

# Document Annotation Based on Query Workload, Content-Value and User Expectation Tracking Form

Alfia A P<sup>1</sup>, Chashu Mol R<sup>2</sup>

<sup>1,2</sup>M. Tech Student, Dept. of Computer Science & Engineering, Mount Zion College of Engineering, Pathanamthitta, Kerala, India

**Abstract:** Document Annotation means these are comments. Annotations are metadata, it give additional information about data. In olden days Form based query interfaces are used for accessing the databases but it has some limitations in the design of a forms-based interface and also they are capable of expressing only limited number of queries. In order to overcome the difficulty most of the high level organizations use the textual descriptions of their product and services. The Textual Descriptions consists of significant amount of structured information. So we use Annotations. If the documents are properly annotated we can increase the quality of searching to some extent. Here we introduce Collaborative Adaptive Data Sharing platform (CADS), which is an “annotate-as-you create” infrastructure that facilitates fielded data annotation. Its main goal is that it reduces the cost of creating annotations and also it is accessible by larger number of queries. Here also we introduce a User Expectation form in the case of Annotations with different meaning in different situation. Actually it is time consuming but while searching sun in Google and get user expected result (micro system, solar system, Newspaper) by taking some minutes it is compensated by the above user expectation tracking.

**Keywords:** Document Annotations, Query workload, Content value, user expectation tracking, Collaborative Adaptive Data Sharing (CADS).

## 1. Introduction

Most of Application Domains uses various techniques for content management and sharing. Most of the data sharing tools uses the annotations for content management but the application domains use the previous annotations for this. The CIRCLE project [1], [2] uses search content item a secure ID and understands content metadata, but it has one disadvantage uses namespaces to uniquely identify types (classes), properties, and resources. Thus, you must have a solid understanding of namespaces. Content management products. Microsoft SharePoint and SAP Net Weaver allow users to share documents, annotate them, and perform simple keyword queries. but the Hard-coded attributes can be added to specialized insertion forms..All the above projects are defeated by ineffective annotation and also the user effort is very high. Here we introduce CADS that improves the platforms by learning the user information demand and adjusting the insertion forms accordingly. At the insertion time annotations are created for each document based on the content value, query workload and user expectation tracking forms. The annotations are created at the insertion time by using Sequential Pattern Matching Algorithm (SPM), it compares all the terms in the document and some terms are taken as annotations based on the content value and query workload. Here also we introduce a User Expectation form in the case of Annotations with different meaning in different situation. Actually it is time consuming but while searching sun in Google and get user expected result (micro system, solar system, Newspaper) by taking some minutes it is compensated by the above user expectation tracking. The remainder of this paper is organized as follows. Section 2 provides related study of previous works. Section 3, reviews proposed system implementation. Section 4 reviews result and performance analysis. Finally, concludes this work in Section 5.

## 2. Related Study

One of the previous work is Managing the content with the CIRCLE. It provides a content-centric storage interface with integrated search and sharing capabilities. CIRCLE gives search content item a secure ID and understands content metadata. The major disadvantage of this scheme is that it uses namespaces to uniquely identify types (classes), properties, and resources. Thus, you must have a clear understanding of namespaces.

Another one is Combining Keyword Search and Forms for Ad Hoc Querying of Databases. It has the advantages of Here uses an approach of using keyword search to lead users to forms for ad hoc querying of databases.

The major disadvantages are hard for users uncomfortable with a formal query language. Keyword query must have schema term(s).

Another one is the Collaborative annotation. There are several system that favor the collaborative annotation of objects and use previous annotations or tags to annotate new objects. The major disadvantages includes Increased user effort and Ineffective annotations.

Another one is Content management products. Microsoft Share point and SAP Net Weaver allow users to share documents, annotate them, and perform simple keyword queries. The major Disadvantages is that Hard-coded attributes can be added to specialized insertion forms. All the existing projects does not provide proper annotations and searching is very difficult and also it needs Increased user effort.

### 3. Proposed System Implementation

#### 3.1 System architecture

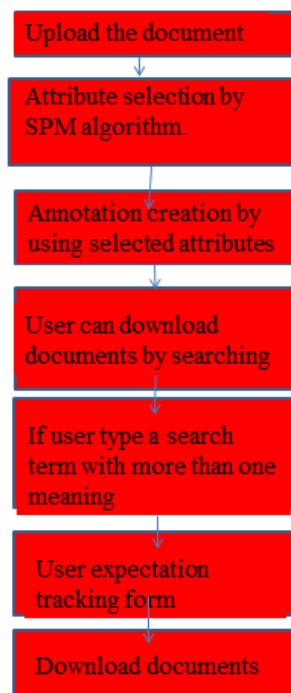


Figure 1: System Architecture

All the previous work use ineffective annotations and also previous annotations are used. CADS improve these platforms by learning the user information demand and adjusting the insertion forms accordingly. It creates the annotations at the insertion time itself.

#### • Document Uploading

Here two login are there admin and user login. Admin has the privilege to upload the document. user registration form is provided. Here user can set up there user name and password and can enter in to site by using the same. It is provided a search button when they click on this, by typing first name itself annotations are displayed. Now user can download pdf document by selecting appropriate annotations.

#### • Attribute selection

Attributes are selected from the document using SPM(sequential pattern matching) algorithm. This algorithm compare each terms in the document and eliminating words like and, is, or, than, over, again etc.

#### • Annotation Creation

Annotations are created based on the selected attribute and query workload obtained from search engine.

#### • User expectation tracking form

If the user wants to search a term having more than one meaning for e.g.: sun(solar system, micro system, Newspaper). By typing such search terms a user expectation tracking form is generated. Here it track the expectation and provide appropriate document to the user.

#### 3.2 System Working

The major steps involved during the design of proposed system are:

**Step 1:** uploading document.

**Step 2:** Attribute selection using SPM algorithm

**Step 3:** Annotation creation by considering content and query value.

**Step 4:** User can download the document by searching.

**Step 5:** User expectation tracking form is generated if the user search term has more than one meaning.

**Step 6:** Then user can download the document based on tge search term.

#### 3.3 Module description

The main modules included in this proposed scheme are: Login, Attribute, selection, Annotation Creation, User expectation tracking form, Download document.

#### • Login

There are two types of login including Admin login and User login. Admin has the privilege to upload the document. User can enter in to the site using user name and password selected by the user in the registration form. Now user can download the document uploaded by the Admin.

#### • Attribute selection

Attributes are selected from the document using sequential pattern matching algorithm. It compare each and every terms in the documents eliminating ‘and, or, this, for etc’ and select appropriative attribute.

#### • Annotation creation

Then annotations are created based on the selected attribute and query workload (frequently asked queries in the search engine).

#### • User expectation tracking form

If the user wants to search a term having more than one meaning,foreg:sun.sun is a News paper, solar system, micro system. By typing such search terms a user expectation tracking form is generated. Here it track the expectation and provide appropriate document to the user.

#### • Download Document

User can download the document uploaded by the Admin using the Annotation.

### 4. Conclusion and Future Work

Compared with existing Annotations, the proposed Document annotation technique reduces the user effort. and provide efficient annotation at the insertion time and leverages the annotations at the query time. Here also we introduce a User Expectation form in the case of Annotations with different meaning in different situation. Actually it is time consuming but while searching sun in Google and get user expected result (micro system, solar system, Newspaper) by taking some minutes it is compensated by the above user expectation tracking. The future work of this paper is, in the proposed system it does not annotate images. In future we can do annotation on images.

### 5. Acknowledgment

The author gratefully acknowledge the valuable comments and suggestions of the reviewers, which have improved the presentation and also thankful to the reference paper authors.

And they are especially grateful to Prof. Smitha Miss for her kind help during the review process of this paper.

## References

- [1] A. Doan, R. Ramakrishnan, F. Chen, P. DeRose, Y. Lee, R. McCann, M. Sayyadian, and W. Shen, "Community Information Management," IEEE Data Eng. Bull., vol. 29, no. 1, pp. 64-72, Mar.2006.
- [2] E.Chu, A. Baid, X. Chai, A. Doan, and J. Naughton, "Combining Keyword Search and Forms for Ad Hoc Querying of Databases,"
- [3] Proc.ACM SIGMOD Int'l Conf. Management Data, 2009
- [4] S.R. Jeffery, M.J. Franklin, and A.Y. Halevy, "Pay-as-You-Go User Feedback for Dataspace Systems," Proc. ACM SIGMOD Int'l Conf. Management Data, 2008.

## Author Profile



**Alfia A P** received degree in Computer Science & Engineering from the Odaiyappa college of Engineering and Technology, Theni, Tamil nadu in 2013. She is now doing M.Tech in Computer Science & Engineering.



**Chashu mol R** received degree in Computer Science & Engineering from VINS Christian college of Engineering, Nagercoil, Tamil nadu in 2012. She is now doing M.Tech in Computer Science & Engineering

