A Semantic Annotation Approach for Professional Literature

WEI Moji†, XU Jianliang†, YU Tao†, YUN Hongyan†

†Department of Computer Science, Ocean University of China, Qingdao, 266100, China

Abstract: This paper presents an automatic annotation method for professional literature. Through comparing with other storage formats and literary styles, we summarize two features of professional literature, and then propose three assumptions. To improve annotation efficiency firstly the method, based on topology structure, partitions the domain ontology into segments which are self-consistent, then locates the most related segment(s) with the keywords extracted from document, finally annotates the document with located segment(s) and expands the annotation scope according to the correspondence between grammatical structure and semantic structure. Experiment results show that the method could improve annotation efficiency and annotation accuracy.

Key words: professional literature, semantic annotation, ontology partition, semantic context, syntax context

1 Introduction

Learning professional literatures is an important way for people to gain knowledge. These documents are written by domain experts, as a result, they are credible and form the basis of study and research. In order to facilitate sharing and searching, professional literatures of various fields are converted to electronic documents and published on the internet. However, current keyword-based search engines are hard to solve the ambiguous problems brought by polysemy and synonym, accordingly the distinction burden falls on human. To capture knowledge we have to try to search synonymous keywords several times and filter the irrelevant information brought by polysemy. This inefficient way would cost much time and effort. With abundance of electronic documents, available resources retrieving has become the bottleneck of knowledge acquisition.

By expanding World Wide Web with semantic annotation, the Semantic Web gradually evolves out[1]. It annotates resources on the web with formal ontology to make resources machine-processable[2], and it also provides knowledge-based search engine. In this improved search engine, agents pre-estimate correlation between resources and keywords according to the semantic, by this way the efficiency would be improved and the time would be shorten as well. The precondition of this approach is the resources on the web have been well-annotated. However, comparing with other technologies of the Semantic Web, semantic annotation appears to be more immature, especially the annotation of professional literatures.

In recent years, automatic semantic annotation has received much attention in the research community. Due to diversity of languages and formats used in the net, there is no such a standard annotation method that can satisfy all the situations. The methods proposed by different literatures have different features, and the targets they oriented are different from each other as well. [3] utilizes semantic pointer to annotate candidate pair using hierarchical relation between concepts and synonymous relation among words. [4] argues that instead files ontologies should also be published as shared services that could be easily and cost-efficiently integrated into applications to extend content management systems with ontological annotation capabilities. [5] makes use of natural language processing based concept recognizers or named entity recognition tools to identify the related concepts in the textual metadata describing a data. [6] proposes hierarchical free tagging method to complement gaps between existing ontologies in annotation process. [7] proposes a tree-structured conditional random field for semantic annotation. The annotated instances would form a hierarchical tree through which we could easily find the sub/sup-concepts or siblings. There are also non-professional documents which contain amounts of

ambiguous words commonly used in everyday language. The mapping of the regions of text to entries in ontology becomes harder. [8] explores how the application of domain models can complement traditional statistical NLP techniques to increase entity spotting accuracy in informal content. [9] and [10] research annotation for Wikipedia which is an important encyclopedia. The former one resorts to relation population to automatically extract relation between entity pair to enrich semantic data on the Wikipedia, and predicate suggestion to recommend proper relation labels to facilitate semantic annotating. [10] proposes BPOL algorithm which extends State-of-the-art algorithm to semi-automatically extract semantic relations between Wikipedia entities from the free text. Semantic annotation is also used in social community. [11] assumes the Desktop reflects all the trends and new interests of a user while it also tracks his/her history. Based on this assumption the paper proposes a novel approach for personalized annotation of Web papers by exploiting Desktop documents. [12] is another semantic annotation paper for social community. It proposes a new method to map tags with terms from thesauri or taxonomies (and vice versa), and gives an information theoretic measure for the quality of that mappings.

Moreover in the study of document classification, there are also literatures about semantic annotation [13-16]. These researches devote to document category management and help allocating the documents to the appropriate fields.

2 Professional literature analysis

2.1 Features of professional literature

The annotation target of this paper is professional literatures which are written by domain experts. Before annotation, we analyze the features of professional literatures first.

For one thing professional literatures are stored in text file such as txt, doc, pdf etc, it is obvious that the structure of this format is quite different from the structure organized by the web page or database. Web page is semi-structured document and the database is structured document, in this semi-structured or structured document, the concepts and the relations between concepts could be described by XML or E-R diagram respectively. On the contrary professional literature is unstructured document where concepts and relations are implicit rather than explicit. However, almost all the actual annotation methods for the semantic web are based on the structures of documents, and then these methods could not be used in professional literatures. Accordingly this paper parses the grammatical structure of document to find out the implicit concepts and explicate relations between them.

For another professional literatures are used to explain knowledge in the fields, therefore they have some special features that novels, poems, essays and other literary styles do not have. Professional literatures focus on interpreting concepts and relations between concepts in the fields, consequently the words in the documents are more normative and the cohesion is higher. In the study of document classification, documents are commonly annotated by lightweight ontology or taxonomy metadata, through annotation professional literatures could be allocated to the appropriate fields. However, in professional literature even the field is same, the concepts and relations involved may be have nothing in common, so it is hard to clearly express the idea of the document just by metadata like title, keywords, abstract, etc. In conclusion the annotation method based on lightweight ontology or taxonomy metadata could not locate concepts of articles precisely. Currently each field has already had mature domain ontology such as GEO-Ontology in the geographic information field, Gene Ontology in the gene field and NCI Ontology in the cancer field. These heavyweight ontologies cover the most important concepts of the fields and also provide abundant semantic information. After several visions upgrading, these ontologies have been widely recognized by domain experts. The annotation method proposed by this paper uses these heavyweight ontologies to annotate professional literatures.

By analyzing the storage format and literary style of professional literature, it can conclude that professional literature holds two features: (1) the documents are unstructured, it means the concepts and the relations between concepts are implicit; (2) the concepts in the documents are dense. Because of these two features, actual annotation method could hardly be used for professional literatures. For the purpose of professional literature annotation, based on the features of content and terms, we propose three assumptions.

2.2 Three assumptions about professional literature

Assumption 1. Professional literature is sequential, each portion of the document presents a subdomain of the field, and namely the document is consistent with local area of the domain ontology.

Each field contains plentiful facets and each facet covers a great deal of concepts. For instance NCI ontology has more than 17000 concepts. While one document impossibly illuminates the whole field, it describes partial concepts and relations instead.

Assumption 2. The portions of the document are relevant, and the connections between the portions are seamless rather than unexpected.

There are strong correlations between the portions of professional literature, in other words the concepts referred in the document are semantically related.

Assumption 3. The words used in the same document are unchanged.

It means that the same word in one document has the same meaning, and to state one meaning the word used in the document is unaltered. The word and the meaning are one-to-one correspondence.

To state the relation between the words and the document portion it belongs to, we formalize our first notion of syntax context as follow:

Definition 1 (Syntax Context). Let D be a document, w be a word, P be a paragraph, and Ch be a chapter. We say that P, Ch, D are syntax contexts of w, if $w \in P$, $P \subseteq Ch$ and $Ch \subseteq D$. In these syntax contexts we say that D is maximum syntax context, P is minimum syntax context, and Ch is medium syntax context.

3 Ontology partition

According the assumption 1, any professional literature could just present partial concepts of the field, while the domain ontology which defines the field is usually large. Because of the conflict if using the whole ontology to annotate the document, a mount of computing resources would be wasted on ontology parsing and concepts locating, which will inevitably reduce the efficiency and accuracy of annotation. Take ACM (Association for Computing Machinery) ontology for example, in document the word “model” is associated with several concepts like Data Models, Process Models, Language Models, etc. The semantic distance between these concepts is large. Annotating the word “model” may easily choose a wrong concept when the context is obscure. To reduce the interference, we assign the concepts which have large semantic distance to different ontology segments.

By assumption 2 it can infer that the concepts in one professional literature are semantically related. In the formal definition of ontology, role has Domain and Range two elements which link two concepts and restrict the relation between the concepts. Through connection of role, the space between concepts which have close semantic distance is close either, and with increment of semantic distance the space would also increase. For improving efficiency and accuracy we would partition the domain ontology and use segments to annotate the professional literature.

In ontology partition there are mainly two approaches. One is splitting the ontology according to the logical structure. The other is partition based on topology structure of the ontology. The first approach has to calculate the weight of relevancy between each two concepts according to axiom and formula. The computation performs exponential growth as concepts grow linearly, thus computational complexity of the first approach is very high, besides this method often fails in large ontology partition. Unfortunately, the domain ontology usually possesses vast concepts, and then it means that the first partition approach hardly split the ontology within acceptable time. As mentioned above we concentrate on the relation between space of the concepts and the semantics, thereby we choose the second approach to partition domain ontology.

Heiner Stuckenschmidt[17] et al proposes a ontology partition method based on the topologic structure. The process is divided into two tasks which contain five steps: Creating Dependency Graph, Determining Strength of Dependencies, Determining Modules, Assigning Isolated Concepts and Merging. Besides they also provide a partitioning tool: Pato (Partitioning Tool for Ontologies). The partition method assigns a concept to a segment if and only if the relationship between the concept and the segment is stronger than the relationship between the concept and other segments. By this way the concepts which have much more links could be centralized to one segment, and the segment divided is much more self-consistent. Finally they have tested the validity of the method by Pato. In our research we adopt Pato as our ontology partition tool as well.

To state the relation between the words and the ontology segments it relates to, we formalize the notion of semantic context as follow:

Definition 2 (Semantic Context). Let C be a concept, w be a word, O be a domain ontology, and Os be a segment of the ontology. We say that Os is semantic context of w, if $C \in Os$, $Os \subseteq O$ and $w^I \in C^I$.

4 Professional literature annotation

4.1 Ontology segment(s) selection

After partition, the domain ontology would be split to many segments. Among these segments we have to choose one segment which has the highest semantic relevancy with the syntax context to annotate the document. The selection process is as follow:

Step1: Extracting keywords from the maximum syntax context by the technology of ontology learning.

Step2: Through comparing keywords and concepts in the segments, selecting one or more semantic contexts which have higher semantic relevancy with syntax context.

Step3: If the keyword-related concepts centralized in one semantic context, we say the semantic context matches syntax context, then using matched semantic context to annotate syntax context.

Step4: Otherwise, if keywords-related concepts evenly scattered to several semantic contexts, the matching process fails.

Step5: If mismatched, shrinking syntax context and extracting keywords from the shrunken syntax context, and then repeating Step2 to Step4, till matching successful or shrinking to the minimum syntax context.

Algorithm 1 gives segment(s) selection algorithm.

```
Algorithm 1 Segment(s) selection
Boolean                        isMinContext = false;
Ontology Segment               OS, MOS[];
SytaxEnviromentStack           stack;
SytaxContext                   curSyCon;

INPUT Ontology Segments;
INPUT document;
push stack(document);                  // push document to the stack
DO WHILE NOT empty stack
   curSyCon = pop stack();
   Extract keywords from curSyCon;
   IF curSyCon equals minimum syntax context
      isMinContext = TRUE;
   ELSE
      isMinContext = FALSE;
   ENDIF
   // the if condition is determined by algorithm 2
   IF keywords-related concepts evenly scattered to several semantic contexts
      IF NOT isMinContext
         shrink syntax context;
         push stack(shrunk syntax context);
      ELSE
         MOS[] = multi semantic contexts;     // MOS[] is MSC[] in algorithm 2
         annotatetransitionalparagraph(curSyEnv, MOS);
      ENDIF
   ELSE
      OS = single semantic context;           // OS is SC in algorithm 2
      annotate(curSyEnv, OS);
   ENDIF
ENDDO
```

In the keywords extraction process, we calculate the relevancy between words and the field by technology of ontology learning. (4-1) shows the relevancy calculation formula.

$$DR_{t,k} = \frac{P(t \mid D_k)}{\max\limits_{1 \le i \le n} P(t \mid D_i)} \qquad (4-1)$$

In (4-1) DS={D_1,D_2,...D_n} is set of fields and D_k is a field to which the document relates. t is a word extracted from the document. $DR_{t,k}$ denotes relevant degree between word t and field D_k. Conditional probability P(t|Dk) could be calculated by formula (4-2).

$$E\left(P\left(t \mid D_k\right)\right) = \frac{f_{t,k}}{\sum\limits_{t \in D_k} f_{t,k}} \qquad (4-2)$$

In (4-2) $f_{t,k}$ denotes the frequency of word t occurring in the field D_k.

We adopt several words which have higher relevant degree as keywords to match concepts. If calculating semantic distance between concept and keyword as relevant degree, we would match the highest relevant concept for keyword quantitatively. However, as the aforementioned domain ontology covers vast concepts, calculating semantic distance as relevant degree results in low efficiency. Therefore in our method we use string matching plus statistics to locate semantic context. The location process is as follow:

Step1: Through string matching, comparing each ontology segment with all the keywords. If the concept of one segment matches the keyword, we say the segment hits the keyword and record the keyword to the segment.

Step2: Counting the keywords that the ontology segment hits, and calculating the hit rate of each segment.

$$\text{Hit rate} = \frac{\text{hit keywords}}{\text{all the keywords}}$$

Step3: Comparing hit rates.

Step3.1: If the hit rate of one segment is much higher than others, then the segment is the very semantic context we want.

Step3.2: If the hit rates of several segments are close, then semantic context location fails. Failures can be divided into two cases.

Case 1: If the repetition rate of the keywords recorded in the different segments is low, it means the problem comes from the syntax context selected. The solution is shrinking syntax context, extracting new keywords and locating semantic context with them once more.

Case 2: Otherwise if the repetition rate of the

keywords recorded in the different segments is high, it means the number of the selected keywords is missed. In that way we would calculate the hit rates by increasing or decreasing the keywords over again.

Annotation experiment shows that there would be several segments having the similar hit rates when the keywords are few. With the increment of the keywords, the gap among hit rates is enlarged and it would reach peak at some point. Afterwards the gap among hit rates would become narrow with the sustained growth of keywords. In conclusion either too few or too many keywords would induce the problem introduced in case 2. To process case 2, it is a good idea by adjusting the amount of the keywords.

Algorithm 2 gives semantic context location algorithm

```
Algorithm 2 Location
Integer                 matched[];
Sematic Contexts        SC,MSC[];
String                  matchedwords[][];
Double                  matchedrate[];
Boolean                 finished=FALSE, sign=FALSE, increase=TRUE;

INPUT maxKeywordnum;
INPUT sematicContexts[];
DO WHILE NOT finished
   INPUT keyword[maxKeywordnum];
   FOR i=1 TO count of semanticContexts[]
       FOR j=1 TO maxKeywordnum
           IF StringMatch(keyword[j], sematicContexts[i])
              matchedwords[i][k++] = keyword[j];
              matched[i]++;
           ENDIF
       ENDFOR
   ENDFOR
   FOR i=1 TO count of semanticContexts[]
       matchedrate[i] = matched[i] / maxKeywordnum;
   ENDFOR
   IF max matched rate >> other matched rate
      SC = max matched rate sematic context;
      finished = TRUE;
      RETURN FALSE;
   ELSE
      IF NOT highoverlap(matchedwords[] which have similar hit rate)
         MSC = ontology segments which hit rates are close;
         finished = TRUE;
         RETURN TRUE;
      ELSE
         IF gaps among hit rates increase
            maxKeywordnum = maxKeywordnum + increasement;
            increase = TURE;
         ELSE
            maxKeywordnum = maxKeywordnum - increasement;
            increase = FALSE;
         ENDIF
         clear matchedwords[][];
         IF sign XOR increase
            sign = sign XOR increase;
            increasement = increasement / 2;
         ENDIF
      ENDIF
   ENDIF
ENDDO
```

4.2 Single semantic context annotation

Annotation experiment shows that with Algorithm 1 the syntax contexts could match its corresponding semantic contexts, and for most cases one syntax context matches just one semantic context, and only in a few cases one syntax context matches two or more semantic contexts. We study how to annotate syntax context by single semantic context first. As above analysis the professional literature is usually an unstructured document, it means that the concepts and the relations between concepts are implicit. Hence the implicit should be explicated first. In document the grammatical structure of the clause interprets the relation of the components, and it gives us a clue to explicate the implicit. Semantic annotation[18] is about assigning the named entities in the document links to their semantic descriptions. According to above illumination we mainly concern the grammatical structure between notional words of the clause.

To parse each clause of the document would cost much time, but feature 2 told that the concepts in the professional literature are dense. So in order to explicate the relation between concepts, it is necessary to parse grammatical structure of each clause. According to assumption 2 the components of the clause are semantically related, and in ontology the semantic relations are defined by role. It is obvious that the grammatical structure of a clause and the semantic structure of the ontology are well corresponded, and therefore we can use the correspondence between grammatical structure and semantic structure to annotate the document. Figure 1 gives the sample sentence to show the flow of annotation method. The sentence is taken from the document about computation theory.

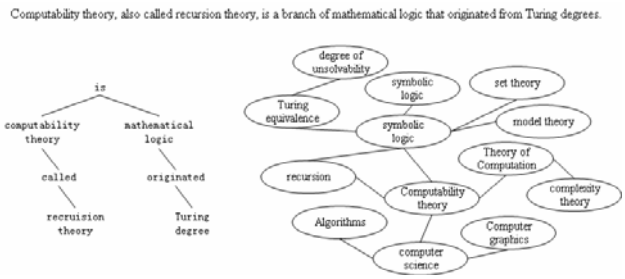

Fig. 1 sample sentence

Firstly the annotation method annotates keywords with located semantic context. In the sample sentence “computability theory” is keyword of the document. Compared with the concepts in the semantic context, the concept “Computability theory” would be picked to annotate the keyword

Then the grammatical structure of the clause which contains the keyword would be parsed. Syntax tree is ideal structure to show the relation between the notional words such as the left part of the figure 1.

Thirdly the annotation method searches the nearest component to the annotated-word, then annotates it

with the concept which neighbors to the concept corresponding to annotated-word, and repeats the process till all the components of the clause are annotated. In figure 1 the nearest component of “computability theory” is “recursion theory”, comparing the semantic distance between word “recursion theory” and the concepts neighbors to “computability theory”, and then choosing concept “recursion” to annotate the word. After that the method annotates word “mathematical logic” with concept “symbolic logic”, and finally annotates word “Turing degrees” with concept “Turing equivalence”.

Finally according to assumption 3, the method expands the concepts to the clauses that state in other syntax contexts, parses the grammatical structures of these syntax contexts, and then repeats former processes. According to assumption 3, we consider that the meanings of the words “mathematical theory”, “recursion theory”, etc. are unchanged, so the concepts “symbolic logic”, “recursion”, etc. can be annotated to the corresponding words in other syntax contexts directly.

According to the correspondence between grammatical structure and semantic structure, the annotation scope could be restricted in a small local area, which would further reduce the concepts participated in comparing process to improve efficiency. By propagating along the grammatical structure and keeping the correspondence between document and ontology, the annotation scope would be expanded gradually and the accuracy would be kept high as well. Finally according to assumption 3, we rapidly annotate the same words in different syntax contexts.

4.3 Multi semantic contexts annotation

In annotation experiment, there are still a few cases that one syntax context matches two or more semantic contexts. If there is just one semantic context corresponding to the syntax context, it can be seen that the idea the syntax context presented is uniformity. If the syntax context which has been shrunk to minimum syntax context still crosses several semantic contexts, it means that the knowledge the paragraph represents is complex and the purpose of this kind paragraph is to state the relations among various knowledge. The paragraph of this kind is transitional paragraph which plays the role of bridging and connecting links, and the concepts it refers to usually state in the boundary of several semantic contexts. So to annotate this syntax context we mainly use the border concepts of the semantic contexts and the roles that connect the border concepts. As a result before annotation we have to gather these concepts and roles first, and then annotate the keywords with them, finally expand annotation scope based on the correspondence between grammatical structure and semantic structure.

5 Experiment results & analysis

ACM ontology is employed to verify our annotation method. The ACM is ontology for Computer Science, built by the Association for Computing Machinery. With ontology partition tool Pato, the ACM ontology is split into 21 segments. We get 8 computer related professional literatures from net, in these documents, 2 documents are about computer graphics, 3 are about theory of computation, and 3 are about Artificial Intelligence. Furthermore in these 8 documents 4 documents are taken from Wikipedia, 2 are come from Journals, and 2 are drawn from E-books. Following two sections give annotation results from two aspects: annotation efficiency and annotation accuracy.

5.1 Annotation efficiency

The third column of table 1 shows the time the original method used, and the fourth column gives the time in our method. Figure 3 is drawn according to the statistic results of table 1. Figure 3 shows that annotating the same document our method could improve efficiency.

Table 1 Annotation efficiency comparison

Field	Vocabulary	Original time	New time	Improved
Computer Graphics	1196	10076	466	9610
Theory of computation	1289	10859	963	9896
Theory of computation	1687	14212	1261	12951
Theory of computation	2111	17784	1578	16206
Artificial intelligence	2215	18660	2843	15817
Artificial intelligence	2341	19722	3005	16717

Computer Graphics	3580	30160	1246	28914
Artificial intelligence	5742	48373	4058	44315

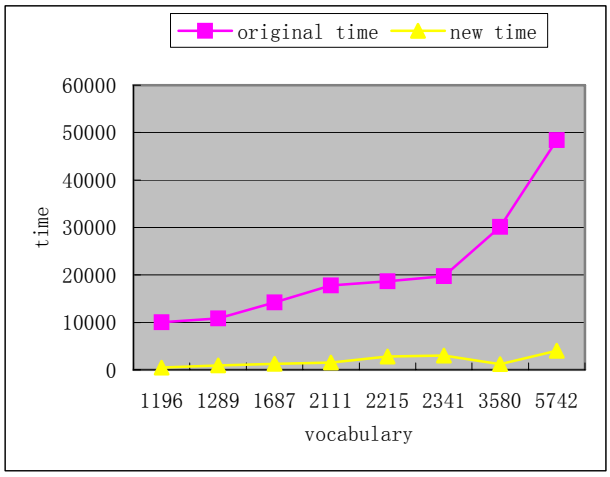

Fig. 2 comparison of annotation time

5.2 Annotation accuracy

The third column to the fifth column of table 2 respectively give the annotation amount, annotation accuracy and annotation accuracy rate without our method, and the sixth column to the eighth column respectively show the annotation amount, annotation accuracy and annotation accuracy rate by our method. Figure 3 and figure 4 are drawn based on the statistic results of table 2, through comparing it can conclude that annotating the same document our method could improve both annotation amount and annotation accuracy.

Table 2 Annotation accuracy comparison

Field	Vocabulary	Original annotation	Original accuracy	Original accuracy rate	New annotation	New accuracy	New accuracy rate
Computer Graphics	1196	187	132	70.59%	169	127	75.15%
Theory of computation	1289	159	119	74.84%	140	110	78.57%
Theory of computation	1687	226	159	70.35%	201	148	73.63%
Theory of computation	2111	243	187	76.95%	217	169	77.88%
Artificial intelligence	2215	307	214	69.71%	276	201	72.83%
Artificial intelligence	2341	322	231	71.74%	295	219	74.24%
Computer Graphics	3580	476	329	69.11%	425	313	73.65%
Artificial intelligence	5742	475	246	52.90%	416	225	54.10%

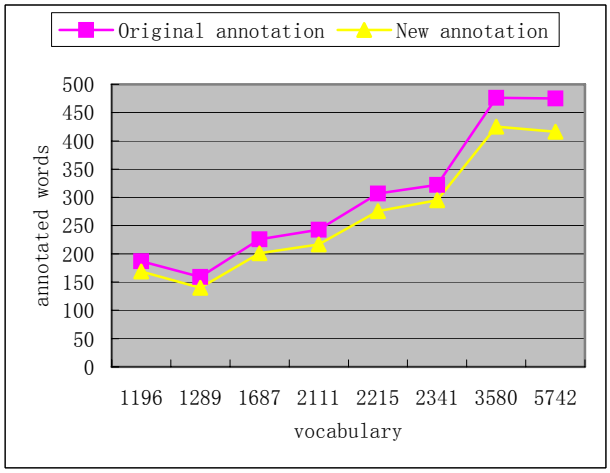

Fig.3 comparison of annotated words

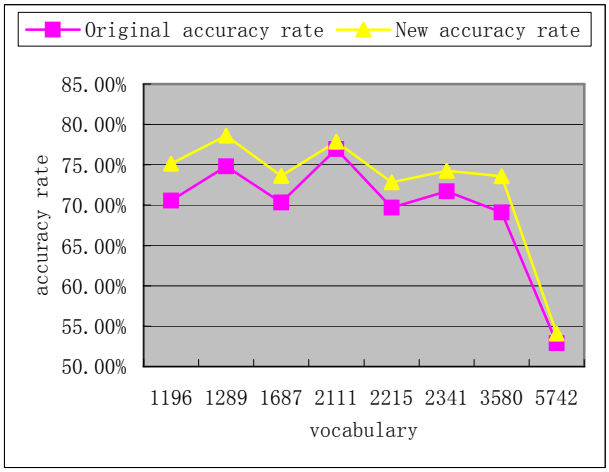

Fig. 4 comparison of accuracy rate

6 Conclusion

The absence of semantic annotations has become the obstacle of semantic web. This paper provides an automatic annotation method for professional literature. Through comparing, this paper concludes two features of professional literature: unstructured and concepts dense, then we propose three assumptions. To improve annotation efficiency our method partitions ontology into segments which are self-consistent, then extracts keywords from syntax context with technology of ontology learning. With these keywords our method would locate semantic context for syntax context, and annotate these keywords with the concepts in the located semantic context. Finally the annotation scope would be expanded based on the correspondence between grammatical structure and semantic structure. We have annotated computer related professional literatures with ACM ontology, the experiment results show that our method could increase the amount of the semantics and improve the annotation efficiency and annotation accuracy.

As annotation of other resources on the net, the annotation algorithm should be suitable for massive documents. This paper merely studies the annotation for single document. The next step of our work focuses on how to distribute annotation tasks to multi-nodes. That is design a task division program to balance load.

References

[1] T. Berners-Lee, J.Hendler, and O. Lassila. The Semantic Web. Scientific American 284 (May 2001). 34-43

[2] Dill, S., Tomlin, J., Zien, J., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Rajagopalan, S. and Tomkins. SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation. Proceedings of the 12th International World Wide Web Conference, WWW(2003). 178-186.

[3] Pazienza M.T., Stellato A., An Environment for Semi-automatic Annotation of Ontological Knowledge with Linguistic Content , 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, June 2006. 442-456

[4] Kim Viljanen, Jouni Tuominen, Eero Hyvönen, Eetu Mäkelä, Osma Suominen. Extending Content Management Systems with Ontological Annotation Capabilities. 6th International and 2nd Asian Semantic Web Conference (ISWC2007+ASWC2007)

[5] Clement Jonquet, Nigam Shah, Cherie Youn, Chris Callendar, Margaret-Anne Storey. NCBO: Annotator Semantic Annotation of Biomedical Data. 8th International Semantic Web Conference (ISWC2009)

[6] Marta Gonzalez, Stefano Bianchi, Gianni Vercelli. Semantic framework for complex knowledge domains. 7th International Semantic Web Conference (ISWC2008)

[7] JieTang, H. Mingcai, L. Juanzi, and L. Bangyong, Tree-structured conditional random fields for semantic annotation. 5th International Semantic Web Conference (ISWC2006). 640-653

[8] Daniel Gruhl, Meena Nagarajan, Jan Pieper, Christine Robson, Amit Sheth. Context and Domain Knowledge Enhanced Entity Spotting in Informal Text. 8th International Semantic Web Conference (ISWC2009). 260-276

[9] Haofen Wang, Linyun Fu and Yong Yu. Bricking Semantic Wikipedia by Relation Population and Predicate Suggestion. 3rd China Semantic Web Conference (CSWS2009)

[10] Gang Wang, Yong Yu, Haiping Zhu. PORE Positive-Only Relation Extraction from Wikipedia Text. 6th International and 2nd Asian Semantic Web Conference (ISWC2007+ASWC2007)

[11] Paul - Alexandru Chirita , Stefania Costache , Wolfgang Nejdl , Siegfried Handschuh, P-TAG: large scale automatic generation of personalized annotation tags for the web, Proceedings of the 16th international conference on World Wide Web. 2007 845 - 854

[12] Christian Wartena, Rogier Brussee. Instanced-based Mapping between Thesauri and Folksonomies. 7th International Semantic Web Conference (ISWC2008) 356-370

[13] Ouyang Ning Bao Ping. The Visualization of Chinese Library Classification by Ontology Tool. LIBRARY JOURNAL. 2008 27(1)

[14] Tian Xin. Studying of Ontology-based Semantic Query System Model--A Case of Library. JOURNAL OF INFORMATION. 2006 25(6)

[15] Dong Hui Yang Ning Yu Chuanming Jiang Ying Xu Guohu Zhang Jidong. Research on the Ontology-based Retrieval Model of Digital Library (Ⅰ)--Explanation of the Architecture. JOURNAL OF THE CHINA SOCIETY FOR SCIENTIFIC AND TECHNICAL INFORMATION. 2006 25(3)

[16] Gao Fan, Li Jing. Ontology and the Ralations between Ontology, Taxonomy and Thesaurus. LIBRARY THEORY AND PRACTICE. 2005 (2)

[17] A. Schlicht and H. Stuckenschmidt. Criteria-based partitioning of large ontologies. In Proceedings of the International Conference on Knowledge Capture (K-CAP), 2007. Poster Contribution.

[18] A. Kiryakov, B. Popov, D. Ognyanoff, D. Manov, A. Kirilov, and M. Goranov. Semantic annotation, indexing, and retrieval. In Proc. of the 2nd Intl. Semantic Web Conference (ISWC2003), 2003. 484-499