How to Organize the Annotation Systems in Human-Computer Environment: Study, Classification and Observations

Anis Kalboussi, Nizar Omheni, Omar Mazhoud, Ahmed Hadj Kacem

How to Organize the Annotation Systems in Human-Computer Environment: Study, Classification and Observations

Anis Kalboussi[1], Nizar Omheni[1], Omar Mazhoud[1], Ahmed Hadj Kacem[2]

[1] ReDCAD Research Laboratory, Sfax University and Higher Institute of Computer Science and Management, Kairouan University, Tunisia
{anis.kalboussi, nizar.omheni, o.mazhoud}@isigk.rnu.tn
[2] ReDCAD Research Laboratory and Faculty of Economics and Management, Sfax University, Tunisia
ahmed.hadjkacem@fsegs.rnu.tn

Abstract. The practice of annotation is a secular and omnipresent activity. We find the annotation in several areas such as learning, semantic web, social networks, digital library, bioinformatics, etc. Thus, since the year 1989 and with the emergence of information technology, several annotation systems have been developed in human-computer environment adapted for various contexts and for various roles. These ubiquitous annotation systems allow users to annotate with digital information several electronic resources such as: web pages, text files, databases, images, videos, etc. Even though this topic has already been partially studied by other researchers, the previous works have left some open issues. It concern essentially the lack of how to organize all the developed annotation systems according to formal criteria in order to facilitate to the users the choice of an annotation system in a well-defined context and according to unified requirements. This problem is mainly due to the fact that annotation systems have only been developed for specific purposes. As a result, there is only a fragmentary picture of these annotation tools in the literature. The aim of this article is to provide a unified and integrated picture of all the annotation systems in human-computer environment. Therefore, we present a classification of sixty annotation tools developed by industry and academia during the last twenty-five years. This organization of annotation tools is built on the basis of five generic criteria. Observations and discussion of open issues conclude this survey.

Keywords: Annotation system, metadata, annotation, tag, Human-Computer environment, classification, survey.

1 Introduction

Digital annotation presents an activity central in many important tasks, such as studying, indexing and retrieving. It is an important modality of interaction with digital objects (web pages, text files, programs, etc.). Annotations cover a very broad spectrum, because they range from explaining and enriching an information resource with personal observations to transmitting and sharing ideas and knowledge on a

subject [10]. They can be geared not only to the individual's way of working and to a given method of study, but also to a way of doing research [18]. Moreover, they may cover different scopes and have different kinds of annotative context: they can be private, shared, or public, according to the type of intellectual work that is being carried out [39]. Furthermore, annotations are not only a way of explaining and enriching an information resource with personal observations, but also a means of transmitting and sharing ideas to improve collaborative work practices [29]. Thus, the possibility of enriching digital contents by adding annotations has attracted many researchers, who have looked at this opportunity from many different perspectives and with a number of purposes in mind [9].

Therefore, since the year 1989, several annotation systems have been developed in human-computer environment adapted for various contexts and for various roles. The list of annotation tools is growing every year. The researchers have mostly focused on providing more and more sophisticated interfaces with several annotation functionalities. These ubiquitous annotation systems allow users to annotate with digital information several electronic resources such as: web pages, text files, databases, images, videos, etc. The proposed annotation tools have been adopted in a variety of different contexts, such as content enrichment, collaborative and learning applications, and social networks, as well as in various information management systems, such the semantic web, digital libraries, and databases [11, 20].

Even though the annotation system is so common and familiar, it turns out to be especially elusive when it comes to explicitly and formally classifying it, mainly because it is applied in very diverse areas. Indeed, we usually derive what an annotation is from the particular task to which it is applied, rather than investigating the annotation system by itself in order to understand its features and how to use it. To cite a few examples, if we deal with the semantic web, annotation systems are considered as makers of metadata [2]; in the field of digital libraries annotation tools are treated as a means of appending additional content [25]; when we talk about databases, annotation systems represent both provenance information about the managed data and a way of creating the database itself [1]; finally, in social networks and collaborative tagging, annotation systems are tools of representing of tags or keywords on different kinds of digital content, e.g. photos, videos or bookmarks [9]. This flourishing of different viewpoints about the annotation systems, which is usually considered as an enrichment of the community of annotation and metadata, reveals a lack of a clear comprehension of how to organize all the developed annotation systems of the literature according to formal criteria in order to facilitate to the users the choice of an annotation system in a well-defined context and according to unified requirements. As a result, there is only a fragmentary picture of these annotation tools with its classifications and features.

The aim of the article is to provide a unified and integrated picture of annotation systems in human-computer environment. This panoramic view is based on a classification of sixty (60) annotation systems developed in the literature during the last twenty five years by industry and academia. This organization of annotation tools is built on the basis of five generic criteria: *annotation type* (computational / cognitive); *category of annotation system* (application/ plugin/ web site*); type of annotative activity* (manual/ semi-automatic/ automatic); *annotated resource type* (text/ web page/ video/ audio/ image/ database/ web service/ Doc/ HTML/ PDFs) and

application domain (semantic web / social networks/ digital library/ learning/ databases/ web services/ computational linguistics/ bioinformatics).

The rest of this article is organized as follows: Section 2 presents a general presentation of annotation systems and a classification of these tools according several criteria; Section 3 draws some key observations and a discussion of open research issues about the annotations systems. Finally Section 4 concludes this survey.

2   Annotation Systems

2.1   Definition of annotation systems

Annotation systems have been developed since the 90th to transpose on electronic document the secular practice of annotation. Then, these systems have gradually taken advantage of processing capabilities and communication of modern computers to enrich the practice of electronic annotation [20]. An annotation system or still called annotation tool or "annoteur", is a system allowing users to annotate various types of electronic resources with different kind of annotations [29]. Many researchers have been interested in the creation of annotation tools to facilitate the annotation practice of the annotator in digital environment. We can find numerous commercial software and research prototypes created to annotate electronic resources (see Figure 1).

Fig. 1. Several annotation systems are developed in different contexts.

2.2   Classification of annotation systems

Annotation systems have been popular among both researchers and practitioners. In the literature the number of annotation systems does not stop increasing every year. The researchers have mostly focused on providing more and more sophisticated interfaces with several annotation functionalities. There are various areas of

continuing research and implementation on annotation systems. Researchers have taken different approaches to develop and implement these applications. However, this diversification about these approaches, which is usually considered as an enrichment of the community of annotation and metadata, reveals a lack of a clear comprehension of how to organize all the developed annotation systems of the literature according to formal criteria in order to facilitate to the users the choice of an annotation system in a well-defined context and according to unified requirements. As a result, there is only a fragmentary picture of these annotation tools with its classifications and features.

To unify and integrate the picture of annotation systems in digital environment, we present a classification of sixty annotation tools developed in the literature during the last twenty five years by industry and academia. This organization of annotation tools is built on the basis of five criteria (see Table 1). Each annotation system should focus on a particular *type of annotation*. In each annotation type corresponding to a set of appropriate *application areas* in which the annotation system is developed. Afterward, an annotation system in a particular application domain can be one of *three categories: (application / website / plugin)*. For each category, the annotation system is necessarily based on a process of *annotative activity (manual / semi-automatic / automatic)* and annotates a particular digital *resource type*.

Table 1. Criteria of annotation systems classification

CRITERIA	SPECIFICATION	
Annotation type	Cognitive	Computational
Application domain	Digital library Learning Social networks	Semantic web Databases Computational-linguistics Web services Bioinformatics
Annotation system category	Application – Website – Plugin	
Annotative activity type	Manual – Semi-automatic – Automatic	
Annotated resource type	Text / PDF/ HTML/ Doc/ Table/ Web service/ Sound / Image/ Video/ Database	

2.2.1. First criterion: Annotation type (computational / cognitive)

- *Computational annotation (metadata):* Annotation can be considered as metadata, that is, additional data which relate to an existing content and clarify its properties and semantics [6]. With this aim, annotations have to conform to some specifications that define the structure, the semantic, the syntax, and even the values that annotations can assume. Computational annotation is used to annotate resources with metadata to facilitate their use by software agents. It is used in the field of information retrieval, summarization, document classification, indexing etc. Computational annotation is applicable to any type of resource: web pages, text files, databases and even for images and videos. The recipients of this kind of annotation are both people and computing devices. On the one hand, metadata can benefit people because, if they are expressed in a human-readable form and

their format and fields are known, they can be read by people and used to obtain useful and well-structured information about an existing content. On the other hand, computational annotations offer computing devices the means for automatically processing the annotated contents. The computational annotations cover essentially the domains of semantic web [2], databases [42], computational linguistics [15], web services [40] and bioinformatics [35].

- *Cognitive annotation (track of reading):* This is an annotation that is processed and manipulated by human agents. This category of annotation requires a cognitive and intellectual effort to be interpreted. It is visible and distinguishable on the document. The annotation is considered as a track of the activity of reading, which means that the reader is active and productive. It is the result of the investment of the reader in the consulted document. Cognitive annotations are regarded as additional content that relates to an existing content, meaning that they increase the existing content by providing an additional layer of elucidation and explanation. However, this elucidation does not happen, as in the case of annotations as metadata, by means of some kind of constrained or formal description of the semantic of the annotated object [6]. On the contrary, the explanation itself takes the shape of an additional content that can help people understand the annotated content. However, the semantic of the additional content may be no more explicit for a computing device than the semantic of the annotated content. Therefore, the final recipients of this kind of annotation are people; because a cognitive annotation does not make the annotated object more readily processable by a computer than the same object without annotations. The cognitive annotations cover essentially the domains of learning [3], digital library [28] and social networks [10].

2.2.2. Second criterion: Application domains

As we saw previously, the electronic annotation can be organized in two main categories: computational annotations which cover essentially the domains of Semantic web, Databases, Computational linguistics, Web services and Bioinformatics; cognitive annotation which stretch essentially the domains of Learning, Digital library and Social networks.

A relevant example of the use of computational annotation is provided by the *Semantic web* [33] initiative promoted by the *World Wide Web Consortium (W3C)*, which aims at enhancing human-understandable data, namely Web pages, with computer-understandable data, namely metadata, so that information is given well-defined meaning, better enabling computing devices and people to work in cooperation. The process of adding metadata to Web pages is called semantic annotation, because it involves the decoration of existing data, such as a piece of text whose content is understandable only to people, with semantic metadata that describe that piece of text, so that computers can process it and thus in turn, offer automation, integration, and reuse of data across various applications [1]. The annotations are represented using W3C recommended standard languages as syntax for describing and exchanging metadata like *RDF (Resource Description Framework)* or *OWL (Web Ontology Language)*. The Semantic Web proposes annotating document content using semantic information from domain ontologies. The result is Web pages with machine

interpretable mark-up that provide the source material with which agents and Semantic Web services operate.

A broader example of the use of annotations as metadata is provided by the *Computational linguistic* domain, wherein linguistic annotation covers any descriptive or analytic notations applied to raw language data. The basic data may be in the form of time functions, audio, video and/or physiological recordings or it may be textual [15]. The added notations may include transcriptions of all sorts (from phonetic features to discourse structures), part of speech and sense tagging, syntactic analysis, reference annotation, and so on. Corpus annotation, sometimes called tagging, can be broadly conceptualized as the process of enriching a corpus by adding linguistic and other information, inserted by humans or machines (or a combination of them) in service of a theoretical or practical goal.

A relevant example of the use of cognitive annotation is presented in the context of a *Digital library (DL)*. Thus, the creation of new information resources is supported by annotations in two ways. First, when users add annotations to existing information resources, these are new information resources themselves [25]. Second, annotations can also assist in the creation of new information resources. Through annotations, new ideas and concepts can be discussed and the results of such a discussion can then be integrated into the newly created object [1].

Cognitive annotations are also used in the context of *Social networks* wherein social tagging on online portals has become a trend now. It has emerged as one of the best ways of associating metadata with web objects. Social networks have become a popular medium for people to communicate and distribute ideas, content, news and advertisements. Also, social content annotation has naturally emerged as a method of annotating, categorization and filtering of online information [10]. With the increase in the kinds of web objects becoming available, collaborative tagging of such objects is also developing along new dimensions. Social tagging became popular with the launch of sites like Delicious and Flickr [43]. Since then, different social systems have been built that support tagging of a variety of resources. On blogging sites like Blogger, Word-press, Live-journal, blog authors can add tags to their posts. On micro-blogging sites like Twitter [8], hash-tags are used within the tweet text itself. On social networking sites like Facebook, Orkut, etc., users often annotate parts of the photos. Users can also provide tagging information in other forms like marking something as 'Like' on Facebook [10]. Social news sites like Digg and Slash-Dot allow users to attach tags to news stories. Yelp, CitySearch and other sites allow users to attach their reviews and other users to select tags to rate reviews too. Multimedia objects like podcasts, live casts, videos and music can also be tagged on sites like YouTube, Imeem, Metacafe, etc [16].

As a final example of the use of cognitive annotations, we present the context of *Learning*. During this process, the learner's activities are numerous and especially varied [3, 22]. Thus, the learner can choose to read, write, listen, discuss, experiment, or annotate various resources to achieve his learning goals. Among these activities the annotation practice is very common and omnipresent because while reading, the learner usually uses comments, highlights, circles sections and posts it to annotate the consulted resources [30]. Annotations give students the chance to summarize new ideas while receiving peer support. Tagging can be considered as an action of reflection, where the tagger can summarize a series of thoughts into one or more

annotations, each of which stands on its own to describe some aspect of the resources based on the tagger's experiences and beliefs [25].

2.2.3. Third criterion: Category of annotation system (application / plugin / web site)

Several annotation systems focus on developing an advanced architecture and building a user-friendly annotation interface to improve annotations of various electronic resources in digital environment. These annotation systems can be classified in three main categories (see Figure 2):
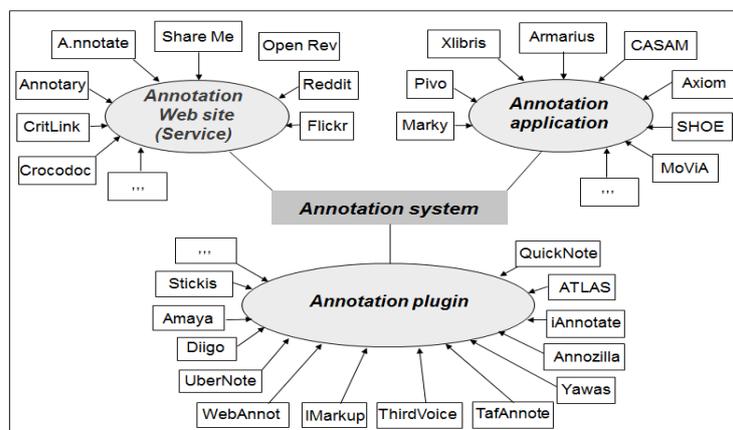

Fig. 2. Category of annotation system (Application/ Plugin/ Website).

- *Annotation application:* It is a whole application that allows its users to annotate the resources consulted via the web. Annotation application offers several features and allows multiple browsers to view and annotate web pages. XLibris [26], PDF Annotator [13], InsightNotes [42] and CASAM [14] are examples of annotation application.
- *Annotation plugin:* It is a toolbar or also called additional module added within application or web browser to annotate web pages consulted through the browser. Several toolbars exist such as Amaya [33], Keeppy [18], OntoMat [27], New-WebAnnot [21] and ATLAS [36].
- *Annotation web-site (service):* It is a specialized web-site for annotating that offers to its registered users the ability to annotate web resources. We quote in this type of annotation system: Delicious, Flickr and Reddit [38].

2.2.4. Fourth criterion: Type of the annotative activity (manual / semi-automatic/ automatic)

Each annotative activity realized by an annotator using an annotation system passes through two complementary processes: choose the anchor and the shape of the annotation in a given resource (*Process 1*) and specify the properties of the annotation

(*Process 2*). Based on these two processes, we can classify the existing annotation systems into three categories:

- *Manual annotation system:* Each of the two previously mentioned processes is manually executed by the user himself. The creation process is completely chargeable to the annotator, who selects the shape and the anchor of the annotation and then specifies properties to give to this note. This type of annotation system tries to simply reproduce the annotation process on paper towards computer. Consequently, when it is a question of annotating a wide collection of digital documents, the annotative activity becomes heavy for the annotator. CATool [4] and Keeppy [18] are examples of manual annotation system.

- *Semi-automatic annotation system:* In this type the first process is performed by the annotator while the second process is executed by the annotation tool. The annotator begins to annotate manually. Mean-while the tool textually analyzes his annotations and generates rules of annotation. Then, the tool uses these rules to deduct passages potentially notables and create candidate annotations. The user can then validate or not the annotations proposed by the tool. The system uses these validations to correct its rules. From a certain level of improvement, the semi-automatic annotation system can continue the annotation process alone. This type of annotation system is used especially in the context of the Semantic Web. Indeed, the semantic annotation of Web resources is painful and heavy if it is a question of annotating a large number of resources. Therefore, various tools are proposed to overcome this problem and assist the annotator in the annotation process. WebAnnot [3], Marky [28], New-WebAnnot [22] and Memory-specs [39] are examples of semi-automatic annotation system.

- *Automatic annotation system:* The automatic annotation means that both processes are executed by the annotation system. Thus, according to certain criteria (given by the user or retrieved from another computer system) the annotation tool selects itself the anchor and the shape of the annotation and then specify its relevant properties [23]. One of the most popular automatic annotation system is the search bar which is an extension of Web browsers that allows to highlight with different colors the keywords typed by the user [41]. Share-Me [39], ATLAS [36] and CASAM [14] are examples of automatic annotation system.

2.2.5. Fifth criterion: Type of annotated resources

Through the annotations systems, annotators consult and annotate varied electronic resources. Such a resource is dematerialized in a particular format: Text, Doc, HTML, PDF, Table, Video, Image, Audio, Source code, URL, Web service and Database.

Table 2 highlights a set of sixty (60) annotation systems found in the literature review classified according to the five (5) criteria presented above. The tools are presented according to the chronological order their publication or update year (from 1989 until 2014), what allows to evaluate the gradual improvements made to these annotation tools. In the case where the annotation system has several versions, we take the recent version to follow the updates in each annotation system.

Table 2. Classification of annotation systems based on five criteria

Name of annotation system	Year	Annotation type (Cognitive / Computational)		Application domain	Annotated resource type	Category of annotation system (Plugin / Site Web / Application)			Type of annotative activity (Manual / Semi-automatic / Automatic)		
		Cognitive	Computational			Plugin	Site Web	Application	Manual	Semi-automatic	Automatic
InterNote	1989	✓		Digital Library	Text			✓	✓		
Mosaic (version 2.0)	1993	✓		Web	HTML		✓		✓		
ComMentor	1994	✓		Digital Library	HTML		✓		✓		
Re:mark	1996	✓		Learning	PDF			✓	✓		
Third Voice	1999	✓		Web	HTML			✓	✓		
AnnoteImage	2000	✓		Bioinformatics	Image			✓	✓		
Ann. Sys. for Sem. Web	2002		✓	Semantic Web	Text			✓		✓	
NCST	2002		✓	Semantic Web	HTML			✓	✓		
IPSA	2003		✓	Digital Library	Image			✓	✓		
UCAT	2003	✓		Learning	HTML			✓	✓		
AnT&Cow	2005		✓	Semantic Web	HTML			✓	✓		
SMORE (version 2.0b)	2006		✓	Semantic Web	Image			✓	✓		
WSMO Studio	2007		✓	Web services	Web service			✓	✓		
Armarius	2008	✓		Digital Library	Image			✓	✓		
Filtered-Push	2009		✓	Digital Library	Database			✓		✓	
@Note	2009		✓	Bioinformatics	Text			✓			✓
eduKEN	2010		✓	Semantic Web	Video			✓	✓		
Diigo	2011		✓	Social Networks	HTML			✓	✓		
Pivo	2011		✓	Semantic Web	Video, Image			✓	✓		

Name of annotation system	Year	Annotation type		Application domain	Annotated resource type	Category of annotation system			Type of annotative activity		
		Cognitive	Computational			Plugin	Site Web	Application	Manual	Semi-automatic	Automatic
AlvisAE	2012		✓	Comp Linguistics	Text	✓			✓		
CAT	2012		✓	Learning	All types	✓			✓		
eHOST	2012		✓	Comp Linguistics	Text	✓			✓	✓	
Surfing Notes	2012		✓	Learning	HTML	✓			✓		✓
WikiLayer	2012		✓	Digital Library	HTML	✓			✓		
BioAnnote	2013		✓	Bioinformatics	Doc			✓	✓	✓	
CATool	2013		✓	Learning	Video		✓		✓		
CLAS	2013		✓	Learning	Video	✓			✓	✓	
Indicoder	2013		✓	Digital Library	Doc, HTML	✓			✓		
In Situ	2013		✓	Semantic Web	Text		✓		✓		✓
Keeppy (ver 2.4.6)	2013		✓	Web pages	HTML	✓		✓	✓	✓	
MeDetect	2013		✓	Bioinformatics	HTML			✓	✓	✓	
Share Me	2013		✓	Semantic Web	Doc		✓		✓		
WebAnnot	2013		✓	Learning	HTML, Doc	✓			✓		
ATLAS	2014		✓	Semantic Web	Video	✓					✓
Axiom	2014		✓	Digital Library	Doc, PDF		✓	✓	✓	✓	
CASAM	2014		✓	Semantic Web	Video		✓			✓	
Crocodoc	2014		✓	Learning	Doc, PDF		✓		✓		
Domeo Annotation	2014	✓		Bioinformatics	HTML	✓		✓	✓		
ndNotes	2014		✓	Digital Library	Doc, PDF	✓				✓	

Name of annotation system	Year	Annotation type		Application domain	Annotated resource type	Category of annotation system			Type of annotative activity		
		Cognitive	Computational			Plugin	Site Web	Application	Manual	Semi-automatic	Automatic
Facebook (ver 2014)	2014		✓	Social Networks	All types	✓			✓		
Flickr (ver 2014)	2014		✓	Social Networks	Image, Video	✓			✓		
Framework ASVA	2014		✓	Semantic Web	Video	✓					✓
Ink Annotation Framework	2014		✓	Digital Library	Doc			✓	✓		
InsightNotes	2014		✓	Databases	Database			✓	✓		
Marky (ver 2.3)	2014		✓	Bioinformatics	All types			✓	✓		
Memory specs	2014		✓	Semantic Web	Text			✓	✓		
New-WebAnnot	2014		✓	Learning	HTML		✓		✓		
Open Rev	2014		✓	Digital Library	PDF		✓		✓		
Org-mode (vers 8.2.10)	2014		✓	Semantic Web	Text			✓	✓		
Prostetnic highlighter (v 2.5)	2014		✓	Web	HTML	✓			✓		
QuickFox-Notes (ver 2.8.5)	2014		✓	Web	HTML	✓			✓		
QuickNote (ver 0.7.5)	2014		✓	Web	HTML	✓			✓		
rbutr	2014		✓	Semantic Web	URL	✓			✓		
Reddit	2014		✓	Semantic Web	URL		✓		✓		
Twitter (ver 2014)	2014		✓	Social Networks	Text		✓		✓		
UberNote (ver 2.0)	2014		✓	Web pages	HTML			✓	✓		
VideoANT (ver 3.0.0)	2014	✓		Semantic Web	Video			✓		✓	
Word Microsoft (ver 2014)	2014		✓	All domains	Doc			✓	✓		
YouTube (ver 3.0.0)	2014		✓	Semantic Web	Video		✓		✓		

3 Observations, limitations and open questions

From the study and the organization of the annotation systems in Human-Computer environment, we can synthesize some observations, limitations and open questions presented as follow:

- *Synthesis 1*

Generally speaking, the term annotation refers to a piece of data associated to another piece of data. Annotations are used in a wide-variety of systems such as blogs, social networks, databases or wikis. Annotations can be defined on every identified resource such as documents, data in a file or in a database, images or video. Annotations can be defined at different level of granularity. For example, in document management systems, annotations can be attached from the whole document to the word level. Annotations can be set manually i.e., made by a user, can be semi-automatic i.e., based on suggestions or fully automated. Annotations can be associated to a group of users (experts, novices, etc.) and can be shared within the group or with other groups. A lot of attempts have been made in the research of annotation systems, many problems, both technical and non-technical, still exist to keep these systems from successful and widely adopted. Marshall [25] argued that "*relatively few researchers are dedicated to understanding what the users actually need*". Azouaou et al. [3] argued that "*there is nowadays no widespread annotation service*". Gabriel et al. [10] claimed that "*there is no wide spread use of web annotation systems*". No single approach is available now that supports all of the features that would work with any standard browser [18]. Also, the existing tools have little connection between each other. Some of them require installation; some require users to log in before use. These factors may all become potential burdens for their users. Many researchers identified several limitations in the implementation of current web annotation systems. The most important of these is the lack of a standardized representation for web annotations, which means that current systems are divergent and proprietary, therefore limiting the possibility for third parties to create clients. Furthermore, they discussed some remaining challenges for annotation systems. The first one is users' privacy. With the current annotation systems' architecture and implementation, users' navigational habits as well as the words they like and dislike can be tracked. When annotations are stored in public annotation servers, the shared or private annotations should be protected from the reach of unauthorized users and the owners of web sites should be able to prevent impropriate annotations to appear on their sites. A best example of this is this problem faced Third Voice [24]. Thus, annotation systems should take more account the security of annotations made by their users. Another issue is interoperability. Current existing annotation systems adopt different strategies to represent annotations, and use different ways to save these annotations. For example, XPointer have been proposed for XML documents and been adopted by Amaya, and Yawas [7] adds the occurrence of the selected text. A detailed discussion appears in [1] where the authors propose new ways to represent the annotation anchor. Their proposition not only applies to XML documents, but also to HTML, PostScript and PDF documents.

- *Synthesis 2*

In another view point, the designers must take into consideration more user centered collaborative design during creation of annotations systems. Annotation can potentially become a bottleneck if it is done by knowledge workers with many demands on their time. Since few organizations have the capacity to employ professional annotators it is crucial to provide knowledge workers with easy to use interfaces that simplify the annotation process and place it in the context of their everyday work. A good approach would be a single point of entry interface, so that the environment in which users annotate documents is integrated with the one in which they create, read, share and edit them. Annotation systems design also needs to facilitate collaboration between users, which is a key facet of knowledge work with experts from different fields contributing to and reusing intelligent documents [4, 14]. Other issues for collaboration include implementing systems to control what to share with whom. For example, in a medical context, doctors might share all information about patients among themselves but only share anonym information with planners [35]. This brings us to issues related to trust, provenance and access rights. An Intranet provides a more controlled environment for tracing the provenance of annotations than the wild Web but access policies are a critical issue to organizations which are invariably concerned with confidentiality issues for client and staff data. As far as cooperation is concerned, almost all of the analyzed systems show that annotations have great potential for supporting and improving interaction among users, and even among computing devices. Therefore, there is a need for modeling and offering different scopes of annotations, for example, private, shared, or public, and managing the access rights of various groups of users.

- *Synthesis 3*

Although highly sophisticated annotation systems exist both conceptually as well as technologically, we still observe that their acceptance is somewhat limited on behalf of the annotator. Studies made in the works of [25, 31, 32] show that many readers prefer to print an electronic document and to annotate it on paper instead of annotating it directly on its electronic format using an annotation system. Therefore, the process of marking a paper document with the tools that we find in our environment, a pen, a highlighter, is most preferred by the reader instead of reading a document on a screen and the mark via a software interface that requires us to use the keyboard, mouse, stylet, etc. Marshall [25] and Omheni et al. [12, 37] study different kinds of people annotating paper texts for a variety of reasons. That's because annotation on paper is a seamless and flexible practice. Annotations on electronic texts have generally been more problematic. On some reading platforms, annotation is clunky, interrupting reading as the reader pulls up menus, makes selections, switches to a keyboard to type in text: the reader's attention is refocused on the user interface rather than on the book's content. Electronic annotation tools may also limit a reader's expressive intent (e.g., forcing a highlight to be continuous when the reader wants to fragment it, or imposing neatness on a reader when she wants to scrawl). Sometimes the electronic annotations are stored in infelicitous ways so they are either gone when the reader returns to the eBook on a different computer or so they are

recoverable when the reader believes them to be deleted and loans the book or document to a colleague. Thus, according to Marshall [25] "*the comfort of annotation on paper cannot be reached. But, the advantages of reading and annotation in digital environment should be able to compensate for this lack of comfort*".

- *Synthesis 4*

The annotative activity in Human-Computer environment can be manual, semi-automatic or automatic. In the cognitive view point annotation systems tend to spend the manual mode to the semi-automatic mode. While in the computational view point annotation tools attempt to spend the semi-automatic mode to the automatic mode. Neither manual nor automated annotation is infallible, and both have advantages. Manual annotation can be performed with relatively little corpus preparation and can be done inexpensively on quite complex phenomena. It is primarily a question of explaining to the annotators the desired annotation scheme and providing the data. But manual work is slow and limited to small results; mounting a large-scale annotation effort that covers tens of thousands of words is a significant undertaking. Automated annotation, on the other hand, requires a considerable investment in corpus preparation and the programming of the automated tagging system, especially if it is first trained on a seed corpus and then applied to a larger one. Its results may be of poor quality. But it is fast, and can produce outputs over very large corpora of all types in very little time [15]. Several efforts have been made towards building scalable, automatic semantic annotation platforms [34]. Most of these systems focus on manual and semi-automatic tooling to improve the productivity of a human annotator rather than on fully automated methods. However, even with machine assistance, annotation of content is a difficult, time consuming and error-prone task. Semantic Annotation faces the challenge to deliver tools capable of full automatic annotation. The question at this moment is if the challenge of semantically annotating pages is solved and we know it still has a long road ahead.

- *Synthesis 5*

In another view point, the aim of the conceptual annotation models is to formalize the main concepts concerning annotations and to define the relationships between annotations and annotated information resources. Therefore, the proposed formal model captures both syntactic and semantic aspects of the annotations. Thus, there are many types of annotation models available in the scientific state of the art and in the already existing end-user applications. We quote in this survey the three main standard frameworks of the annotation: the W3C annotation project Annotea [19], the IEEE Learning Object Metadata (LOM) [17] and the Dublin Core [5]. Using standard formats is preferred, wherever possible, as the investment in marking up resources is considerable and standardization builds in future proofing because new tools, services etc., which were not envisaged when the original semantic annotation was performed may be developed. For annotation systems, standards can provide a bridging mechanism that allows heterogeneous resources to be accessed simultaneously and collaborating users and organizations to share annotations. Because of the need for interoperability, identification and access rights, annotation systems should use basic

annotation framework to model the annotation. But despite the success of RDF technology in the Semantic Web; the Dublin core in Web technology and IEEE-LOM in the world of learning, few annotation systems adopt these standards to model annotations. This raises a big problem of interoperability and compatibility between these systems since each of them is based on a particular personal annotation model.

- *Synthesis 6*

Although the existing annotation tools offer good user interaction interfaces, project management and quality control abilities are still limited. Therefore, some recent research works focus on these challenges. For example Marky [28] introduces, a new Web based document annotation tool equipped to manage multi-user and iterative projects, and to evaluate annotation quality throughout the project life cycle. At the core, Marky is a Web application based on the open source CakePHP framework. User interface relies on HTML5 and CSS3 technologies. Rangy library assists in browser-independent implementation of common DOM range and selection tasks, and Ajax and JQuery technologies are used to enhance user-system interaction. Marky grants solid management of inter- and intra-annotator work. Most notably, its annotation tracking system supports systematic and on-demand agreement analysis and annotation amendment. Each annotator may work over documents as usual, but all the annotations made are saved by the tracking system and may be further compared. So, the project administrator is able to evaluate annotation consistency among annotators and across rounds of annotation, while annotators are able to reject or amend subsets of annotations made in previous rounds. As a side effect, the tracking system minimizes resource and time consumption.

- *Synthesis 7*

Finally, most of the annotation systems described previously serve their intended purpose quite well however to deal with the overwhelming size of the web, new approaches need to be considered. In any computerized system, resources (such as processing power, secondary storage, main memory, etc.) are always scarce. No matter how much we have available, a system could always make use of more. These systems are very demanding, much more than normal ones since they need to process huge amounts of texts using complex linguistic tools in the least possible time. Physical resources are not the only bottleneck in the whole system. The raw material which the systems utilize is digital resources, downloaded from the internet. Unfortunately, this process is still extremely slow, especially when several pages are requested simultaneously from the same server. In synthesis, we can conclude that such systems are only usable by top of the range computers having a high bandwidth connection to the internet. A possible solution is to exploit new technologies in distributed computing such as cloud computing [8]. The benefits of such an approach are various. Since the system is utilizing resources elsewhere on the internet, there is no need of a powerful machine with huge network bandwidth. The client is just a thin system whereby the user defines the seed data and simply initializes the system which is then executed somewhere remotely in the cloud.

4   Conclusions and Future Work

Based on an overview of existing annotation systems, both in research and industry, this article proposes a unified and integrated picture of annotation systems in several areas of technology. This panoramic view is based on a classification of sixty annotation systems developed in the literature during the last twenty five years by industry and academia. This organization of annotation tools is built on the basis of five criteria: annotation type (computational / cognitive); category of annotation system (application/ plugin/ web site); type of annotative activity (manual/ semi-automatic/ automatic); type of annotated resource (text/ web page/ video/ image/ database) and application domain (semantic web / social networks/ digital library/ learning/ databases/ Web services/ computational linguistics/ bioinformatics). The presented list of annotation systems is not exhaustive, but it contains the majority of annotation systems encountered in our survey of annotation tools. Even if there are other annotation systems developed in the literature which are not mentioned in this article, it is certain that these systems can be easily integrated into our classification since the categorization technique is based on cross-cutting criteria applicable for any annotation system.

Nevertheless, the outcome of this article has been limited by the inadequate information about the annotation systems that were discussed. Some of the systems are open source; therefore it is possible to study its documentation and code to explore the structure. However, for many of the other systems, it is very difficult, if not impossible, to get to know their strategies of implementation. Therefore, in the next phase, we will aim to reach a thorough understanding of the implementation and structure of the annotation systems. This tools studied above are necessarily based on annotation models to conceptualize their properties in a formal way to be exploited by computer systems. Thus, we are also planning to propose a survey of conceptual annotation models in digital contents. In another perspective, based on the five criteria of annotation systems classification, we will try to propose a service of annotation systems research. This service presents a user interface providing the possibility of looking for an annotation system which meets the requirements of the user according to the criteria of classification.

References

1. Agosti, M., Ferro, N.: A Formal Model of Annotations of Digital Content. ACM T. Inform. Syst. 26 (1), Article 3 (2007)
2. Andrews, P., Zaihrayeu, I., Pane, J: A classification of semantic annotation systems. Semantic Web Journal 3(3), pp.223–248 (2012)
3. Azouaou, F., Mokeddem, H., Berkani, L., Ouadah, A., Mostefai, B.: WebAnnot: learner's dedicated web–based annotation tool. Int. Journal of Technology Enhanced Learning 5(1), pp.56–84 (2013)

4. Barret, A., Celevenger, J., Martini, J.: Open Sourcing Harvard University's Collaborative Annotation Tool. Academic Technology Services, Harvard University (2013)

5. Carmichael, P.: Learning how to learn: using the Dublin Core metadata element set to support teachers as researchers. In: Proc. of the Int. Conf. on Dublin Core and Metadata Applications: Metadata for e-Communities: Supporting Diversity and Convergence, pp.201–203 (2002)

6. Caussanel. J., Cahier J.P., Zacklad M., Charlet J.: Cognitive Interactions in the Semantic Web. International Workshop on the Semantic Web, Workshop at WWW (2002)

7. Denoue, L., Vignollet, L.: An annotation tool for Web browsers and its applications to information retrieval. In: Proc. of Int. Conf. on Information Retrieval and its Applications, pp.180–196 (2000)

8. Dingli, A.: Knowledge Annotation: Making Implicit Knowledge Explicit. Intelligent Systems Reference Library (2011)

9. Ferro, N.: Digital Annotations: a Formal Model and its Applications. The Information Retrieval Series 22, pp.113–146 (2008)

10. Gabriel, H.-H., Spiliopoulou, M., Nanopoulos, A.: Summarizing dynamic Social Tagging Systems. Expert. Syst. Appl. 41(2), pp.457–469 (2014)

11. Gilbert, M., Morgan, J., Zachry, M., McDonald, D.: Indicoder: An Extensible System for Online Annotation. In: Proc. of the Conf. on Computer Supported Cooperative Work Companion, pp.139–142 (2013)

12. Omheni, N., Kalboussi, A., Mazhoud, O., Kacem, A.H.: Modelling Learner's Personality Profile through Analysis of Annotation Digital Traces in Learning Environment. In: Proc. of the 15th IEEE International Conference on Advanced Learning Technologies (ICALT'15), IEEE, in press, (2015)

13. GRAHL: PDF Annotator manual. GRAHL software design. (2014) Available online at http://www.pdfannotator.com/

14. Hendley, R., Beale, R, Bowers, C., Georgousopoulos, C., Vassiliou, C., Sergios, P., Moeller, R., Karstens, E., Spiliotopoulos, D.: CASAM: collaborative human-machine annotation of multimedia. Multimed. Tools Appl. 70(2), pp.1277-1308 (2014)

15. Hovy, E., Lavid, J.: Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. International Journal of Translation 22(1), (2010)

16. Hwang, W.Y., Wang, C.Y., Sharples, M.: A Study of Multimedia Annotation of Web-Based Material. Computers & Education 48(4), pp.680–699 (2007)

17. IEEE-Standards Association: IEEE Draft Standard for Learning Object Metadata - Corrigendum 1: Corrigenda for 1484.12.1 LOM (Learning Object Metadata). IEEE, Piscataway, pp.1–24 (2010)

18. Kadam, S., Bajpai, S., Yelmar, P.: Annotation: an Investigative Survey of Annotation Types and Systems. In: Proc. of the Int. Conf. on Advances in Engineering and Technology, pp.102–105 (2014)

19. Kahan, J., Koivunen, M.R.: Annotea: an open RDF infrastructure for shared web annotations. In: Proc. of the 10th Int. Conf. on World Wide Web, pp.623–632 (2001)

20. Kalboussi, A., Mazhoud, O., Omheni, N., Kacem, A.H.: An approach of assistance of learner's annotative activity through web services. Australian Journal of Intelligent Information Processing Systems 13(3), pp.15–22 (2013)

21. Kalboussi, A., Mazhoud, O., Omheni, N., Kacem, A.H.: A New Annotation System based on a Semantic Analysis of Learner's Annotative Activity to Invoke Web Services. International Journal of Metadata, Semantics and Ontologies 9(4), pp.350–370 (2014)

22. Kalboussi, A., Omheni, N., Mazhoud, O., Kacem, A.H.: An Interactive Annotation System to Support the Learner with Web Services Assistance. In: Proc. of the 15th IEEE International Conference on Advanced Learning Technologies (ICALT), in press, (2015)

23. Kharkate, S.K., J. Janwe, N.: Automatic Image Annotation: A Review. International Journal of Computer Science & Applications (TIJCSA)1.12 (2013)

24. Margolis, M., Resnick, P. 'Third Voice: Vox Populi Vox Dei?. First Monday J. 4.10 (1999)
25. Marshall, C.C.: Reading and Writing the Electronic Book. Marchionini, G. (Ed.), Morgan & Claypool, Chapel Hill, NC (2009)
26. Morgan, N.P., Schilit, B., Golovchinsky, G.: XLibris: the active reading machine. In: Proc. of the Conf. Summary on Human Factors in Computing Systems, pp.22–23 (1998)
27. Nacer, H., Aissani, D.: Review: Semantic web services: Standards, applications, challenges and solutions. J. Netw. Comput. Appl. 44, pp.134–151 (2014)
28. Perez, M., Glez-Peña, D., Fdez-Riverola, F., Lourenço, A.: Marky: a tool supporting annotation consistency in multi-user and iterative document annotation projects. Computer Methods and Programs in Biomedicine (2014)
29. Kalboussi, A., Mazhoud, O., Kacem, A.H.: Annotative activity as a potential source of web service invocation. In: Proc. of the 9th Int. Conf. on Web Information Systems and Technologies, pp.288–292 (2013)
30. Kalboussi, A., Mazhoud, O., Hadj Kacem, A., Omheni, N.: A formal model of learner's annotations dedicated to web services invocation. In: Proc. of the 21st International Conference on Computers in Education (ICCE'13), pp.166–169 (2013)
31. Omheni, N., Mazhoud, O., Kalboussi, A., HadjKacem, A.: Prediction of human personality traits from annotation activities. In: Proc. of the 10th International Conference on Web Information Systems and Technologies (WEBIST'14), pp.263–269 (2014)
32. Omheni, N., Mazhoud, O., Kalboussi, A., HadjKacem, A.: The annotation: a track of reader's personality traits on paper. In: Proc. of the 2014 ACM Southeast Regional Conference (ACMSE'14), article No.10 (2014)
33. Quint, V. and Carcone, L.: Project Amaya W3C. Available online at: http://dev.w3.org/Amaya/doc/WX/Annotations.html (accessed on October 2014)
34. Reeve, L., Hyoil H.: Survey of semantic annotation platforms. In: Proc. of the ACM symposium on Applied computing, ACM (2005)
35. Rinaldi, F.: Semi-Automated Semantic Annotation of the Biomedical Literature. In: Proc. of the 13th International Semantic Web Conference (ISWC), 473-476 (2014)
36. Shah, R., Yu, Y., Shaikh, A., Tang, S. and Zimmermann, R.: ATLAS: Automatic Temporal Segmentation and Annotation of Lecture Videos Based on Modelling Transition Time. In: Proc. of the ACM International Conference on Multimedia, pp.209–212 (2014)
37. Omheni, N., Kalboussi, A., Mazhoud, O., Kacem, A.H.: Automatic Recognition of Personality from Digital Annotations, In Proc. of the 11th International Conference on Web Information Systems and Technologies (WEBIST'15), Portugal, in press (2015)
38. Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., Strohmaier, M. 'Evolution of reddit: from the front page of the internet to a self-referential community?. In: Proc. of the companion publication of the 23rd Int. Conf. on World wide web companion, pp.517–522 (2014)
39. Tanaka, K., Kunze, K., Iwata, M., Kise, K.: Memory specs: an annotation system on Google Glass using document image retrieval. In: Proc. of the ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14), pp.267–270 (2014)
40. Tosi, D., Morasca, S.: Supporting the semi-automatic semantic annotation of web services: A systematic literature review. Information and Software Technology 61 (2015)
41. Tsai, C.F., Hung, C.: Automatically annotating images with keywords: A review of image annotation systems. Recent Patents on Computer Science 1(1), pp.55-68 (2008)
42. Xiaon, D., Eltabakh, M.: InsightNotes: summary-based annotation management in relational database. In: Proc. of the ACM SIGMOD International Conference on Management of data, pp.661–672 (2014)
43. Zheng, N., Bao, H.: Flickr Group Recommendation Based on User-Generated Tags and Social Relations via Topic Model. In: Proc. of the 10th Int. Symposium on Neural Networks, pp.514–523 (2013)