Technology 41 (04): 151-154.

ISSN: 0011-9342
Issue: 11 | Pages: 685 - 695

[11] Chen Jiangtao, Lv Jianqiu. (2018) Research status and hot spot analysis of agricultural technology extension in China based on knowledge graph. Research on Science and Technology Management 38 (004): 175-180.

[12] Xiao Ding, Wang Ganyu, Cai Ming, et al. (2019) Research on implicit conflict detection method based on knowledge graph in smart home scene linkage. Acta Computer Science 15 (6): 1190-1204.

[13] Liu Jie, Liu Linzhi. (2018) Research on industrial innovation evaluation and forecast model based on knowledge graph analysis. Modern Intelligence 38, 324 (06): 140-145.

[14] Liu Ziquan, Wang Huifang. (2018) Retrieval method of power equipment defect records based on knowledge graph technology. Power System Automation 042 (014): 158-164.

[15] Lu Qiang, Liu Xingyu. (2018) Semantic matching model of knowledge graph question and answer based on transfer learning. Computer Applications 38 (07): 1846-1852.

[16] Huang Lingyun, Dai Yongwu. (2018) Evolution of research frontier of forestry finance at home and abroad in the past 30 years-based on visualization of knowledge graph. Forestry Economic Issues 038 (001): 87-98.

[17] Yang Jingzhao, Zang Min, Zhen Meirong. (2018) Team fault zone and innovation based on knowledge graph: analysis of hot spots and trends. Scientific and Technological Progress and Countermeasures 035 (015): 144-151.

[18] Guan Saiping, Jin Xiaolong, Jia Yantao, et al. (2018) Research progress of knowledge reasoning based on knowledge graph. Acta Software Sinica 29 (10): 2966-2994.

[19] Sun Qiangqiang, Jiang Wanbei, Sun Danfeng. (2018) Scientific environmental knowledge mining based on knowledge graph and syndrome--a case study of desertification in Minqin. Arid Region Geography 041 (002): 426-434.

[20] Peng Min, Yao Yalan, Xie Qianqian, et al. (2019) Joint knowledge representation model based on CNN with attention mechanism. Chinese Journal of Information Technology 33 (2): 51-58.

[21] Zhao Dongmei, Li Ya, Tao Jianhua, et al. (2018) Affective analysis model based on collaborative filtering Attention mechanism. Chinese Journal of Information Technology 032 (008): 128-134.

[22] Zheng Xiongfeng, Ding Lixin, Wan Runze. (2018) Hierarchical BGRU model based on user and product Attention mechanism. Computer Engineering and Application 054 (011): 145-152.

[23] Zhao Yunshan, Duan Youxiang. (2019) Convolution neural network text classification model based on Attention mechanism. Journal of Applied Sciences 37 (004): 541-550.