Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2306321, 11 pages
https://doi.org/10.1155/2022/2306321

Research Article

English Translation Template Retrieval Based on Semantic Distance Ontology Knowledge Recognition Algorithm

Yu Chen

Zhejiang Yuexiu University, Shaoxing 312000, Zhejiang, China

Correspondence should be addressed to Yu Chen; 20082028@zyufl.edu.cn

Received 30 March 2022; Accepted 30 April 2022; Published 31 May 2022

Academic Editor: Xiantao Jiang

With the rapid development of the Internet, data information has begun to spread rapidly on the Internet, and the types and fields involved are becoming more and more diverse. With the deepening of internationalization, English information has also become a common entry for tourism. Relatively speaking, the translation of English documents is also on the agenda. The intricate literature information has put forward a more specialized and vertical demand for retrieval. Traditional retrieval systems usually use string matching algorithms to search, but this technology cannot perform semantic expansion of query conditions. In the case of "multiple words with one meaning" and "polysemy of one word," the retrieval accuracy of traditional retrieval systems decreases. In order to solve the previously mentioned problems, this study proposes semantic retrieval based on domain ontology technology. The purpose of semantic retrieval is to semantically expand the user's query conditions on the basis of accurately understanding the retrieval conditions. At the same time, with the help of the constructed domain ontology, it provides more accurate and comprehensive retrieval services for document retrieval. The experimental results of this study show that, for the clustering of the entire dataset, the number of ontology concepts in each article is 61% of that, before merging, which effectively reduces the complexity of the semantic network model. Therefore, the validity of the ontology concept merging algorithm proposed in this study is verified.

1. Introduction

The scale of online resources continues to expand. On the one hand, it creates convenient conditions for people to acquire knowledge and is no longer limited by factors such as time and place. On the other hand, it also makes it more difficult for people to acquire the knowledge they need accurately [1, 2]. The traditional retrieval adopts the string matching technology based on query conditions, ignoring the semantic relationship that the query conditions may contain. Sometimes it is not possible to accurately understand the user's query intent [3, 4]. When users conduct queries, they always hope to obtain as much information as possible related to their query conditions. However, due to the lack of domain knowledge, sometimes users cannot fully describe the content they want to query [5, 6]. At this time, it is hoped that the search engine can automatically perform appropriate query expansion according to some of the input retrieval conditions. This paper applies ontology technology to semantic retrieval [7, 8]. Through experiments, it can be seen that the semantic retrieval system based on ontology has significantly improved the recall rate and accuracy rate. At present, the more unified view of ontology concept is to explain the explicit formal specification of the shared concept model. However, because the construction of ontology is a time-consuming and difficult task, this paper will analyze the shortcomings in the construction process of ontology and the realization of the system [9, 10].

Ontology clearly defines concepts and the relationships between concepts, so that they can be accepted and recognized by everyone in a certain field. Compared with traditional information retrieval, ontology-based semantic retrieval can avoid the problem of low retrieval performance caused by lack of professional knowledge and inaccurate retrieval conditions. In addition, semantic retrieval can also sort the output according to the similarity between the

retrieval conditions and the retrieved information. In summary, semantic retrieval can make query results more accurate and comprehensive and more in line with the professional query needs of semantic retrieval.

The current information retrieval system generally adopts the string matching algorithm based on query conditions. However, the algorithm ignores the implicit information characteristics of the query conditions and cannot return the retrieval results accurately and comprehensively. Under this search algorithm, the text will be successfully retrieved if and only if the query information completely contains the query conditions. In addition, because of the inability to understand the semantic relationship of the retrieval conditions, it is also impossible to effectively retrieve "polysemous words" and "polysemy." Therefore, depending on the current retrieval system, it is often impossible to accurately query the required literature information. Based on the above problems, this paper proposes realizing the semantic retrieval of documents based on domain ontology technology and implements a semantic retrieval system based on computer-related documents as a model.

2. Related Works

With the introduction of ontology concept, researchers have done a lot of related research on ontology concept. Among them, Sang Q proposed defining a subgraph consisting of points and their neighbors. The edges of the two subgraphs are arranged according to length and angle to obtain the corresponding relationship, and, finally, the local support probability based on the corresponding relationship is calculated. The performance of the proposed method was validated on different synthetic and real data, showing that the proposed method can improve the robustness and accuracy of traditional techniques. Stylianopoulos proposed a new probabilistic clustering algorithm suitable for identifying linear elements in datasets containing linear clusters. The algorithm is an expectation-maximization-like procedure applied to mixed probability density functions. Each function models a line segment. Experimental evaluations show that the proposed method is relatively good or better than related state-of-the-art cluster-based methods and traditional line detection methods. Zhang analyzed the concepts of machined surface and machined features and proposed a new semantic method for automatic recognition of machined features. Semantic methods provide an ontology-based conceptual model to represent machined surfaces and machined features. Furthermore, an automatic feature recognition method based on semantic query and reasoning was proposed. Case studies show that the proposed method can effectively identify and interpret interaction features with good openness and scalability. Ramirez-Amaro proposed a framework that uses semantic representations to infer human activities from observations. The proposed framework can be used to address difficult and challenging problems of transferring tasks and skills to humanoid robots. The results show that the built system correctly identified human behavior in about 87.44% of the cases in real time. This is even better than random participants recognizing another person's behavior. Maksimov considered ontology methods for semantic recognition and representation of text documents relevant to interactive information retrieval problems. Interactive tools are presented. The tool uses the operation of constructing aspect projections for the graphical representation of the ontology. This makes it possible to reduce the dimensionality of the graph to an acceptable level from a display and perception point of view. Most of the existing efficient and reliable ciphertext search solutions are based on keyword or shallow semantic parsing, which are not intelligent enough to satisfy users' search intentions. Therefore, Fu proposes a content-aware search scheme that can make semantic search more intelligent. Experimental results show that the proposed scheme is effective. Sun proposed a real-time fusion semantic segmentation network called RFNet. It effectively utilizes complementary cross-modal information. A comprehensive set of experiments demonstrate the effectiveness of the proposed framework. On Cityscapes, their method outperforms previous state-of-the-art semantic segmenters. It has excellent accuracy and 22 Hz inference speed at full 2048×1024 resolution, outperforming most of the existing RGB-D networks. There is still a mismatch between text-based and knowledge-based retrieval methods. The former does not take into account complex relationships, while the latter does not properly support keyword-based query and ranking retrieval. Therefore, Devezas studied text-based methods and how they evolved to exploit entities and their relationships in the retrieval process. Handa has proposed an efficient method to perform searches on encrypted data using clustering. Experimental results using real datasets show that, compared with the state of the art, the proposed multikeyword ranking search scheme on encrypted cloud data significantly reduces the number of comparisons and search time, while maintaining 100% recall and 82% accuracy. How to discover the most common neighbor words from huge stream data is very interesting. Chen developed a novel and effective mechanism to solve this problem, including a quadtree-based index structure, an index update technique, and a best-first-based search algorithm. An empirical study shows that the proposed technique is effective and meets user requirements by varying many parameters. Most of the above scholars' research on information retrieval is based on keyword matching, and the retrieval results are easily affected by user input bias.

3. Retrieval Method Based on Ontology Knowledge Recognition Algorithm

Conceptual Principles of Ontology. Regarding the study of ontology, the concept of ontology is developed from the field of philosophy. It is a philosophical question that studies the nature of existence [11]. In recent decades, ontology has become a focus of attention. After ontology was introduced into the computer field, it has been widely used in artificial intelligence, computer network programming language, information retrieval, and other fields [12]. Ontology is

playing an increasingly important role in research and development in various fields. However, so far, the computer field has not formed a unified view on the definition of ontology.

Ontology has applications in many aspects of the library and information field, such as information system modeling, information extraction, semantic web, search engine, knowledge management, knowledge base construction, and library information resource construction. The construction process of ontology can be expressed as

$$\text{ontology} = \text{concept} + \text{Attributes} + \text{kilometer} + \text{value} + \text{name}. \quad (1)$$

3.1. Functions and Features of Ontology

Function of the Body. According to different ontology classification rules and different application fields of ontology, ontology also has many different functions and characteristics. But no matter what category the ontology belongs to or what field it is applied to, the basic main functions of the ontology are basically the same. According to the literature reviewed, it can be concluded that the main basic functions provided by ontology include the following: it provides a roadmap for a field or the relationship between fields, and an ontology provides a clear formal specification for a concept. Ontology concept definitions are correlated in different domains [13].

The basic function of ontology is to provide a common vocabulary. It can be used to describe the vocabulary required by the target world and recognized by the majority of users [14]. Ontology can provide the basis of semantic interoperability, so it can realize the effect of semantic interoperability. Ontologies make implicit concepts explicit. Since ontology is a formal structure, by using the rules of this structure, implicit concepts and relationships between concepts can be explored. In addition, the use of high-level concepts in ontology helps to understand low-level concepts; that is to say, ontology systematically describes the concepts in the domain. For systematic knowledge, the organizational structure of various knowledge concepts in a good knowledge system must be systematic. The use of ontology can effectively describe these knowledge concepts systematically.

3.2. Semantic Search Engines. The massive resources of the Internet have caused a serious information overload. As far as ordinary users are concerned, how to filter irrelevant information and find a satisfactory answer is a difficult problem. The emergence of keyword-based search engines partially solved this problem. However, the search return set contains a lot of information: one is inaccurate and may not contain the required web page, and the other is that the user still needs to manually identify and select the correct result. To this end, semantic search appears, based on the ontology knowledge base, to correctly understand the semantics of documents, obtain the deep meaning of documents, and return query results more accurately. In layman's terms, semantic search unites the Semantic Web. Through ontology knowledge base reasoning, combined with traditional keyword-based search, a result set is finally formed [15]. Among them, the reasoning part of Ontology, the result is definite. Because it uses formal reasoning, such as the Bool model, the traditional search part is the returned set obtained after sorting. Table 1 compares traditional search engines and semantic-based search engines. By comparison, it can be concluded that semantic search is more suitable for language specifications and can understand deeper meanings.

The overall architecture of semantic search is shown in Figure 1. It is mainly divided into three parts. The first part is the establishment of domain ontology knowledge base, the second part is ontology reasoning and traditional search engine retrieval, and the third part is server and user interface.

3.3. Ontology Knowledge-Based Construction Algorithm

3.3.1. Traditional Ontology Construction Method. In the actual construction process of ontology, due to factors such as knowledge characteristics, ontology scale, and the actual use of building ontology, there is no unified standard specification for ontology design and development. Each ontology builder will choose and adjust according to his own project needs. At present, the more mature and commonly used construction methods mainly include TOVE method, IDEF-5 method, skeleton method, prototype evolution method, and seven-step method.

(1) IDEF-5. The development process of this method to build ontology is shown in Figure 2.

(2) TOVE method. The development process of this method to build ontology is shown in Figure 3. TOVE refers to the Toronto Virtual Enterprise especially used to build an ontology about the enterprise modeling process.

(3) Prototype evolution method. This method completes the construction of ontology through six steps: First, analyze the needs of the constructed ontology. Second, plan the ontology construction process. Third, obtain ontology information. Fourth, determine the ontology concept and relationship. Fifth, formalize the ontology. Sixth, evaluate the ontology. The established ontology is evaluated.

(4) Seven-step method. This method is a relatively common ontology construction method, which completes the ontology construction through seven steps. First, determine the application purpose and direction of building an ontology. Second, consider reusing existing ontologies. Third, extract the proper nouns of the domain. Fourth, define the subordinate structure of the class. Fifth, define the data attributes of the concept itself and the object attributes between the concepts. Sixth, define the limitations of attributes. Seventh, build the instance.

Table 1: Comparison of semantic search and traditional search engines.

Engine tips	Traditional search	Semantic search
Polysemy	Based on keywords	Problem-based
Synonym	Not support	Support
Particle	Cannot understand	Understand
Phrase	Particles such as "of" and "is" are not considered	Consider particle
Long tail query	Do not understand	Understand
Engine tips	Cannot handle	Can handle

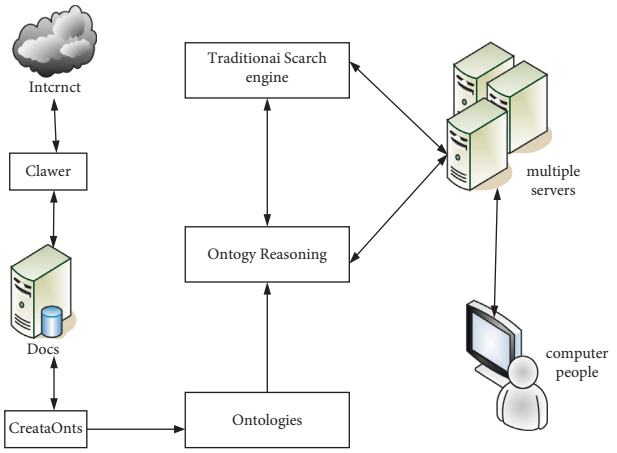

Figure 1: Architecture of semantic search.

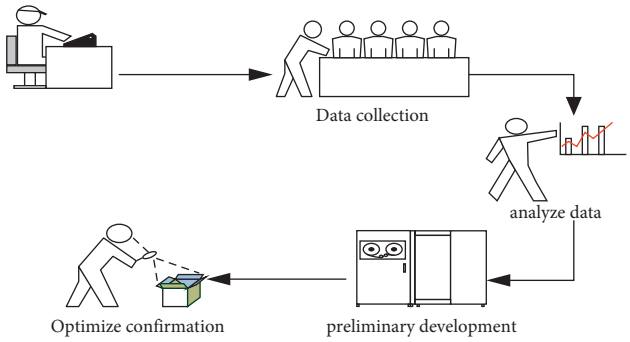

Figure 2: IDEF-5 ontology development process.

3.3.2. Improved Construction Method. It can be seen from Figure 4 that the main steps of ontology construction are the same, but the specific implementation varies from person to another. After analyzing and summarizing the existing methods, it is found that the construction of ontology has something in common with object-oriented development. Therefore, this paper combines the "object-oriented" idea with the traditional "seven-step construction method" and proposes a new construction method. This approach treats each ontology concept as an abstract class. The methods in each abstract class correspond to the relationship of the ontology, and the constants of each class correspond to the data attributes of the ontology. This paper relies on object-oriented method to construct document ontology and semantic dictionary ontology in computer field. Compared with

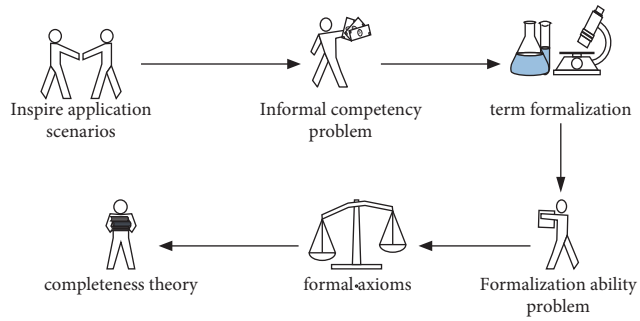

Figure 3: The ontology development process of TOVE.

the traditional construction method, this method is simpler and more convenient to implement. At the same time, because the concept in object-oriented method has a good relationship between the superior and the subordinate, the relationship extracted by this method is more complete and rich. Therefore, the ontology constructed based on this method is of higher quality, which in turn can improve the retrieval performance in the computer field to a certain extent. The specific development process is shown in Figure 5.

4. Design and Implementation of Ontology-Based Semantic Retrieval System

By combining ontology technology and search technology, this paper uses ontology concepts with semantic relations to express query conditions input by users. When a user conducts a query, the search content is not directly matched and retrieved through query conditions in a traditional way. Instead, it first determines whether the query conditions correspond to the concepts in the ontology, and, if not, it directly performs a matching query. Otherwise, according to the created concept semantic dictionary ontology, first, the synonyms, Chinese-English contrast words, and hyponyms of the query conditions are extracted. Then, together with the original input conditions, new retrieval conditions are formed, and then the new retrieval conditions are used to query and retrieve documents. In order to verify the proposed theory, this paper designs and implements an ontology-based document semantic retrieval system, which provides a semantic search function for document query in the computer field.

Implementation of Semantic Retrieval System. The retrieval system initially implemented in this paper mainly includes

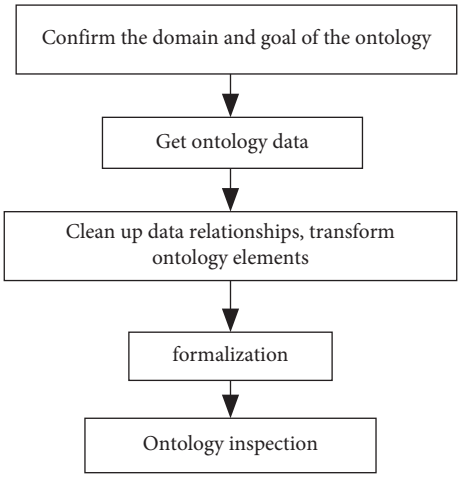

Figure 4: Ontology construction requirements.

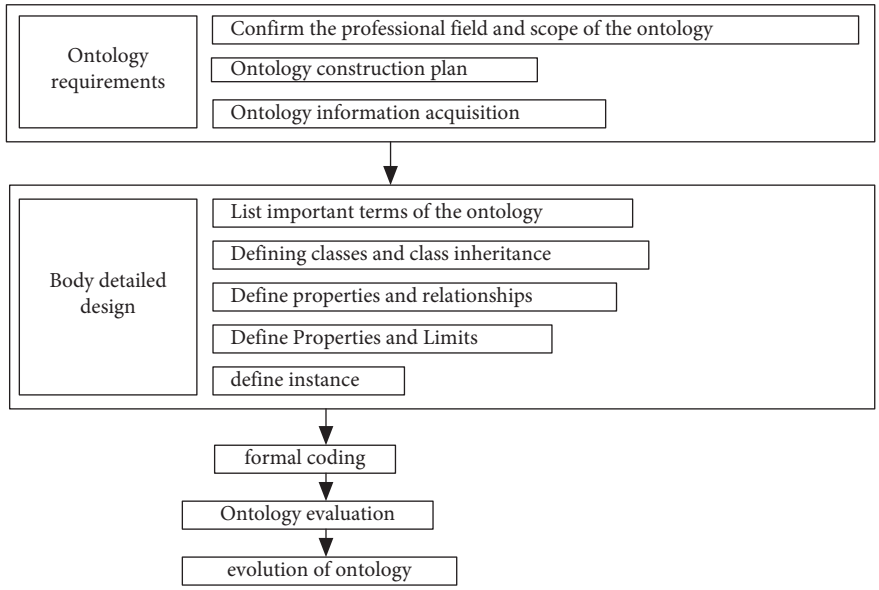

Figure 5: The construction process of domain ontology.

two steps: building an index and retrieving according to query conditions. The first is to refer to authoritative materials to construct domain ontology and collect original data. At the same time, the collected data is marked and indexed according to the constructed domain ontology. Then, the query conditions input by the user are received, and appropriate expansion is carried out with the help of domain ontology. Finally, the expanded query conditions are used to match and compare the resources in the index library, and the query results are returned. The overall framework is shown in Figure 6.

4.1. Similarity Calculation Based on Document Domain Ontology. Word length is an important indicator of the difficulty of studying texts. It is well known that since language is ambiguous and obscure, the distribution in the examination of word length is very helpful for us to discern its lexical characteristics. Those words that are 2–5 words in length are considered "small words." The smaller the text, the easier the words. Conversely, the more "large" (long words) the text is, the more difficult it is to contain. The distribution also affects the formal hierarchy of the text. If a text uses many small words, it tends to be more informal; vice versa, if a text uses many large words, it will be more formal. Figure 7 shows the distribution of word lengths in the corpus.

From the data in Figure 7, it can be seen that the average word length of the traditional knowledge base in Figure 7(a) is higher than the number of word lengths in Figure 7(b). If it is broken down, it can be seen that these differences are mainly from 1-letter and 9-letter words.

When using ontology to realize information retrieval, the query conditions will be expanded. In this way, the

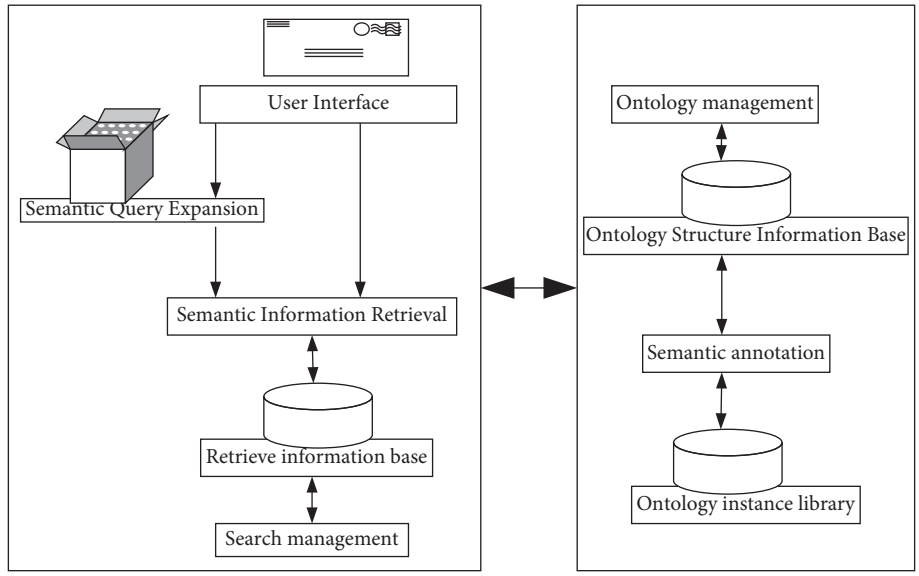

Figure 6: Semantic retrieval model framework.

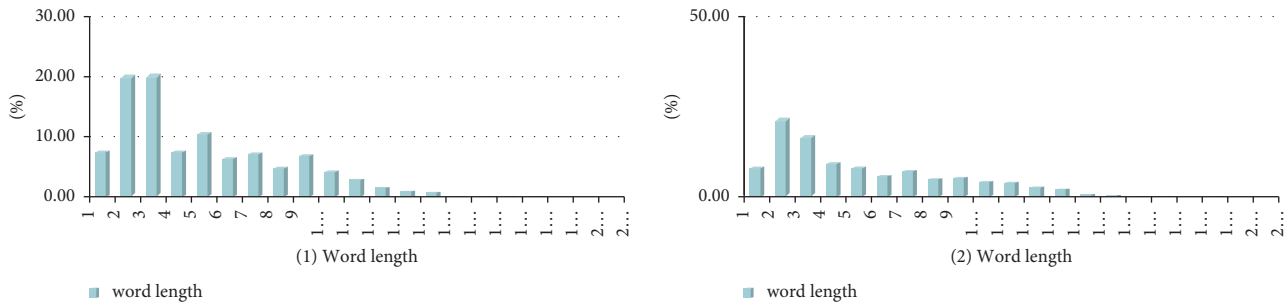

Figure 7: Comparison of word lengths in the knowledge base.

problem of decreasing accuracy and recall caused by "polysemous words" can be avoided. However, due to the expansion of the query conditions, it may result in too much information retrieved. Some documents are closely related to the query conditions, and some may be partially related to the query conditions. In order to solve the above problems, the queried documents should be sorted during the retrieval output, and the documents with a higher degree of relevance should be output first. Because this article implements semantic retrieval based on ontology, the similarity between ontology concepts and query conditions is calculated as the weight value of documents when querying. This paper mainly considers synonyms, hypernyms, and hyponyms when performing query expansion. Therefore, based on the above three basic structural relationships, the similarity between semantics is defined as follows:

$$\mathrm{Sim}(A_1, B_2) = \left\{ \begin{array}{c} 1; \text{synonym relationship} \\ 1: \text{English} - \text{Chinese translation relationship} \\ 0.5; \text{direct subordinate relationship} \end{array} \right\}. \tag{2}$$

In the above expression, A_1 represents the query condition, and B2 represents the concept in the ontology. For nondirect superior and inferior relationships, the semantic distance between the extension concept and the query concept needs to be considered. The closer the distance is, the higher the weight value should be. Therefore, when it does not belong to the above three basic structural relationships, the similarity between semantics is defined as follows:

$$\begin{aligned} \mathrm{Sim}(A_1, B_2) &= \frac{\mathrm{dep}(A_1) + \mathrm{dep}(B_2)}{M \times N}, \\ M &= \mathrm{Dist}(A_1, B_2) \cdot 2 \cdot \mathrm{Max}_{\mathrm{dep}}, \\ N &= \mathrm{Max}\left(\left|\mathrm{dep}(A_1) - \mathrm{dep}(B_2)\right|, 1\right). \end{aligned} \tag{3}$$

In the above formula, $\mathrm{dep}(A_1)$ and $\mathrm{dep}(B_1)$, respectively, represent the hierarchical depth of the query condition and the expanded concept word ontology. $\mathrm{Dist}(A_1, B_2)$ represents the shortest path in the ontology hierarchy tree. $\mathrm{Max}_{\mathrm{dep}}$ represents the maximum depth in it.

Since this paper adopts ontology-based semantic information retrieval, the similarity between query conditions and extension conditions is used as a measure of web page weight. Because, during query expansion, multiple query terms may be expanded for the same query condition, the final weight formula of a certain document is as follows:

$$F_{\text{core}} \notin \text{Sim}(A_1, S_n) + \text{Sim}(B_2, S_n) + \ldots \text{Sim}. \tag{4}$$

In the above formula, A_1 is the representation of the query condition, S_n is the extended concept in the ontology, and F_{core} is the final score of the text knowledge. Figure 8 is a flow chart of concept similarity calculation.

Using the page score calculation process shown in Figure 8, the final output order of each document can be obtained, as shown in Table 2.

4.2. Improved Semantic Expansion Algorithm. The semantic expansion algorithm based on word vectors is to obtain query expansion words through semantic matching and obtain query vectors and document vectors through association matching. The improvement direction and key technology of this algorithm are introduced below.

4.2.1. Query Expansion Algorithm Improvement. The acquisition method of semantic expansion word set is divided into two steps: The first step is to obtain query expansion words based on Word2vec and automatic threshold screening method. Word2vec is a group of related models used to generate word vectors. These models are shallow, two-layer neural networks that are trained to reconstruct linguistic word texts, and the second step is to introduce improved LSF technology on the basis of the first step to further filter query expansion words. Among them, LSF is a tool for distributed resource management, which is used to schedule, monitor, and analyze the load of networked computers.

The Word2vec algorithm mainly includes two models: Skip-gram and Continuous Bag of Words (CBOW). The model composition of CBOW is divided into three layers.

Input layer is as follows:

$$U(\text{Context}(\alpha)1), U(\text{Context}(\alpha)2), \ldots, U(\text{Context}(\alpha)2\ D) \in K^x. \tag{5}$$

Among them, the word vector of the 2D words in the input layer is Context(α), U is the word vector, and x is the length of the word vector.

Projection layer, summation, and accumulation are performed on all word vectors of the input to obtain a new projection layer vector. The calculation formula is as follows:

$$T_\alpha = \sum_{n=1}^{2D} U(\text{Context}(\alpha)n) \in K^x. \tag{6}$$

In the output layer, corresponding to a Huffman tree, a Huffman tree, also known as an optimal binary tree, refers to a binary tree with the shortest weighted path length constructed for a set of leaf nodes with certain weights. Corresponding to a Huffman tree, the words appearing in the training samples are used as leaf nodes, and the word frequencies in the training samples are used as the weights to construct the Huffman tree.

The network structure of the Skip-gram model, like the CBOW model, also includes an input layer, a projection layer, and an output layer. The Skip-gram model is built on the premise that the current word α is known, the output layer of its context is predicted, and the construction of the conditional probability function Q(Context$(\alpha)/\alpha$) is defined as

$$Q\left(\frac{\text{Context}(\alpha)}{\alpha}\right) = \Pi K \in \text{Context}(\alpha) Q\left(\frac{U}{\alpha}\right). \tag{7}$$

According to softmax, the $Q(U/\alpha)$ probability formula is

$$Q\left(\frac{U}{\alpha}\right) = \prod_{i-2}^{\alpha} Q\left(\frac{b_i^u}{U(\alpha)}\right), \quad \beta_{i-1}^x. \tag{8}$$

$Q(b_i^u/U(\alpha), \beta_{i-1}^x$ is shown in the following formula:

$$Q\left(\frac{b_i^u}{U(\alpha)}, \beta_{i-1}^x\right) = \left[\chi(U(\alpha)^a \beta_{i-1}^x\right]^{1-b_i^u} \cdot \left[1 - \chi(U(\alpha)^a \beta_{i-1}^x\right]^{b_i^u}. \tag{9}$$

The log-likelihood function of the objective function of Skip-gram is

$$H = \sum_{\alpha \in D} \log Q\left(\text{Context}\left(\frac{\alpha}{\alpha}\right)\right). \tag{10}$$

The stochastic gradient method is used to maximize $H(\alpha, u, i)$, and the gradient formula for β_{i-1}^u is

$$\frac{\delta H(\alpha, u, i)}{\delta \beta_{i-1}^u} = \left[1 - b_i^u - \chi(U(\alpha)^a \beta_{i-1}^u)\right] U(\alpha). \tag{11}$$

The update formula of β_{i-1}^u is

$$\beta_{i-1}^u = \beta_{i-1}^u + \theta\left[1 - b_i^u - \delta(U(\alpha)^a \beta_{i-1}^u)\right] U(\alpha). \tag{12}$$

In the above formula, θ represents the learning rate.

Due to symmetry, the gradient formula of $H(\alpha, u, i)$ with respect to $U(\alpha)$ is

$$\frac{\delta H(\alpha, u, i)}{\delta \beta_{i-1}^u} = \left[1 - b_i^u - \chi(U(\alpha)^a \beta_{i-1}^u)\right] \beta_{i-1}^u. \tag{13}$$

The update formula of $U(\alpha)$ can be expressed as

$$U(\alpha) = U(\alpha) + \theta\left[1 - b_i^u - \delta(U(\alpha)^a \beta_{i-1}^u)\right] \beta_{i-1}^u. \tag{14}$$

Algorithm Comparison. The evaluation standard of the experiment is an important tool to measure the quality of the experimental results. Therefore, some popular evaluation indicators are selected to measure the advantages and disadvantages of the algorithm.

(1) Clustering Error and Evaluation. This paper uses ontology topic concept clustering algorithm. For the clustering algorithm, two indicators of error sum and time complexity are selected for measurement. Obviously, it is difficult to control to achieve error and minimum at the same time, and

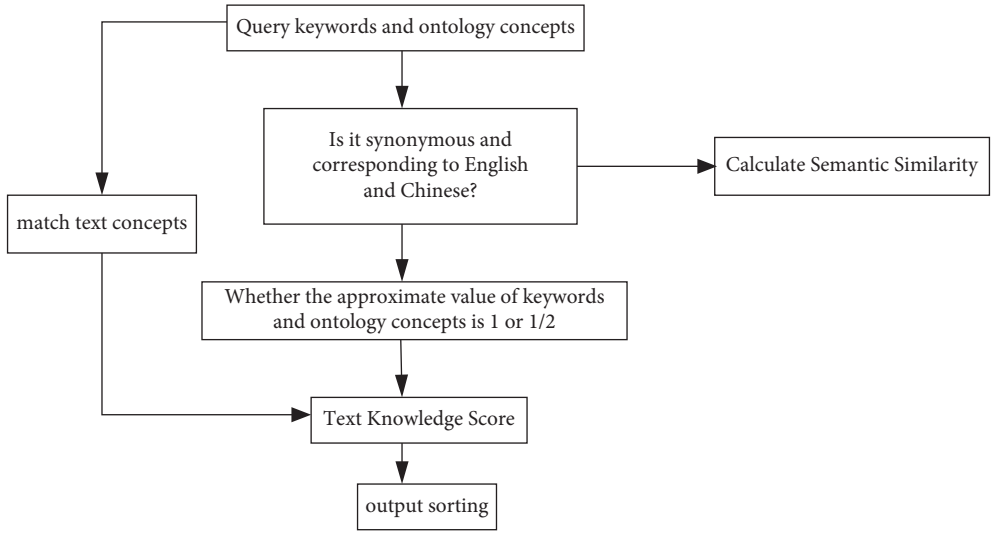

Figure 8: Schematic diagram of conceptual similarity calculation process.

Table 2: Schematic representation of scoring results.

Literature number	Search condition	Extended concept	Score	Sort
1	Hide information	Hide writing, data, watermark	2	1
2	Hide information	Digital watermark	0.5	5
3	Hide information	Hide information, digital watermark	1.5	2
4	Hide information	Hide information	1	3
5	Hide information	Digital, audio	0.65	4

the time complexity is the lowest, usually a compromise between the two. Since machine learning is used for clustering and manual retrieval is used for identification, the error sum can be appropriately relaxed to reduce the time complexity. Therefore, the clustering of multiple texts, the number of texts, will affect the time of the clustering algorithm to a large extent.

The main point of clustering evaluation is to carry out the criterion L and the square of the error, where n represents the number of clusters and I represents the sample point. The sample set in the a-th cluster is denoted as R_a, and N_a is the center point of the a-th cluster sample.

$$L = \sum_{a-1}^{n} \sum_{j \in R} \left\| I - N_a \right\|^2. \tag{15}$$

(2) Time Complexity (Average Running Time). From the point of view of the design of the algorithm, it is difficult to directly describe the time complexity with mathematical notation. To this end, the algorithm's start time and end time and the method of measuring multiple sets of data are used to determine it and compare it with other mainstream algorithms in terms of time complexity.

At present, the clustering algorithms with unknown number of categories mainly include the adaptive sample construction method (AdaMethod) and the density-based spatial clustering algorithm DBSCAN. By comparing the time complexity with the objective function, since the purpose of establishing the ontology is to expand the search performance, the search system will rescore and sort the search results, so the objective function can be appropriately relaxed. For a large amount of data, it is necessary to increase the time complexity. Because a compromise is made under the optimal time complexity and objective function, the experimental results are as follows.

Figure 9 compares the time complexity and error sum of the algorithm in this paper with other clustering algorithms and compares the time complexity and error sum of the three methods. On the whole, there is a relatively obvious upward trend, and the curve rising speed of the two methods except the algorithm in this paper is relatively obvious. Curves are twists and turns.

5. Experimental Results

Development Environment. The whole system is developed using Java language, and the specific development tools involved and the corresponding function descriptions are shown in Table 3.

Since this paper uses ontology to achieve semantic information retrieval, in the case of query expansion, the query results will be sorted according to the similarity between expansion conditions and ontology concepts.

Performance Evaluation and Comparison. The advantages and disadvantages of a system usually need to be

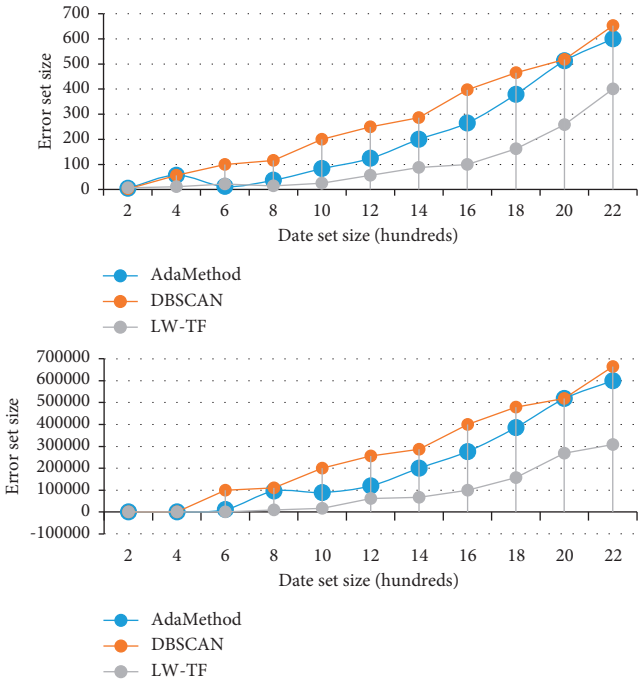

Figure 9: Comparison of the time complexity and error sum of the algorithm in this paper and other algorithms.

comprehensively judged from many aspects. The most common ones are functional verification and performance verification. Functional verification is to determine whether the target system has completed its own development requirements. Performance verification is to determine whether the time overhead and space overhead of the target system are low.

For an information retrieval system, the performance of the system can usually be evaluated from three perspectives: accuracy, recall, and F value. The accuracy rate focuses on the returned results and determines how many of the returned results are related to the query conditions. The recall rate focuses on all relevant documents and looks at the ratio of the number of relevant documents in the returned result set to the number of all relevant documents. Assuming that, in the returned results, the quantity related to the query is recorded as Hf and the number of documents not related to the query is recorded as Tp, in the unreturned results, the number of documents related to the query is recorded as Te, and the number of documents not related to the query is recorded as Hk. The formula for calculating the accuracy is as follows:

$$\mathrm{Pr} = \frac{Hf}{Hf + Tp} \cdot 100\%. \tag{16}$$

Calculation of recall rate is as follows:

$$\mathrm{Re} = \frac{Hf}{Hf + Te} \cdot 100\%. \tag{17}$$

The above calculation formula shows that the recall rate and the accuracy rate cannot be satisfied simultaneously in a retrieval system. When a retrieval system pursues high accuracy, the system will match the documents with the highest relevance to the query conditions to ensure that the retrieved data is the data relevant to the query conditions. However, the negative effect of this strategy is to ignore the data that may be related to the query conditions, resulting in the overall small scale of the retrieved data, which in turn affects the recall rate. Conversely, when the system pursues a high recall rate, it cannot guarantee that the retrieved data are all related to the query conditions. Although the overall size of the retrieval results has increased, the accuracy has decreased. Therefore, in order to reconcile the contradiction between the precision rate and the recall rate and take into account both, the F value can be used for evaluation. This value is the harmonic mean of the precision rate P and the recall rate R. The formula for calculating the F value is as follows:

$$\begin{aligned} F &= \frac{1}{\theta(1/Q) + (1-\theta)1/K} \\ &= \frac{\left(\varphi^2 + 1\right)\mathrm{QK}}{\theta^2 Q + K}. \end{aligned} \tag{18}$$

A in this represents the regulator B. According to the requirements, when C is greater than 1, the table recall rate corresponds to a high weight, and when C is less than 1, it indicates that the accuracy rate corresponds to a high weight. When it is equal to 1, the weights of the two are equal. The calculation formula of the F value at this time can be simplified as

$$F1 = \frac{2\mathrm{QK}}{Q + K}. \tag{19}$$

Accuracy Rate Histogram. Under multiple query conditions, first, the accuracy rate of each query condition under each retrieval algorithm is calculated separately. Then the difference in accuracy of different retrieval algorithms for the same retrieval condition is calculated. Finally, the difference is represented in the form of a histogram. In order to better distinguish between experiments, this paper selects relevant literature in the field of computer and selects literature in other fields that are not related to computer as a comparison. This paper randomly selects five retrieval conditions of "information security," "database," "network," "storage," and "operating system" and conducts inquiries in the two retrieval systems, respectively. According to the search results, the specific accuracy rate is calculated according to the above formula as shown in Figure 10.

It can be seen from Figure 10 that, among the five selected query conditions, the accuracy rate corresponding to

Table 3: Development tools.

Tool name	Tool use
Silr4.6.1	Search frame
JDK1.7	Java development kit
Tomcat7.0	Web server
Eclidse3.2	Development environment
Protégé	Ontology construction
Jena	Ontological reasoning

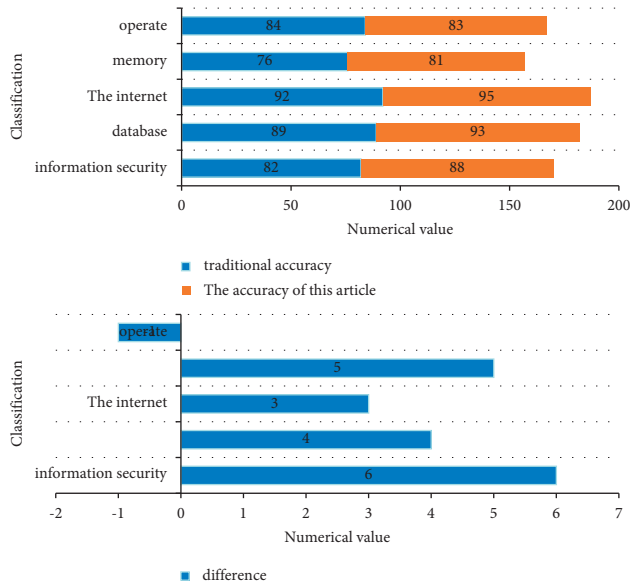

Figure 10: Accuracy comparison chart and accuracy difference histogram.

Table 4: Average precision under standard recall.

(%)	Information security		Database		The Internet		Memory		Operate	
	Tradition	This article	Tradition	This article	Tradition	This article	Tradition	This article	Tradition	This article
0	0	0	0	0	0	0	0	0	0	0
10	95	90	95	66.7	100	100	75	100	100	100
20	100	100	80	80.5	92	80.6	75	100	100	100
30	82	90	68.5	90.9	89	85	65.8	100	100	100
40	75	100	71.7	93	88	59	80	100	100	100
50	82	86	86.3	90	87	52	75	92.4	100	100
60	82	88	80	90	84.3	42.7	65.2	92	100	100
70	85	88	78	88.7	77.6	36	65	90.5	100	100
80	78	78	80	87	63	39	64.5	88.9	100	100
90	75	76	78	85.9	0	38.6	64	87.6	100	100
100	71	83	80	85	0	0	50	64.7	100	100

four query conditions in the retrieval system implemented in this paper is higher than that of the traditional retrieval system. The difference of the fifth query condition is −1, indicating that the retrieval method in this paper is lower than the traditional retrieval system under this conditional query. At the same time, it further illustrates that the retrieval system based on ontology can improve the retrieval accuracy to a certain extent.

Performance Test. According to the retrieval results, the average precision rate under the standard recall rate is obtained by sorting and analysis, as shown in Table 4.

According to the data in Table 4, the corresponding accuracy curve graph is drawn under the 11-point standard recall rate, as shown in Figure 11.

It can be seen from Figure 11 that the retrieval system implemented in this paper is overall better than the traditional retrieval system in terms of accuracy. When the recall rate is between 10% and 40%, the retrieval system achieved in this paper has consistently higher accuracy than traditional retrieval systems. It is shown that, by expanding the query conditions, more matching documents can be retrieved. When the recall rate is between 40% and 80%, the

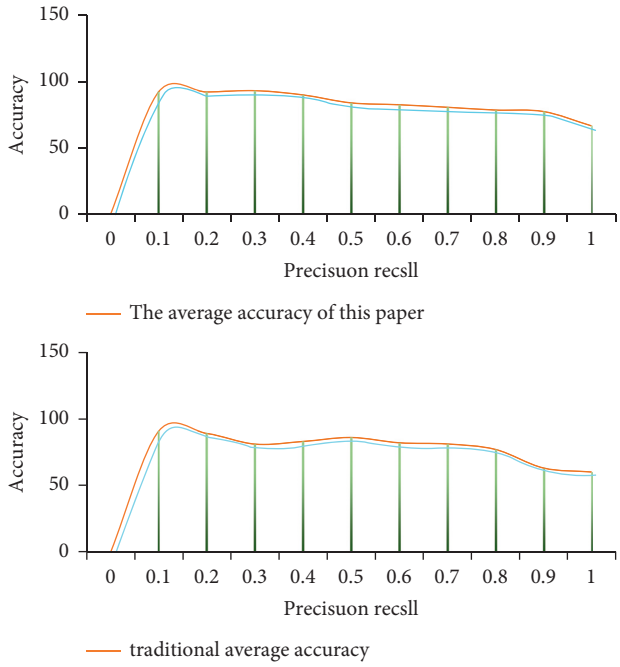

Figure 11: Accuracy at standard call rate.

accuracy of the retrieval system implemented in this paper is not much different from the traditional retrieval system. However, when the recall rate exceeds 80%, the accuracy of the retrieval system implemented in this paper is nearly 15% higher than that of the traditional retrieval system. It can be concluded that the retrieval system implemented in this paper has a certain degree of improvement in accuracy compared with the traditional retrieval system.

6. Conclusions

This study mainly verifies the algorithm proposed in this paper from the perspective of experiments and analyzes different problems, such as recall rate, accuracy rate, and average error, using no evaluation method. The experimental results show that the establishment of the automated method of news domain ontology proposed in this paper is effective, and the validity of the prokernel of semantic search system built on the basis of news domain ontology knowledge base is verified.

Finally, a semantic retrieval system for computer-related documents is initially implemented. This system is used to evaluate the advantages and disadvantages of ontology construction, and it also verifies whether the combination of ontology technology and literature retrieval is feasible. Through functional and performance experiments, it is proved that the ontology constructed in this paper is effective. At the same time, it is proved that it is feasible to combine ontology technology and literature retrieval. When constructing the ontology, only part of the information is used, resulting in a small scale of the ontology constructed. At the same time, as the scale of online literature data continues to expand, the domain ontology constructed based on the old literature may not fully reflect the current knowledge characteristics. At the same time, the definition of the object attributes of ontology is relatively simple and rough, and it cannot fully reflect the relationship between document concepts. Therefore, it is necessary to modify and update the established ontology in time in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This research study was sponsored by 2020 Zhejiang Provincial First Class Online Course: Business Translation. The author acknowledges the project for supporting this article.

References

[1] Q. S. Qiang, T. Huang, H. Tang, and P. Jiang, "An improved non-rigid point set registration algorithm by preserving local topology," *Pattern Recognition and Image Analysis*, vol. 31, no. 4, pp. 646–655, 2021.

[2] K. Stylianopoulos and K. Koutroumbas, "A probabilistic clustering approach for detecting linear structures in two-dimensional spaces," *Pattern Recognition and Image Analysis*, vol. 31, no. 4, pp. 671–687, 2021.

[3] Y. Zhang, X. Luo, B. Zhang, and S. Zhang, "Semantic approach to the automatic recognition of machining features," *International Journal of Advanced Manufacturing Technology*, vol. 89, no. 1-4, pp. 417–437, 2017.

[4] K. Ramirez-Amaro, M. Beetz, and G. Cheng, "Transferring skills to humanoid robots by extracting semantic representations from observations of human activities," *Artificial Intelligence*, vol. 247, pp. 95–118, 2017.

[5] N. V. Maksimov, O. L. Golitsina, K. V. Monankov, A. A. Lebedev, N. A. Bal, and S. G. Kyurcheva, "Semantic search tools based on ontological representations of documentary information," *Automatic Documentation and Mathematical Linguistics*, vol. 53, no. 4, pp. 167–178, 2019.

[6] Z. Fu, F. Huang, K. Ren, J. Weng, and C. Wang, "Privacy-preserving smart semantic search based on conceptual graphs over encrypted outsourced data," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1874–1884, 2017.

[7] L. Sun, K. Yang, X. Hu, W. Hu, and K. Wang, "Real-time fusion network for RGB-D semantic segmentation incorporating unexpected obstacle detection for road-driving images," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5558–5565, 2020.

[8] J. Devezas and S. Nunes, "A review of graph-based models for entity-oriented search," *SN Computer Science*, vol. 2, no. 6, pp. 437–536, 2021.

[9] R. Handa, C. R. Krishna, and N. Aggarwal, "Document clustering for efficient and secure information retrieval from cloud," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 15, Article ID e5127, 2019.

[10] L. Chen, S. Shang, B. Yao, and K. Zheng, "Spatio-temporal top-k term search over sliding window," *World Wide Web*, vol. 22, no. 5, pp. 1953–1970, 2019.

[11] S. R. Scott, "Tropes and some ontological prerequisites for knowledge," *Metaphysica*, vol. 20, no. 2, pp. 223–237, 2019.

[12] G. Ren, R. Ding, and H. Li, "Building an ontological knowledgebase for bridge maintenance," *Advances in Engineering Software*, vol. 130, pp. 24–40, 2019.

[13] C. Toraman and F. Can, "Discovering story chains: a framework based on zigzagged search and news actors," *Journal of the Association for Information Science and Technology*, vol. 68, no. 12, pp. 2795–2808, 2017.

[14] B. Wang, L. Ming, and H. Wang, "Circular range search on encrypted spatial data," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 704–719, 2017.

[15] K. Yongzhen, M. Weidong, and Q. Fan, "Image forgery detection based on semantic image understanding," *International Journal of Hospitality Information Technology*, vol. 7, no. 2, pp. 109–124, 2017.