Optimization of Computer Aided Query System for Tourism Information Based on Internet of Things

Bin Liu[1] and Weiwei Jiang[2*]

[1]Huihua College, Hebei Normal University, Shijiazhuang, Hebei 050091, China,
liubin0773@163.com
[2]Bussiness school, Lingnan Normal University, Zhanjiang, Guangdong 524000, China,
jiangweiwei1371@163.com

Corresponding author: Weiwei Jiang, jiangweiwei1371@163.com

Abstract. This paper analyzes the application status of information retrieval and the technology used, and conducts in-depth research on ontology-oriented information retrieval technology. Aiming at the problem of low efficiency of information retrieval in the tourism field, a tourism information retrieval system based on domain ontology has been researched and implemented. The tourism information retrieval system based on the Internet of Things solves the problem of standardized description of domain knowledge and also solves the problem of semantic heterogeneity in network information sharing. Through the logical description and semantic reasoning of the set of conceptual relations generated by the abstraction of domain things, domain information can be effectively expressed at the semantic level. Therefore, the research of this thesis has laid a theoretical foundation for the further optimization of information retrieval technology.

Keywords: Internet of Things; Computer Aid; Travel Information
DOI: https://doi.org/10.14733/cadaps.2022.S6.124-135

1 INTRODUCTION

Modern society belongs to the era of the big explosion of information technology, and Web 2.0 has achieved rapid development. Tourists can publicize their location information and geographic tags, and can share tourist attractions that they have already experienced and communicate with friends. Compared with the traditional tourist forecast information, the tourist information forecast assisted by the computer system is a personalized forecast based on the combination of related information and location information. By optimizing tourism information through the tourism information forecast system, people can get better services from it. Moreover, in the optimization information of the tourism information forecast system, collaborative filtering technology is also introduced.

Considering that when tourists optimize tourism forecast information, they will be affected by geographic location and popularity. This article combines the location information of the CSN that tourists have visited to achieve a reasonable forecast of the range of tourist activities. The regions can be divided, and the probability level of visits by tourists in each region can be analyzed based on relevant information. The distance between the center of the region and the CSN and its popularity can be used to generate the probability level of the CSN in each region. Finally, by fully integrating the influence of regional popularity and combining with the existing CATIS forecasting system, a reasonable forecast of tourist information optimization problems can be realized.

Li [1] mentioned the World Wide Web invented by Tim Bemer S Lee marks the entry of mankind into the age of network information. With the rapid development of computer technology, the Internet and even the mobile Internet, Internet has long been known as the world's most important massive information tower. The diversification of information presentation and the swiftness of information release have promoted the prosperity of Internet information. At the same time, how to accurately retrieve useful information and knowledge from the massive information resources according to the user's query requirements is a problem that needs to be continuously studied in information retrieval today, as shown in Figure 1.

On January 19, 2011, CINIC released the 27th Statistical Report on Internet Development in China. According to data, as of December 2010, there were more than 340 million search engine users and 303 million mobile Internet users in China. Search engine users accounted for 81.9% of the total number of Internet users in China, a year-on-year increase of 33.1%. Search engines have become mainstream information acquisition methods. World-renowned search engines, including Baidu, Sogou, Google, Bing, and Yahoo, are constantly optimizing the information retrieval experience driven by emerging concepts and technologies such as cloud computing and box computing. Internet users rely on search engines in their lives and work. Yeal et al. [2] thought the degree is getting higher and higher. However, in many professional fields, some or even mainstream information retrieval systems use keyword matching technology, and the retrieval results they provide generate too much spam, so that the services and experience they provide cannot satisfy users.

In recent years, the integration of traditional industries and the Internet industry has brought explosive growth of information to traditional industries, while also bringing new and broad business opportunities to traditional industries. Filip and Andrea [3] mentioned the rapid development of the "tourism + Internet" model has also made it easier to release and obtain information in the tourism field. With the rise of a series of tourism information websites such as Ctrip, Qunar, Dianping, Ding Ding Life and China Tourism Information Network, vertical search, information, and retrieval in the tourism field have gradually become an important way for Internet users to obtain tourism information. But taking the travel information that users are most concerned about as an example, the information retrieval services provided by the numerous traffic map query and travel strategy query websites in Fangzhi travel field are not high in recall and intelligence. The current problems of the tourism field information query website are:

(1) Single query demand matching method: There are many tourism information query systems based on keyword matching technology to achieve information retrieval in the tourism field. These systems lack semantic analysis of search terms, cannot determine their conceptual similarity, and optimize the retrieval experience Caused great obstacles;

(2) The sorting method of information retrieval results needs to be optimized: there is a lack of a perfect result sorting mechanism, and the results are simply shifted or ranked by bidding. Too much spam and noisy data increase the difficulty for users to obtain available information;

(3) The cost of information integration is too high: retrieval intelligence is insufficient, and it is common for users to perform secondary retrieval by reconstructing keywords to obtain relevant information of retrieval results, resulting in a lot of time being spent on information integration of retrieval results. The concept association rules when the search results are presented need to be optimized.

Cheng and Jin [4] thought it can be seen that how to improve the shortcomings of traditional semantic retrieval methods has become a major problem facing the information retrieval technology research institute in a specific field. The proposal of ontology solved the problem of standardized description of domain knowledge, and also solved the problem of semantic heterogeneity in network information sharing, which directly led to the birth of the semantic web. The Internet of Things ontology technology can effectively express the domain information system at the semantic level through the logical description and semantic reasoning of the set of conceptual relations generated by the abstraction of domain things, laying a theoretical foundation for the further optimization of information retrieval technology. Therefore, information retrieval oriented to ontology technology has become an important research topic [6-8].

Figure 1: Schematic diagram of the internet of things.

Baltasar and Ortin [5] have made some achievements in applying ontological concepts to information science. Ontology theory is not only widely cited in the fields of information science such as semantic retrieval, semantic extraction, query expansion, intelligent navigation of digital libraries, and information integration, but also in traditional disciplines such as food safety, tourism emergency warning, network public opinion monitoring, and situational awareness. There are more and more applied researches in intersecting fields. Ontology description languages such as XML, RDF, OWL, and ontology reasoning engines such as Racer, Jess, Pellet, etc., provide standardized description standards and semantic reasoning rules for the research of ontology in the field of information science.

At present, well-known ontology-based information retrieval projects abroad include (onto)ZAgent, ontoseek and sKc, which represent three different areas of information retrieval: (Onto) recognizes that gent is based on network agents to search for its own ontology and each ontology on the Internet. The object reference tree established by conceptual metadata helps users to retrieve useful ontology information on the Internet. Ontoseek is a semantic information system used for searching yellow pages and product catalogs. It uses SENSUS to match user search strategies and actual data to retrieve web pages that contain useful information for users. sKc enables different ontologies in various semantically heterogeneous autonomous systems to achieve communication and interoperability through algebraic systems. (onto)ZAgent, ontoseek and sKe are major research results in ontology application and information retrieval. These three projects have their own advantages and disadvantages in retrieval efficiency. How to better utilize the advantages of ontology to improve information retrieval efficiency is also a follow-up study The focus of the system's attention [9,10].

2 QUERY SYSTEM REQUIREMENT ANALYSIS

The tourism information retrieval system based on domain ontology (noTIRS) designed and implemented in this paper, as shown in Figure 2, is an information retrieval system based on the domain ontology of Beijing tourism information. By improving the key technology of ontology semantic retrieval, explore to provide better tourism information retrieval services.The key technologies of semantic retrieval include semantic similarity calculation, query expansion, semantic expansion, and inverted indexing of domain documents. These technologies are closely related to the class hierarchy, conceptual relationships, and attributes of the domain ontology. Therefore, the domain ontology that has been constructed needs to be improved for a long time, and constantly supplemented with relevant examples and attributes of the domain, in order to continuously improve the effect of semantic retrieval.On the one hand, the system to be implemented in this paper has the function of maintaining, expanding and improving the domain ontology, and it needs to preprocess the domain documents and training documents; on the other hand, through improved key technology of semantic retrieval, it can provide recall and accuracy. Higher rates and more optimized information retrieval results for sorting schemes.

This thesis divides the use objects of the DOTIRS system into: ontology developers who maintain and expand the domain ontology and ordinary users who perform travel information retrieval, and analyze and organize the needs of the two types of users. After constructing the Beijing tourism information domain ontology according to the research content, a visual way is needed to facilitate the editing operations of the domain ontology concepts such as classes, relationships, and attributes, so as to continuously improve the domain ontology examples and the comprehensiveness of the structure. For ordinary users, they need to enter keywords to obtain query expansion results and semantic inference retrieval results processed by key semantic retrieval techniques.

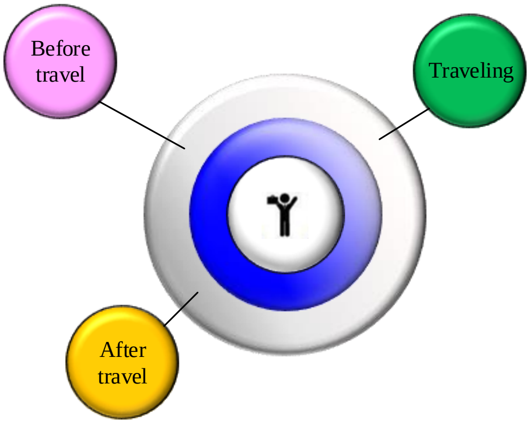

Figure 2: Tourism information system.

3 RETRIEVAL SYSTEM

According to the demand analysis of Beijing Tourism Information System, the system can be divided into three functional modules: ontology management, semantic retrieval, and query expansion. The

DOTIRS system adopts a three-tier architecture of data layer, logic layer, and presentation layer, and its system frame diagram is shown in Figure 3.

Presentation layer: the user interface layer, this layer mainly provides the page display function of the DOTIRS system, as well as the knowledge base management interface and the query expansion interface. Through the knowledge base management interface, users can add and delete instances and edit attributes of the ontology tree through visual operations. Through the query expansion interface, users pass in search keywords for semantic retrieval or query expansion retrieval.

Logical layer: This layer is the core business layer of the DOTIRS system, which consists of two modules: knowledge base management and query expansion. Semantic retrieval is an important function in the query expansion module. The DOTIRS system also regards semantic retrieval as a separate functional module. On the one hand, it provides users with semantic inference results and on the other hand provides semantic expansion search terms for query expansion.

Data layer: The data layer includes domain ontology datasets and index document collections. The domain ontology data set is composed of domain ontology database, domain ontology tree structure and domain reasoning rule documents.

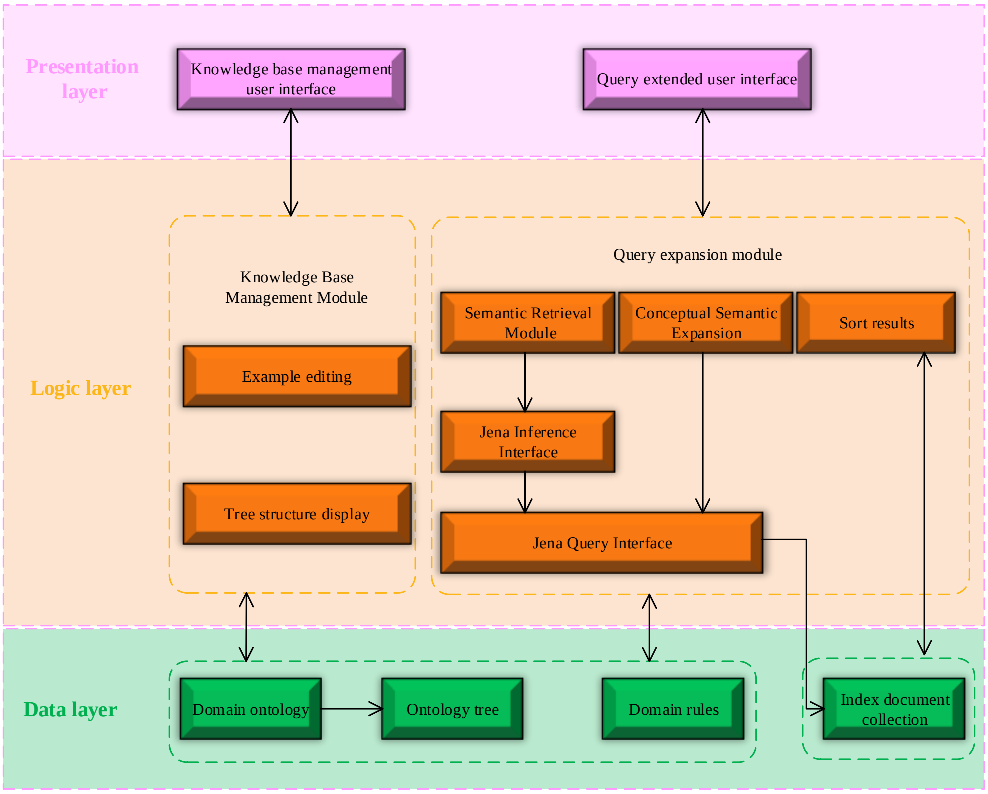

Figure 3: Frame diagram of DOTIRS system.

The knowledge base management module provides the visual operation of the domain ontology tree. The knowledge base management implemented by the DOTIRS system is used to visually display the concepts, attributes and instances of the "scenic spot" subcategories, and provide a manual editing interface.

The framework of the knowledge base management module is shown in Figure 4. This module is composed of ontology tree representation, ontology loading and instance editing sub-modules, and knowledge base management user interface. In the DoTIRs system, the functions of ontology tree representation, ontology loading and instance editing of the knowledge base management module are realized respectively through the class DisPlayO, Loadont, and Manageront.

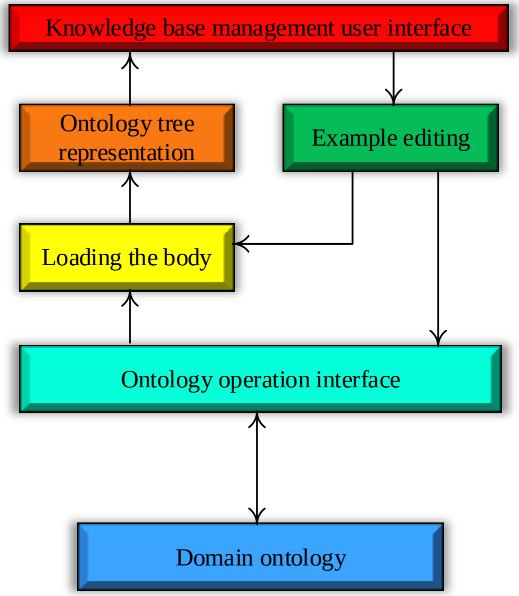

Figure 4: Structure diagram of knowledge base management module.

The class DisPlayont responds to the mouse click on the interface of the knowledge base management module, and displays the ontology tree node List, the instance LISt of the node and the instance attribute List generated by the class Loadont in real time.

The class Loadont and the class Manager operate on the owL file of the domain ontology through the extension package provided by the Jena inference engine. The extension package used in this module is mainly OntModel, Class, Modelspec, property, and property under com.shuai.hP!jena.ontofogy. Resouree et al. Apl.

The main functions and functions of the Loadont class are as follows:

(l) readowL function: Load the owl ontology file whose path is owlFileDir into the OntModel class through the OntModel.read(owlFileoir) interface.

(2) ListClass function: through the OntModel.listNamedClass() and OntModel.listsubClass() interfaces, the classes and subclasses in the domain are extracted hierarchically and encapsulated in the ontology tree structure.

(3) Listxns function: Obtain an instance of the class through the OntC-ass.listxnstance() interface.

(4) getelassPro function: get the properties (object properties) of the class through the ontClass.listDeclaredproperties() interface.

(5) getlnsPro function: the ontResource.listproperties() interface reads the RDF triples described in section 2.2.2, and extracts them through URI segmentation and regular expressions to obtain the data properties of the instance.

The main functions and functions of the Manageront class are as follows:

(1) addInsPro function: add the attributes of the instance. Obtain the instance that needs to add attributes, create a triple RDF node, and add attribute values to the individual individuals converted into the instance U.

(2) addClass function and deleteClass function: add and delete classes through the addsubClass and removesubClasS interfaces provided by ontClass;

(3) addins function: create RDFstatement, add instances in the form of triples.

(4) deleteeIns function: This function is used for instance deletion operations. Before deleting an instance, you need to clear the instance of the attribute.

(5) writeOWL function: write the modified ontology structure into the ontology OWL file, and call the Loadont class to refresh the ontology tree.

The constructed ontology in the field of tourism information in Beijing contains many semantic relationships such as scenic spots, geographic locations, traffic information, and service facilities. These relationships need to be reasoned by semantic rules to achieve semantic retrieval of tourism, transportation, and service information. D O T I R S system Je na inference engine realizes the semantic retrieval module of the system.

The function of the semantic retrieval module is divided into the following three steps: use the domain rules established by Jena to perform semantic inference on the parsed domain ontology based on OWL or RDF. The basic framework of this module is shown in Figure 5.

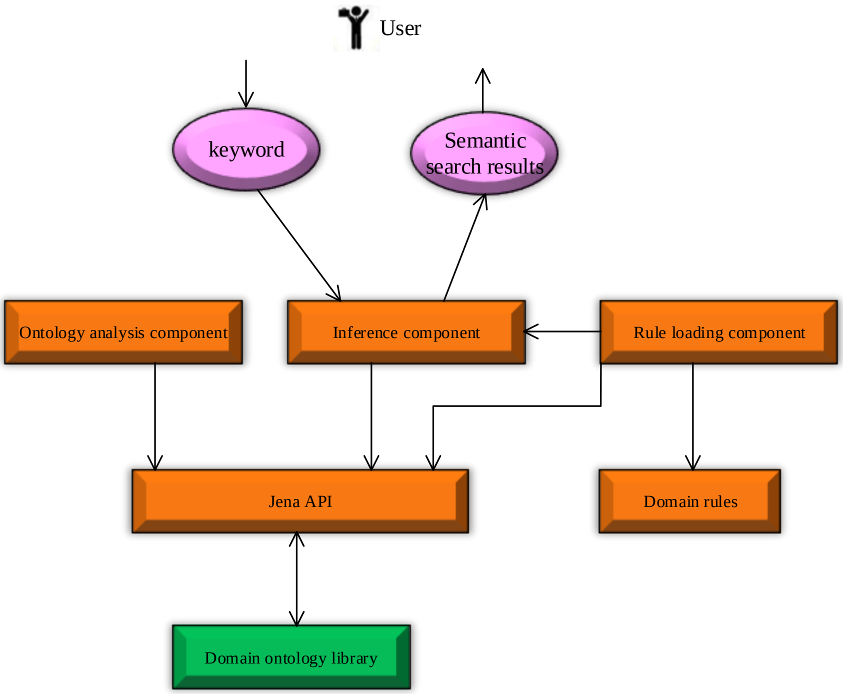

Figure 5: Framework diagram of the semantic retrieval module of the DOTIRS system.

The ontology analysis component, query component, reasoning component and rule loading component all operate on domain ontology through the API interface provided by Jena inference engine.

The ontology analysis component realizes the reading and analysis of the classes, relationships, and attributes in the domain ontology library, and can occasionally store the ontology data model in the memory into a relational database.

Ontology parsing component reads and parses domain ontology through JenaAPI, including operations on RDF language model and operations on OWL/RDFS ontology.

Analyze the relationship between the component and the domain ontology to construct an RDF language model, namely {resource, attribute, value} triplet, and the specific process is shown in Figure 6.

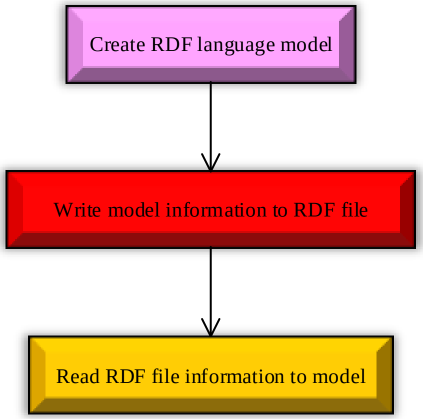

Figure 6: RDF language model construction flow chart.

The reasoning component can search the domain ontology for concepts and ontology level between concepts based on the general reasoning functions (parent class, whole and part, etc.) that Jena comes with and the domain rules parsed by the rule loading component.

The query component queries the Jena reasoning module through the query engine, returns the semantic search results to the user, and transmits them to the query expansion module at the same time, preparing expansion keywords for query expansion

Geographical location is the key information that needs to be considered in the tourism field. When the user enters the search keywords and the tourist attractions in the ontology, the DOTIRS system needs to provide the neighboring attractions, surrounding accommodation services, catering services and transportation facilities of the attractions. Based on the constructed ontology in the field of tourism information in Beijing, the following semantic inference rules are established using RDF rule extensions and Jena inference engine:

Rule 1: Scenic spot A is located at location P, and there is scenic spot B in location P, then scenic spot A and scenic spot B are in the same place. Rule 2: Scenic spot A is located at point P, and there is scenic spot B near point P, then scenic spot A and scenic spot B are neighboring scenic spots. Rule 3: Restaurant R belongs to scenic spot A, scenic spot A is located at location P, then restaurant R enters location P.

DoTIRs system implements query expansion function through objects of class QueryExPansion. This class declares objects of class search protege, Jenaontology, search Manager and Ranking to realize query expansion keyword generation sub-module and result retrieval and sorting sub-module in the query expansion module framework.

The function of the class search protege is to use the snRPeM semantic similarity calculation model to combine the keyword set together, perform semantic similarity calculation with the concepts in the domain ontology tree generated by the class TostarismTree, and return the concept similarity to expand keywords. The function of Jenaontology-like is to realize semantic inference, semantic extension keywords inferred from basic rules of domain ontology and custom rules.

Class IndexManager is used for indexing of training documents. The establishment of the index is completed by the three main member functions of this class:

The infoRetrieval function extracts the structure of a web page document through a regular method, and calls the MM Analyze: tokenizer provided by Lucence to segment the title and content to generate index items. The termPath function compares the word segmentation result with the ontology tree to generate a path set of terms. The path index function constructs an entity-based inverted path index structure. The search Manager class expands the keywords according to the incoming query, queries the search results, and passes the result document set to the object of the class Ranking.

The ranking object uses the ILSS sorting algorithm, and the result document set is sorted by semantic score, and finally the retrieval result of query expansion is obtained.

4 EXPERIMENT

The Foursquare site enables computer social networking, and visitors can travel anywhere on the network site. Collect and sort this tourism information to get information about time, text and other aspects. After this information is effectively sorted, it can effectively analyze and evaluate people's tourism information forecasts, and realize an optimized analysis. The Foursquare data set is selected in this experimental study. The time span selected is from September 2011 to November 2012. The travel records analyzed are tourists from the United States, Europe and other regions, and the data contained in the data set is very rich. , Including social information, as well as geographic location information, time, related content, etc. corresponding to travel. This paper conducts research based on 5,468 tourists in the Foursquare data set, and these tourists reasonably visit 7,286 OTIFS.

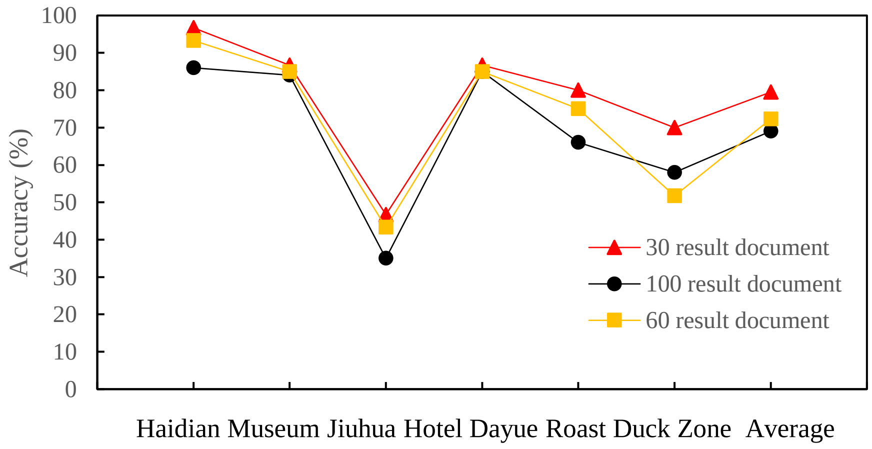

Figure 7: Precision of the first 30, 60 and 100 result documents.

A comparative experiment was carried out, and the optimization information of the tourism information forecast system was reasonably evaluated through the following two indicators, namely Precision@ k and recall rate Recall@ k.

In this paper, the five baseline methods corresponding to the CATIS forecast system are compared and analyzed, namely TSCP, GSCP, GTCP, GTSP, GTSC.

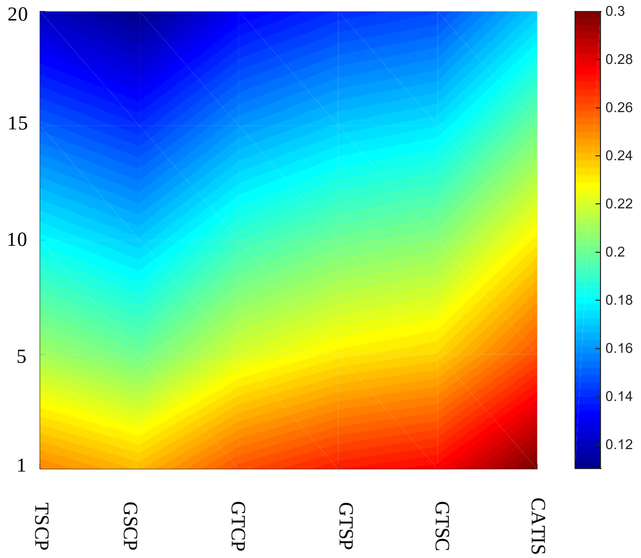

Figure 8: Different factors on the foursquare data set affect the accuracy of remote forecasting.

From the analysis of the comparison results of Figure 7, Figure 8, Figure 9 and Figure 10, it can be seen that, first of all, it can be concluded that the CATIS designed in this paper has better performance no matter it is in a remote or local scene. It can also be concluded that in different scenarios, the influence of various factors on the accuracy of the forecasting system is different.

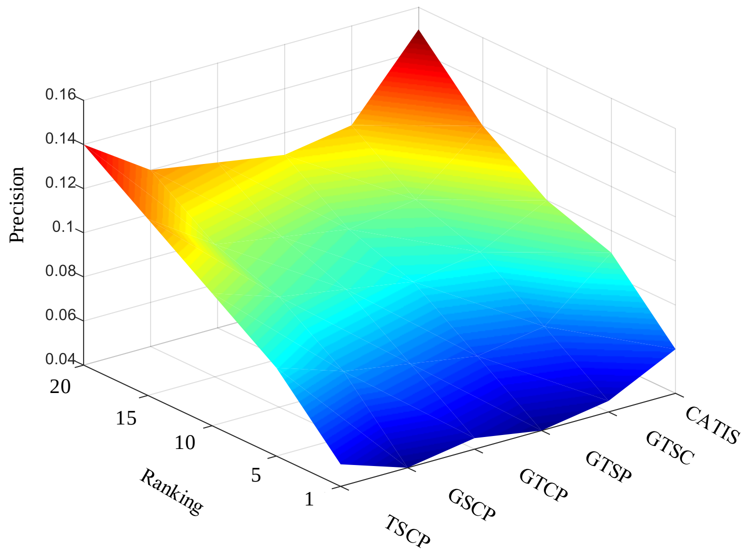

Figure 9: Different factors on the Foursquare data set affect the local forecast accuracy.

It can be seen from Figure 7 that if it is an off-site tourism information forecast, it can be ranked according to the importance of factors. The tourism content has the greatest impact, followed by popularity, followed by social relevance and time effects, and the least impact is geographic location.

However, it can be seen from Figure 8 that if it is a local forecast scenario, the ranking will be changed at this time. The importance of tourism content will be the lowest, and the highest is the time effect, followed by geography and popularity. Wait. After comparative analysis, it can be concluded that if it is a remote scene, the problem of data sparseness can be effectively overcome through content information; but in a local scene, the time effect is more prominent.

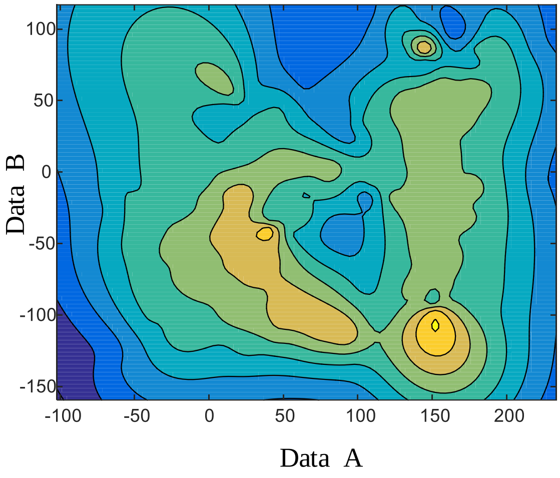

Figure 10: Data comparison.

5 CONCLUSION

The paper analyzes the application status of information retrieval and the technology used, and conducts in-depth research on ontology-oriented information retrieval technology. Aiming at the problem of low efficiency of information retrieval in the tourism field, a tourism information retrieval system based on domain ontology has been researched and implemented. The tourism information retrieval system based on the Internet of Things solves the problem of standardized description of domain knowledge, and also solves the problem of semantic heterogeneity in network information sharing. Through the logical description and semantic reasoning of the set of conceptual relations generated by the abstraction of domain things, domain information can be effectively expressed at the semantic level. Therefore, the research of this thesis has laid a theoretical foundation for the further optimization of information retrieval technology. In this paper, a CATIS forecast system is constructed under the computer system-assisted environment. In the forecast system, information such as time, geography, and content are effectively integrated. By considering these factors, problems such as data sparseness and interest drift are solved. Very good solution. Combined with the real data set, the experimental analysis was carried out. Through comparative experiments, it is verified that the proposed CATIS forecast system has higher accuracy than other methods.

Bin Liu, https://orcid.org/0000-0002-9600-9071
Weiwei Jiang, https://orcid.org/0000-0002-0953-4485

REFERENCES

[1] Li, S.: Trends, Development and Issues in Tourism Information System, The Frontiers of Society, Science and Technology, 14(2), 2020, 1-9. http://doi.org/10.25236/FSST.2020.021402

[2] Yael, R.; Ayelet, G.; Amit, R.: Identifying attributes of public transport services for urban tourists: A data-mining method, Journal of Transport Geography, 93, 2021, 134-144. https://doi.org/10.1016/J.JTRANGEO.2021.103069

[3] Filip, E.; Andrea, H.: Big Data: a Source of Mobility Behaviour and a Strategic Tool for Destination Management, Czech Journal of Tourism, 8(2), 2019, 85-102. https://doi.org/10.2478/CJOT-2019-0006

[4] Cheng, M.; Jin, X.: What do Airbnb users care about? An analysis of online review comments, International Journal of Hospitality Management, 76, 2019, 58-70. https://doi.org/10.1016/j.ijhm.2018.04.004

[5] Baltasar, G.; Ortin, F.: A didactic object-oriented, prototype-based visual programming environment, Science of Computer Programming, 176, 2019, 1-13. https://doi.org/10.1016/j.scico.2019.02.004

[6] Wenjuan, L.: The Integration of Contemporary Art Visual Elements in Visual Communication Design, Journal of Frontiers in Art Research, 1(3), 2021, 4-7. https://doi.org/10.23977/jfar.2021.010302

[7] Fan, M.; Li, Y.: The application of computer graphics processing in visual communication design, Journal of Intelligent & Fuzzy Systems, 39(4), 2020, 5183-5191. https://doi.org/10.3233/JIFS-189003

[8] Lin, Y.; Hu, Y.; Chen, W.; Biswaranjan A.: Reversible self-verifying and self-recovering technique for color image demosaicking, Multimedia Tools and Applications, 2021, 1-20. http://doi.org/10.1007/S11042-021-10914-Z

[9] Gilbert, T.: Looking at Digital Art: Towards a Visual Methodology for Digital Sociology, The American Sociologist, 49(4), 2018, 569-579. https://doi.org/10.1007/s12108-018-9384-2

[10] Le, D.; Biswaranjan, A.; Ajaya, K.; Jyotir, M.; Raghvendra, K.: NoSQL Database Classification: New Era of Databases for Big Data, International Journal of Knowledge - Based Organizations, 9(1), 2019, 50-65. https://doi.org/10.4018/IJKBO.2019010105