Annotation: an Investigative Survey of Annotation Types and Systems

Sonali T. Kadam, Sanchika Bajpai, ¸Prashant M.Yelmar

Abstract - **Annotation plays a important role in our daily life and study. Mostly documents exist in digital format on the web, people spends a large amount of their time on searching the web browsers for looking useful information. This is the unidirectional interaction with user.Annotations grow geometrically because of the reflections on documents shown by different writers and with their writing time. Firstly we have introduced the history of the annotation then some research related work done in the annotation results. Annotation can be done for the Web, java, pdf, text, xps (XML Paper Specification), mobile, image, multimedia etc. We focus on the survey to see the use of annotation in different areas along with the usage in different scenarios. In Information research for decision making and integrating an annotation database can be founded on the parameters such as document, user and time. Website which supports annotation systems, provide user friendly interfaces, easy-to-use structural and layout annotation functions.Web database helps to generate query result pages or search result records based on a user's query. Automatic extraction of the data from such query result pages is very important for different applications, such as data integration, meta querying cooperates with multiple web databases.Researchers have shown the many approaches to develop and implement web annotation systems.Our attention is toanalyse the annotation and need in web databases to improve a search result research.**

Keywords: **Information search, Information retrieval, Annotation, Metadata**

Sonali T. Kadam

Department of Computer Engineering, Bhivarabai Sawant Institute of Technology and Research ,University of Pune, India.

Sanchika Bajpai

Department of Computer Engineering, Bhivarabai Sawant Institute of Technology and Research ,University of Pune, India.

Prashant M.Yelmar

Department of Information Technology, MIT College of Engineering ,University of Pune, India.

I. Introduction

The Internet gives a huge amount of useful information which is usually formatted for different users, which makes it difficult to extract relevant data from various sources. So, the availability of robust, flexible information retrieval systems that translates the Web pages from web databases into userfriendly structures such as a relational database,structured data becomes a need. Information retrieval includes the representation, storage, organization of, and access to information items such as documents, Web pages, online catalogs, structured and semi-structured records, multimedia objects.

The representation and organization of the information items should provide the users with easy access to information of their own interest. As a human point of view, information retrieval deals with mainly to study the behaviour of the user, understanding their main requirements, anddetermines the effects of organization and operation of the information retrieval system. Recent advanced changes in IR to research efforts conducted by pioneers such as Hans Peter Luhn, Eugene Garfield, Philip Bagley, and CalvinMoors, this last one having allegedly coined the term information retrieval (1692) [1].
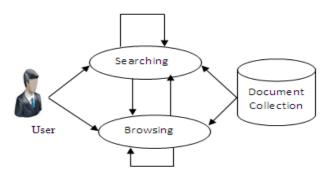

Fig 1 The tasks of the user.

As a result, millions of developers have created billions of documents,web pages, books, multimedia object that made the largest human repository of knowledge. An immediate response is that searchinguseful information on the web which is difficult task and obviously requires a search query to run a search engine. When user query is given to search engine, the

key goal of the IRsystem is to retrieve information that is useful or relevant to the user and it is not as structured and maybe semantically ambiguous. In traditional IR, the important role of user was well-defined in that a user formed a query, viewed the results,offeredfeedback,reformation of query and repetition.IR process is shown in Fig.1.

An increasing number of databases are becoming web accessible through search interfaces. A standard technique for improving the effectiveness of information retrieval is to combine the various text representations and search strategies.The DARPATIPSTER project (Harman, 1992) and TREC evaluations (Harman, 1995) was introduced the idea of information integration of different search systems.Combining representations, retrieval algorithms, queries, and search systems produces better effectiveness than a single system.The view of combinationis similar to the inference net model which is a new approach to retrieval based on language models that supports integrated combination as shown in[2].

We have noticed that while studying or reading text, reader needs more information or requires the text to be better explained and directly assigned to this word. So our goal is to study the annotation through which it will be possible to automatically add definitions or links to related pages to keywords in text of web pages.It will give the motivations for mobileannotation, multimediaannotation,web annotation, text annotation, JAVA annotation, IT based annotation, image annotation, multimedia annotation (Audio , video , animation, noisy) [3]. If we discuss in brief,for instance ,is an online communication tool that allows an personally to post comments. In addition, users can share, text chat and opinions through discussion board. When feedback need to be pointed to specific part of documents discussion may not be effective. Blog and wiki has a great influence way of online collaboration and sharing. Hybrid mechanism that allows annotations to provide the benefits of standardannotations to increase the interaction between users.

Annotation is an interesting research issue since the invention technologies such as HTML, XML, Wiki, JAVA, IT based,content-based image retrieval (CBIR).etc. In this paper, we havedefined theannotations as:

"Online annotations associated with web resources such as web pages, with which users can add, update or delete a text, image,comment which include highlights or underlining, footnotes, tags, and links from a web page without modifying the page itself". We are going to study the core areas of annotation that includes: interaction between users and the web, annotation and annotation systems, system requirements, limits of existing tools, related technologies, and areas where annotations would help.

This paper provides the contribution in evolution of annotation, different annotation systems; current and future services in annotation and collaboration, review and comparison of existing annotation systems.Further paper will also introduce some of the currently existed annotation systems.

II.   Related Work

The survey includes several sections each representing an area of research and application in annotation. Important areas to be discussed are the annotations, existing types, systemsand their comparison, limits of systems, related technologies. Each section is demonstrated as follows.

A. Annotations

An annotation is a comment, explanation, presentational mark-up i.e. metadata attached to text, image, or other data [4].Annotation can be conceptualized to assistin the field of information systems and decision making and database is founded on the three parameters of document(different document requires different annotation), user(multi users cannot make the same type of annotation on the same document) and time(at the same time a specific user may not annotate the same document) [5].Annotation contents are easily under stable to readers [6], it may have personal annotation[7], different forms such as adding a text or note, image, underlining, colouring, highlighting, commentary, linking [8] with additional functionalities like reading ,editing, linking and sharing.Annotation system consists of three modules: the first is to view existing annotations, the second to create new annotations, and the third to store the annotations [9].

B. Existing Types Of Annotation And Their Comparative Study

Following information gives the comparison of the existing annotation types and systems in terms of their meaning, technologies used, supported functions/features.

1) *Web Annotation*: It is an online annotation deals with a web resource in which a user can add, update or delete information from a web resource without modifying the resource itself in which existing systems are Firefox (built-in),Diigo, Keeppy, IBMmarkup, CoNote and approach or methods or technologies are ontology, REODRUNNER ,TMSUMS, wrapping approach ,annotation approach , SemTagetc[4] .Data is aligned by using data alignment algorithm[12], then in annotation phase annotators are used to create annotations and after that wrapper induction systems are used to generate an wrappers. Web annotation is very much important in deep web database collection as well as in online internet shopping.

2) *Java Annotation*: It is a form of syntactic metadata that can be added to Java Classes, methods, variables, parameters and packages and It is possible to create meta-annotations in thepresenting ones in Javae.g. built in annotation and custom annotation [4],[13],[14],[15],[16]. Annotations can be read from source files, class files, or reflectively at run time.

3) *Image Annotation*: Image annotationor automatic image tagging is the process by which a system automatically assigns metadata in the form of caption or keywords to a digital image used in image retrieval systems to organize and locate images of interest from a database. In this annotations machine leaning techniques are used   with the help of training annotation which are used to translate the textual vocabulary with the visual vocabulary or clustered regions[4],[17],[18]. Image annotation becomes widelyimportant for clinical applications and medical research. Medical image annotation includes metadata annotation as well as multimedia annotation, such as graphical annotation [19].

4) *Text Annotation*:Text annotation is the result of adding a note to a text, which may include highlights or underlining, comments, footnotes, tags, and links.Text annotations have the functions for both private and public reading and communication practices which includes four major functions such as facilitat[ing] reading and later writing tasks, eavesdrop[ping] on the insights of other readers, provid[ing] feedback to writers or promote communication with collaborators, call[ing] attention to topics and important passages. [4]. Text annotations can include notes written for a reader's personal (annotation)[7] purposes, as well as shared annotations written for the purposes of collaborative writing and editing, commentary, or social reading and sharing. Text annotations are also known as marginalia, though the meaning of this term specifically for hand-written notes made in the margins of books.

5) *PDF Annotation*: For this annotation iAnnotate PDF (from Branchfire) and GoodReader (from Aji) allow annotation of PDFs as well as exporting summaries of the annotations. To already created PDF filesXournal , Qiqqa , PDF-XChange Viewer , Apple's Mac OS X'ssoftware's allows the user to annotate, highlight, and add notes [4].

6) *XML Paper Specification (XPS) Annotation*
:This annotator is a document management software program for Microsoft Windows that allows users to view, annotate, convert, sign and print XPS documents [4],[20],[21].

7) *IT-based Annotation***:** These systems are consists of standalone and client-server systems which utilize footnote interfaces, interlinear annotations, voice annotations, digital-ink based interfaces. Areas of educational research investigate the benefits of socially shared text annotations for IT -based annotation sharing in collaborative learning, critical reading , argumentative writing, critical thinking[4].

8) *Multimedia Annotation*:In video annotation [21], the number of video clips and hours of video footage consumed and produced has been continuously growing with the self-broadcasting on-line services like YouTube. Only some of video players offer annotation features those are suitable for research, teaching or learning. Annotation capabilities in the popular video players are launched for marketing purposes, which includes advertisements or for social commentaries. Collaborative annotation in management environmentconsider the management of sources of information as multimedia information management.Multimedia describethe documents which include text, audio, video and images. In area of multimedia work done search includes NIX (NASA Image eXchange2) [22].This search enable users to search in NASA's online image and photo collections over the web.

III. Advantages of Annotation Systems

Metadata can supply along with data. Annotations have a lot of advantages such as follows.
1. *Static Type Checking*- the compiler will check for you where the annotation is applicable and how.
2. *Clean Code*- its much easier to see the meta data defined in annotations.
3. *Readabilit*y - For example in JPA configuration its much more cleaner to declare new entities by Annotations instead of hbm.xml files. These things change only in development stage so there is no problem with recompiling code. When we use xml files we have to often open both- entity file to make changes.
4. *Flexibility* - In XML files we have to write all configs in "only one proper way". It is advantage and disadvantage at the same time.
5. *Length* - XML-based configs are often very long . Annotations are much simpler to use.
6. *Active Process* - Annotating is an active process. It's easy to become inattentive when simplyhighlighting.
7. *Ideas* - Annotating helps you learn to pick out main ideas.
8. *Easy to Understand* - Annotating helps user to understand.
9. *Easy to Remember* - Annotating helps user remember.
10. *Additional Information* - In web searching and dynamic link generation annotations provide the third party, relative metadata about the contents of web page to get additional information.

IV. Conclusion

Customer requirements have far outgrown the initially envisaged usage of the internet.Web infrastructure is missing an important feature, such as hypermedia (e.g., structuring and annotation).To review the previous work on annotation we employed a information research, its   need ,annotation and implementation approach.

We studied that, till the internet incorporates the building blocks like hypermedia,comments,notes, images, there is much more need of research to bridge the gap between customers expectations and web databases. Significant and satisfactory research efforts remain to beinvested, to the open

deep web database issues and interface designs for collaboration using annotation.

From this paper, we have beenable to show, how an annotation is a great need of the current situation to enhance informationresearch with different types of annotations.Annotations can improve information retrieval by using automatic extraction of data records .Our annotation survey is for internal, educational, IT basedas well as social use and our experience will inform some of these investigations for web annotation, java annotation , image annotation and multimedia annotation in the future work.

Acknowledgment

The research presented in this paper would not have been possible without the support and advises of professors of Bhivarabai Sawant Institute of Technology and Research, S.B. Patil College of Engineering and MIT College of Engineering for helpful discussions about this work.

References

[1] G. Salton and M. McGill, Introduction to Modern Information Retrieval. McGraw-Hill, 1983.

[2]W. Bruce Croft, "Combining Approaches for Information Retrieval," *Advances in Information Retrieval: Recent Research from theCenter for Intelligent Information Retrieval*, Kluwer Academic, 2000.

[3]PeiwenZhu,"Web Annotation Systems: A Literature Review and Case Study ",A Master's Paper for the M.S. in I.S. degree ,Chapel Hill, North Carolina pages 35,April, 2008.

[4]http://en.wikipedia.org/wiki/Annotation

[5]R. Charles and D. Amos,"Annotation and its application toinformation research in economic intelligence",*In Advances inKnowledge Organization*, volume 10, pages 35.39, Wuzburg,Germany, 2006. Getersloh, Germany, ErgonVerlag.

[6]O'Hara, K., &Sellen, A." A comparison of reading paper and online documents", *CHI97 Conference Proceedings, ACM Press*, Atlanta, GA,1997.

[7]Fu, Xin and Ciszek, Tom and Marchionini, Gary and Solomon, Paul,"Annotating the Web: An Exploratory Study of Web Users' Needs for Personal Annotation Tools", *(ASIST)*, Charlotte (US), 2005.

[8] Marshall, C. "Toward an ecology of hypertext annotation", *Proceedings of ACM Hypertext '98*, PA ,pages 40-49 , 1998.

[9]Denoue, L., &Vignollet, L."An annotation tool for web browsers and its applications to information retrieval", *Proceedings of RIAO200*, Apr. 2000.

[10] Masashi Inoue, "On the need for annotation based image retrieval",*Proceedings of the workshop on information retrieval* , 2004.

[11] R.Guha and R. McCool. Tap: Towards a web of data.http://tap.stanford.edu/.

[12] Yiyao Lu, Hai He, Hongkun Zhao, WeiyiMeng, and Clement Yu,"Annotating Search Results from Web Databases" *IEEE Trans. Knowledge and Data Eng*., vol. 25, no. 3, pp. 514-527, Mar. 2013.

[13]http://docs.oracle.com/javase/1.5.0/docs/guide/language/annotations.html

[14]http://www.25hoursaday.com/CsharpVsJava.html#attributes

[15]http://www.javabeat.net/annotations-in-java-5-0/

[16] Vincenzo Gervasi, Giacomo A. Galilei, "Software Manipulation with Annotations in Java", *Advances in Software Engineering,LNCS 5316*, , Springer,pp 161-148,2008.

[17]S Feng, R Manmatha, and V Lavrenko"Multiple Bernoulli relevance models for image and video annotation". *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1002–1009,2004.

[18]G Carneiro, A B Chan, P Moreno, and N Vasconcelos ,"Supervised Learning of Semantic Classes for Image Annotation and Retrieval". *IEEE Trans. on Pattern Analysis and Machine Intelligence*. pp. 394–410,2006.

[19]Fusheng Wang ,Rabsch, C. , Peiya Liu, "Native Web Browser Enabled SVG-based Collaborative Multimedia Annotation for Medical Images" ,*Data Engineering*, Pages 1219 - 1228 ,ICDE 2008[20]http://www.xpsdev.com/xps-annotator

[21]http://www.annotations.harvard.edu/icb/icb.do?keyword=k80243&pageid=icb.page466612.

[22]Charles Robert, Frederic Andres, Kim V eltman ,'' Advances in Collaborative Annotation in Semantic Management Environment", 2nd International conference on Digital Information Management(ICDIM'07),Vol.1,pages 336-344 ,2007.

About Author (s):

Sonali T. Kadam completed her B.E. degree in Computer Engineering from VPCOE, University of Pune in 2009. She is pursuing M.E. in Computer Engineering from BSIOTR, University of Pune, Maharashtra, India. Her research interest includes Knowledge and Data Engineering, Database Management System and Data Mining.

Sanchika Bajpai is working as Assistant Professor at JSPM's BSIOTR, Pune, Maharashtra, India. She completed her B.Tech degree in IT from Chaudhary Charan Singh University in 2009 and M.Tech. in Computer Science from Amity University, Lucknow. Her research interest includes Software Testing and Database Management System.

Prashant M. Yelmar is working as Assistant Professor at MAEER'S MIT College of Engineering, Pune, Maharashtra, India. He completed his BE in Information Technology from Mumbai University and M. Tech in Computer Science and Engineering from Walchand college of Engineering Sangli. His research interest includes Data Mining, Information Retrieval, Geographic Information Systems, Social Networks, and Soft Computing.