**Abstract**

With the advent of the big data era, knowledge engineering has attracted wide attention, as mining knowledge from large-scale data is critical for big data analysis. Knowledge graph techniques provide a way to extract structured knowledge from large-scale texts and images, thus having a wide application prospect. The data of the military has a high sensitivity and confidentiality, therefore it has high value. Through a large amount of military data to achieve a human-computer interaction platform has an important strategic significance. Through the association of military equipment information, military base information and military combat force information in the form of a knowledge graph, it can not only recommend operational strategies, but also provide better training for military personnel. In this article, we realize a semantic search and question-answering system based on the military knowledge graph by crawling the semi-structured and unstructured military data, storing them in different databases and achieving knowledge fusion.

**1.Related Work**

Knowledge graph is a technical method that uses graph models to describe the relationship between knowledge and modeling the world [1]. The early idea of the knowledge graph comes from Semantic Web [2,3]. Its original idea is to transform the World Wide Web based on text links into a semantic web based on entity links. In November 2012, Google took the lead in proposing the concept of Knowledge Graph (KG), indicating that it would add the function of Knowledge Graph to its search results. Its original intention is to improve the capabilities of search engines and enhance the search quality and search experience of users. According to statistics in January 2015, the KG built by Google has 500 million entities and about 3.5 billion entity relationship information, which has been widely used to improve the search quality of search engines. Although the concept of Knowledge Graph is relatively new, it is not a completely new research field. As early as 2006, Berners Lee proposed the idea of ​​linked data, calling for the promotion and improvement of related technical standards such as URI (Uniform resource identifier), RDF (Resource discription framework), OWL (Web ontology language), prepare for the arrival of the semantic network. Subsequently, the knowledge graph is gradually used in semantic search [4,5], intelligent question and answer [6-8], auxiliary language understanding [9,10], auxiliary big data analysis [11-13], and enhance the interpretability of machine learning [14], combined with graph convolution to assist image classification [15, 16] and other fields play an increasingly important role.

Knowledge-based Question Answering (KBQA) is the core function of the intelligent question answering system and a natural way of human-computer interaction. Knowledge Q&A relies on a large knowledge base (knowledge graph, structured database, etc.), which converts users' natural language questions into structured query sentences (such as SPARQL, SQL, etc.), and directly derives the answers required by users from the knowledge base.

In recent years, with the strong demand of users for intelligent applications, many companies and institutions such as Google, Baidu, Wiki, etc. have adopted automatic or semi-automatic methods to design a series of complete knowledge graphs for the high-quality data obtained. For example, DBpedia [17], Freebase [18], YAGO [19], etc., this type of graph is composed of a large number of entities, relationships and attributes. At the same time, the development of machine learning and deep learning has laid a research foundation for intelligent question answering. For example, Baseball [20], Lunar [21] realized intelligent question answering with limited domains, Paralex [22], SEMPRE [23], ParaSEMPRE [24], STAGG [25] tried more challenging open domain intelligent question answering research.

In 2014, Ye et al. [26] pointed out that the key to solving single-knowledge point question and answer is to decompose the original task into two subtasks-topic word recognition and relationship detection. In 2015, Ye et al. [27] emphasized that it is more difficult to find the relationship that matches the meaning of the question directly from a large knowledge base. In the thesis, Entity Linking is first used to locate topic words, and then from the subset of relations related to topic words to find the relationship that matches the meaning of the question, the question is parsed into a structured query. In 2016, Ye et al. [28] inherited the open source WebQuestions data set of the Stanford Natural Language Processing Group, and based on this, annotated the semantic analysis results of the question (SPARQL query), and contributed to the WebQuestionsSP data set. In 2019，Yang et al. [29] demonstrate an end-to-end question answering system that integrates BERT with the open-source Anserini information retrieval toolkit. In 2020，Ma et al. [30] proposes an answer quality prediction model based on the question-answer joint learning (ACLSTM) which can effectively improve the prediction effect of answer quality.

1. **Preliminary design**

2.1 System architecture

As is shown in Figure.1, our system architectures can be seen as a system in three layers, among which exists multiple interfaces to help different layers conenect with each other as well as transporting information. In concrete, we capture and colloect semi-structured and unstructured data from external data in websites. Then, the DBMS(Database Management System) layer store these data into different kinds of database like mysql, mongodb and neo4j in terms of data features. Next we construct the military knowledge graph in KG construction layer by using millions of entities, 10 millions of triples and extra ontology like conceptions, properties and axioms. And in the top application layer, KG can be utilized to develop some useful applications like search engine, query answering robot and data visualization, etc.

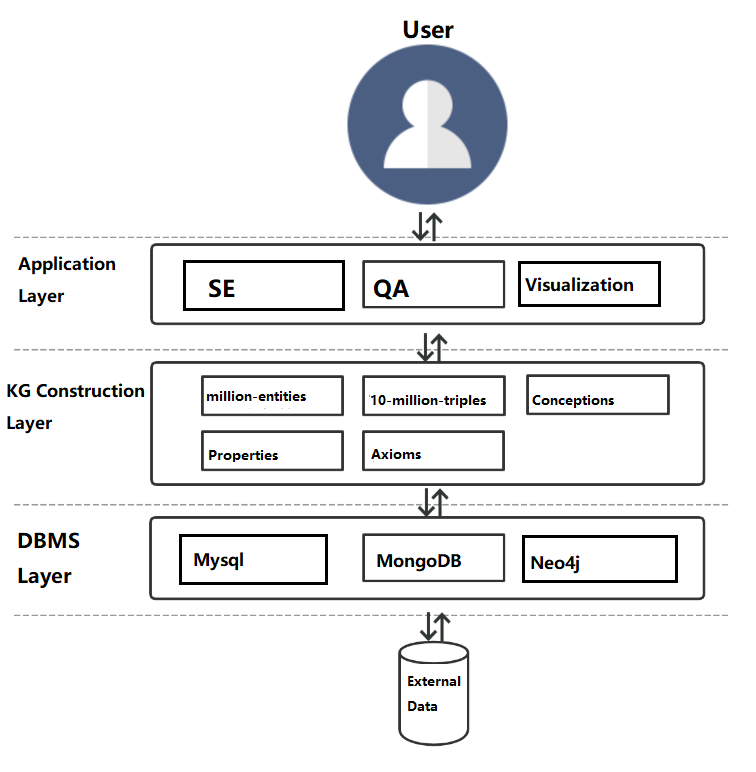


Fig.1: System architecture

2.2 Key modules and their functions

According to the military KG , we plan to develop three modules in our system (Fig.2), which are dependent from each other. In information retrieval module, the user can search for the information especially those in military field, like news and entities. In query answering module, a cute robot is happy to answer questions from users. In data visulization module, questions about an entity can be linked to its related entities to show more details.

Fig.2: Key modules

2.3 Key techniques and challenges

The research content of this project aims to realize knowledge inference and prediction of military knowledge graph by studying rule-based parallel inference technology and knowledge graph embedded representation learning technology through multi-source heterogeneous knowledge such as military knowledge graph and military detection log. In order to effectively improve the accuracy and scalability of reasoning.This project will focus on modeling of machine learning algorithms such as knowledge graph representation learning and Markov chain Monte Carlo algorithm, and research will focus on semantic model and representation learning technology.The specific difficulties of the current research are as follows:

(1) the uncertain and inconsistent reasoning in heterogeneous knowledge reasoning: on the one hand, the data source is multiple in the field of military, which leads to task concurrency and incompatible phenomenon, at the same time, uncertain and inconsistent reasoning tends to bring high complexity. How to improve the compatibility efficiency of uncertain and inconsistent reasoning is a difficulty. On the other hand, the traditional interpretable military reasoning methods, such as language description, attack graph and attack tree, are insufficient in terms of extensibility and global nature. They can only make preliminary analysis, but can't locate and grasp complex problems from a global perspective.In addition, at present, both uncertain and inconsistent reasoning are non-monotonous reasoning with great complexity. Therefore, the construction of high-performance inference interpretation model, especially the design of minimum inference process model, is a difficulty.

(2) Representation and reasoning of military framework knowledge: Knowledge map is mainly composed of framework knowledge (e.g., ontology axiom) and factual knowledge (e.g., triple).At present, most representation methods based on knowledge graphs focus on the factual knowledge in knowledge graphs rather than the framework knowledge in knowledge graphs.Framework for knowledge reasoning and prediction technology, knowledge of the relationship between semantic and characteristic is very important, military coalition reasoning being applied in the scene is very common, the semantic properties can not only help to improve the accuracy of military knowledge graph and the scope of the cover but also help to complete knowledge graph structure to evaluate its analysis.

(3) Mixed reasoning based on symbolic rules and machine learning: military knowledge usually consists of military knowledge base and military rule base.The former benefits from the data accumulation of information words, while the latter is manually edited and maintained by experienced commanders or combat experts.However, the reasoning before the two is independent of each other.If these two kinds of knowledge can be combined with constraints effectively, then more meaningful battlefield information can be mined and deduced, which is a major difficulty in the current joint model design.

In view of the above difficulties, the following key technologies are adopted to solve the problem:

(1) In view of the high complexity of uncertain and inconsistent reasoning, this subject establishes a parallel reasoning theory compatible with uncertain and inconsistent reasoning, and designs corresponding reasoning algorithms on this basis.The existing knowledge division theory and method are used to support the knowledge parallel inference technology, and the knowledge stratification theory and method are designed to support the iterative execution of knowledge inference.For minimum inference and reasoning process interpretation model of the existing problems of methods through the establishment of the facts of the input information one step, the key points in the knowledge base and inference conclusion as explanation space node, as the relationship between nodes in a single step reasoning explanation, on the edge of space, and explain the reasoning conclusion will be as the root, with the facts of the input information as leaves, interpretation tree dug up from the explanation space, according to the explanation in the tree node on the uncertainty of measurement explain the size of the tree, from explanation space heuristic search algorithm to efficiently search the minimum interpretation tree form minimum inference process,Then it provides a reasonable explanation for battlefield commanders in the face of complex battlefield environment.

(2) In view of the representive learning integrated with frame knowledge, this project is put forward based on Bert + BiLSTM + CRF frame knowledge representation method, and the integration of military knowledge graph within the context information and external text information to complete the modeling of hybrid knowledge representation and reasoning and prediction, so as to improve the accuracy of knowledge graph and the scope of cover.In addition, this project further improves the reasoning and prediction accuracy of the mixed embedding model by using the internal context information and external text information of the military knowledge map, so as to solve the problems of low accuracy and poor anti-noise ability of the present representation learning technology.

**References**

1. Singhal,Amit.Introducing the Knowledge Graph:Things,Not Strings.Official Blog （of Google）, 2012.
2. Berners-Lee Tim,James Hendler,Ora Lassila.The Semantic Web.Scientific American,2001.
3. Shadbolt Nigel,Wendy Hall,Tim Berners-Lee.The Semantic Web Revisited.IEEE Intelligent Systems, 2006.
4. Guha R,McCool R,Miller E.Semantic search.Proceedings of the 12th international conference on World Wide Web.ACM,2003:700-709.
5. DONG X,Gabrilovich E,Heitz G,et al.Knowledge vault:A Web-Scale Approach to Probabilistic Knowledge Fusion.Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2014:601-610.
6. CUI W,XIAO Y,WANG H,et al.KBQA:Learning Question Answering Over Qa Corpora and Knowledge Bases.Proceedings of the VLDB Endowment,2017,10（5）:565-576.
7. YAO X,Van Durme B.Information Extraction over Structured Data:Question Answering with Freebase[C]//ACL.2014:956-966.
8. HAO Y,ZHANG Y,LIU K,et al.An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge.Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics （Volume 1:Long Papers）,2017:221–231.
9. YANG B,Mitchell T.Leveraging Knowledge Bases in LSTMs for Improving Machine Reading.Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics （Volume 1:Long Papers）,2017:1436–1446.
10. WANG J,WANG Z,ZHANG D,et al.Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification.Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence,2017:2915–2921.
11. Kaminski M,Grau B C,Kostylev E V,et al.Foundations of Declarative Data Analysis Using Limit Datalog Programs,2017（2）:1123–1130.http://arxiv.org/abs/1705.06927.
12. Bellomarini L,Gottlob G,Pieris A,et al.Swift Logic for Big Data and Knowledge Graphs.IJCAI2017, 2017:2–10.
13. CHEN J Y,Freddy Lécué,Jeff Z Pan,CHEN H J.Learning from Ontology Streams with Semantic Concept Drift.IJCAI 2017,2017:957-963.
14. CHEN J Y,Freddy Lécué,Jeff Z Pan,et al.Transfer Learning Explanation with Ontologies.International Conference on the Principles of Knowledge Representation and Reasoning .KR2018.
15. WANG X,Ye Y,Gupta A.Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2018: 6857-6866.
16. Lee C W,FANG W,Yeh C K,et al.Multi-Label Zero-Shot Learning with Structured Knowledge Graphs.arXiv preprint.arXiv:1711.06526.2017.
17. Bizer C, Lehmann J, Kobilarov G, et al. DBpedia - a crystallization Point for the Web of Data[J]. Journal ofWeb semantics, 2009, 7(3):154-165.
18. Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008: 1247-1250.
19. Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge[C]//Proceedings of the 16th international conference on World Wide Web. ACM, 2007: 697-706.
20. Green B F, Alice K W, Chomsky, et al. Baseball: An Automatic Question Answerer[C]//Proceedings of the Western Joint Computer Conference. IEEE Computer Society, 1961:219-224.
21. Woods W A. Progress in natural language understanding: an application to lunar geology[C]//International Computer Conference, 1973: 441-450.
22. Fader A, Zettlemoyer L, Etzioni O, et al. Paraphrase-Driven Learning for Open Question Answering[C]// Meeting of the Association for Computational Linguistics, 2013: 1608-1618.
23. Berant J, Chou A K, Frostig R, et al. Semantic Parsing on Freebase from Question-Answer Pairs[C]//Conference on Empirical Methods in Natural Language Processing, 2013: 1533-1544.
24. Berant J, Liang P. Semantic Parsing via Paraphrasing[C]// meeting of the association for computational linguistics, 2014: 1415-1425.
25. Yih W, Chang M, He X, et al. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base[C]//international joint conference on natural language processing, 2015: 1321-1331.
26. Yih W–t,He X,Meek C.Semantic Parsing for Single-Relation Question Answering.Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,2014,2:643-648.
27. Yih W-t,CHANG M W,He X,et al.Semantic Parsing Via Staged Query Graph Gen- Eration:Question Answering with Knowledge Base.In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing,2015,1:1321-1331.
28. Yih W-t,Richardson M,Meek C,et al.The Value of Semantic Parse Labeling for Knowledge Base Question Answering.Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,2016,2.
29. Yang W, Xie Y, Lin A, et al. End-to-End Open-Domain Question Answering with BERTserini. 2019. doi:10.18653/v1/N19-4013.
30. Weifeng Ma, Jiao Lou, Caoting Ji, & Laibin Ma. (2021). ACLSTM: A Novel Method for CQA Answer Quality Prediction Based on Question-Answer Joint Learning. Computers, Materials & Continua, 66(1), 179.