ABSTRACT— The Internet presents a huge amount of useful information which is usually formatted for its users, which makes it difficult to extract relevant data from various sources. Therefore, the availability of robust, flexible Information Extraction (IE) systems that transform the Web pages into program-friendly structures such as a relational database will become a great necessity. The motivation behind such systems lies in the emerging need for going beyond the concept of “human browsing.” The World Wide Web is today the main “all kind of information” repository and has been so far very successful in disseminating information to humans[5].

The Web has become the preferred medium for many database applications, such as e-commerce and digital libraries. These applications store information in huge databases that user’s access, query, and update through the Web. Database-driven Web sites have their own interfaces and access forms for creating HTML pages on the fly. Web database technologies define the way that these forms can connect to and retrieve data from database servers.[3]

In this paper, we present an automatic annotation approach that first aligns the data units on a result page into different groups such that the data in the same group have the same semantic. And then we assign labels to each of this group. Keywords—Data alignment, data annotation, web database, wrapper generation.

I. INTRODUCTION

The Internet super highway is widely used as a vehicle to information sharing across the globe. People across the globe, of all walks of life, are accessing Internet resources through search engines. The search engines provide web based interface for information search. Search engines return huge amount of data which is presented in encoded format through web pages. However, the data comes from underlying database. The search results from web databases that can be used further in applications like price comparison, data collection, and other related applications. When the search word “LAP TOP PRICES” is given in Google, it returned 24,60,00,000 result pages. Some of the results obtained are as follows.

![Figure 1: Some of the search results](image)

As can be seen in Figure 1, it is evident that there are many search results that are associated with different web pages. The URL associated with each search result is different and the results came from many underlying databases in the web. The search results are to be made machine processable in order to use them further in real world applications. With the annotations, it is possible to process the web pages returned by search engines. For instance, the prices of various companies pertaining to a product can be compared. The comparison web sites can exist that derive data from across the pages returned by search engines. By providing price comparison, the web sites over Internet can help netizens to make well informed decisions. With many processing techniques, the search engines are presenting the results in meaningful way. Earlier the case was different. The results needed much human effort in order to annotate it manually. Recently Lu et al. [1] presented various ways of annotating the search results. They developed a mechanism that will automatically annotate the search results getting rid of manual labeling of web pages. Their solution
The paper is structured as follows. Section II reviews the literature that focuses on the prior work pertaining to automatic annotation of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.

II. IMPLEMENTATION

Our automatic annotation solution consists of three phases. They are illustrated in Figure 2.

The first phase is known as alignment phase where data units are organized into groups based on different concepts. The phase 2 is known as annotation phase which takes care of making annotators that annotate web documents automatically. The phase 3 is known as annotation wrapper generation phase where an annotation rule is generated for each identified concept. Annotation wrapper is the collection of all the rules for all groups which have been aligned. Annotation wrappers help improve the process of annotation. A clustering based scripting technique is used to achieve this.

In this paper we implemented few annotators presented by Lu et al. [1]. We built a prototype application that takes care of automatic annotations of search results. The research results are obtained through Google search. The results are automatically annotated using the mechanism proposed in [1]. The empirical results revealed that our prototype is useful and can be used in the real world. The remainder of the paper is structured as follows. Section II reviews literature that focuses on the prior work pertaining to annotation of search results. Section III presents the proposed approach to achieve automatic annotation of web search results. Section IV presents experimental results while section V concludes the paper.
than 90% precision and recall were recorded for both the performances such as data alignment and annotations. The table also shows the performance of annotation with wrapper. The results are presented in the following graphs.

As shown in Figure 5, Figure 6, and Figure 7, it is evident that the prototype application is capable of producing annotations automatically given search results of Google. The performance of the application is encouraging and the application can be used in the real world applications.

IV. CONCLUSION

In this paper, we automatically constructing an annotation wrapper for annotating the search result records retrieved from any given web database.

REFERENCES

[1] Annotating Search Results From Web Databases Yiyao Lu, Hai He, Hongkun Zhao, Weiyimeng, Member, Ieee, And Clement Yu, Senior Member, Ieee Ieee Transactions On Knowledge And Data Engineering, Vol. 25, No. 3, March 2013


[6] Wang Computer Science Department University of Science and Technology Clear Water Bay, Kowloon Hong Kong Computer Science Department University of Science and Technology Clear Water Bay, Kowloon Hong Kong


