International Journal of Multidisciplinary Research and Development 2015; 2(3): 609-611

International Journal of Multidisciplinary Research and Development

IJMRD 2015; 2(3): 609-611
www.allsubjectjournal.com
Impact factor: 3.672
Received: 08-03-2015
Accepted: 23-03-2015
E-ISSN: 2349-4182
P-ISSN: 2349-5979

Naikwade Namrata
(B.E. Computer Engg. Pune University) PREC, Loni, Maharashtra, India

Lolage Pooja
(B.E. Computer Engg. Pune University) PREC, Loni, Maharashtra, India

Sonawane Jui
(B.E. Computer Engg. Pune University) PREC, Loni, Maharashtra, India

P.B. Vikhe
(M.E Computer Engg.Pune University) PREC, Loni, Maharashtra, India.

Correspondence:
Naikwade Namrata
(B.E. Computer Engg. Pune University) PREC, Loni, Maharashtra, India

Tag and value based annotator

Naikwade Namrata, Lolage Pooja, Sonawane Jui, P.B. Vikhe

Abstract
The databases on web servers are accessible through HTML based search engines. The information mined from these web servers are mostly in unstructured format called as Search Result Record (SRR). Whenever a web user wants to search for a specific product, he/she reviews or goes through many sites to obtain a relevant result. This result is obtained in multiple data units. The data units returned from the underlying database are usually encoded into the result pages dynamically. The data which is essential for the user by comparing information from different sites are given meaningful labels.
We present an automatic annotation in which clustering-based shifting technique is used. First of all, our approach handles all types of relationships between text nodes and data units. Secondly, we use a variety of features together such as data units, their data types (DT), data contents (DC) and adjacency (AD) information to align data units on a result page into different groups such that the data in the same group have the same semantic. Then using clustering based shifting algorithm, for each group we annotate it from various aspects and aggregate the various annotations to predict a final annotation label.

Keywords: Web Database (WDB), Search Result Record (SRR), Data Content (DC)

1. Introduction
The use of Internet has been increased widely over a period of time. Also, the use of E-Commerce has increased rapidly since a decade. The Web Databases are accessed through HTML based search engine. The result returned from web database is in the form of Search Result Record (SRR). SRR contains text nodes and data units.
Here, we perform data unit level annotation. This is a high demand data of interest from multiple Web Databases (WDBs). For example, a book comparison shopping system collects multiple result records from various sites, it is necessary to check whether any two SRRs refer to the same book. The system also needs to list the prices given by each site. Hence, the system needs to know the semantic of each data unit. The annotation and data alignment are performed automatically using their respective algorithms.
The system is sometimes unable to distinguish among data tags of different web based languages. So, we are overcoming by distinguishing among the data tags. The main application of this system is E-Commerce.

2 Working Model:
In this paper, we consider how to automatically assign labels to the data units within the SRRs returned from WDBs. In this system, we perform data unit level annotation. Annotation of web pages is necessary for applications such as comparison of book shopping, deep web collection etc. For example, a book comparison system collects various search records from various websites and it finds whether two records are from the same book or not. Due to this automatic comparison can be easily done if the data units are assigned with meaningful labels. Annotation is essential for easier storage of data into tables. The result page consists of various search result records i.e. SRRs.
Once the search result records have been extracted from the result page, annotation gets performed in two different phases. The first phase is the annotation phase which performs different types of annotations for the assignment of labels. The second phase is the alignment phasewhich organizes the data units into groups of the same semantic.
As shown in the figure 1.Architectural Model consists of three different parts which are Input, Process and Output.The Input Part consists of the query entered by the user from web database. The second part that is the Process consists of three phases i.e. SRR Extraction, Annotation and Alignment. And after this process the result page gets generated in the form of output in the table format.

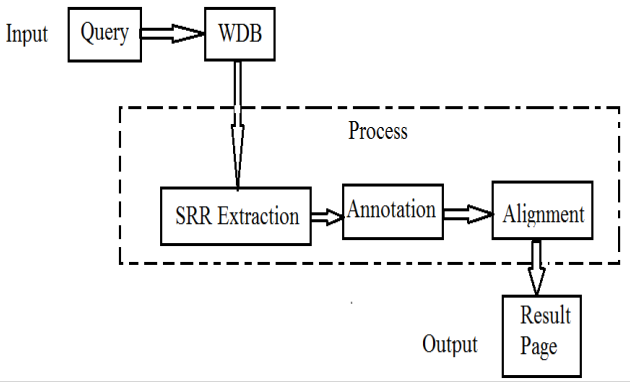

Fig 1: Architectural Model

2.1 Input:
Query entered by the user on web page.
Example:-User entered a Query for Searching book "Revolution 2020".

2.2 Output:
We get a well aligned and structured result in the table format by using alignment and clustering algorithm.
As per the Query entered by user the result will be displayed in the following format.

PUBLICATION	AUTHOR	NAME	YEAR	PRICE
Rupa	ChetanBhagat	Revolution 2020	2011	140
Rupa	ChetanBhagat	Revolution 2020	2011	130
Rupa	ChetanBhagat	Revolution 2020	2011	100

3 Algorithms:
Alignment Algorithm:

```
Align (SRRs, Annotator List)
{
 Display all Annotator on page
 For each annotator in Ai
{
  For i<-1 to no. of SRRs
{
  Flag=0;
  For j<-1 to SRR[i].length
{
Ej<-SRR[i][j]
     //jth element in SRR[i]
  If Ej==null then
  Continue;
 Else
  Al<-get annotator List from Ai
  If AL contains (Ej) then
 Display Ej+1 element on page
 Flag=1;
Break;
}
  If (Flag==0) then
  Display NIL on page
  Insert new Row for next SRR
}
}
}
```

Explainationof Algorithm:
The above algorithm shows alignment algorithm in which we are going to align SRRs and Annotator List. Here SRRs and Annotator List are the given inputs.i and j are the variables used in this algorithm.Here we first align SRRs and Annotator List.Then we display all the annotator on page for each annotator in Ai.Variable i contains number of SRRs.Variable Ej contains SRR[i][j] then if condition is applied where it checks whether Ej is equal to null i.eEj contains no SRR,if it is so then continue else get annotator list from Ai to A1.If AL contain Ej then display Ej+1 element on page.Then assign flag is equal to 1 and break the loop.If flag==0 then display NIL on page insert new row for next SRR.In this way by using the alignment algorithm we align SRR and annotator list.

4. Advantages:
1. Performs annotation quickly:
   Every annotatoris used to produce a label for the units within their group and is used to determine the most appropriate label for each group.
2. Automatic assigning of labels:
   Here, assigning of the label is done automatically as every SRR going through the annotation phase, is automatically assigned as the labels. Thus, automatic assigning of labels is possible.
3. Easy to search the data:
   When the user enters any query, the data to be searched from various websites becomes easy to search. Thus, it becomes easy for the user to search the data from various database.
4. Saves time:
   As the user gets the overall information from all the websites in a single table format as the output, it becomes very easy and efficient for the user to search and compare the data that he wants. Thus this saves thetime of the user.

Conclusion
Annotating large data in a single website may lower the processing speed and so it becomes difficult for the user to search the data efficiently. We analyzed data annotation problem and proposed a multi annotator approach to automatically constructing an annotation wrapper for annotating the search result records retrieved from any given web database. Accurate alignment is critical for achieving accurate annotation. Our method is a clustering method utilizing automatically obtainable features. Using clustering based shifting algorithm, for each group we annotate it from different aspects and aggregate the different annotations to predict a final annotation label for it.

References
1. S. Handschuh, S. Staab, and R.Volz, "ON DEEP ANNOTATION," Proceeding 12th International Conference World Wide Web (WWW), 2003.

2. H. Zhao, W. Meng, Z. Wu, V. Raghavan, and C. Yu, “FULLYAUTOMATIC WRAPPER GENERATION FOR SEARCH ENGINES,” Proceeding International Conference World Wide Web (WWW), 2005.
3. Y. Lu, H. He, H. Zhao, W. Meng, and C. Yu “ANNOTATING SEARCH RESULTS FROMWEB DATABASES” proceeding IEEE,2013.