

Annotations on Digital Contents

Maristella Agosti, Nicola Ferro

Department of Information Engineering – University of Padua
Via Gradenigo, 6/a – 35131 Padova (PD) – Italy
e-mail: {agosti, ferro}@dei.unipd.it

Submitted: 12th July 2005 / Received: date / Revised version: date

Abstract. This paper provides a comprehensive study on annotations, which defines its contours and complexity. This work adds to the usual case and user studies a new complementary approach, which investigates also the history in order to capitalize on previous knowledge, and on our cultural heritage. This study points out an aspect that has never previously been taken into account that is the temporal dimension involved in annotations. Moreover, it discusses both the notion of hypertext between documents and annotations, and the idea of annotations as context for documents. A set of key features of annotations is presented; those features need to be taken into account when designing systems that have to support the management of digital annotations on digital contents.

Key words: annotations – contents – metadata – hypertext – context – temporal dimension

The study and the research concerning the annotation of digital contents is an active field of investigation, which is faced from many different perspectives. The previous research work on annotations has left us with many open issues. These issues concerned the lack of clarity about: what an annotation is, what the features and the way of using an annotation are, and what architecture and functionalities a system with annotation capabilities has to provide. These issues are mainly due to the fact that, up to now, models and systems for annotations have been developed for specific purposes. As a result of this, there is a fragmentary picture on the annotation and its management, which are tied to specific usages and lack a general validity.

The aim of this paper is to discuss the different perspectives regarding annotations in order to gather some key features about them. These key points can help us to

better distinguish among the different usages of annotations and to understand what is the case at hand, if we come to deal with annotations. Furthermore, they can be used as a support if we need to take design choices for developing an annotation management system. Finally, they can also be seen as the groundwork needed to properly delineate the complexity of this problem, which is often underestimated and only partially addressed, and to open the way for defining a comprehensive formal model of annotations on digital contents, which was not dealt with in previous research on this topic and even today is still lacking. In conclusion, we aim at contributing to solve the issue pointed out by [16]:

strangely enough, there is not an agreement yet on the definition of digital annotation, or on how to distinguish it from other digital entities (e.g. hyperlinks, metadata, newsgroup messages). Furthermore, an analysis of the basic operations, to be enabled by a digital annotation system, seems to be lacking.

This paper aim neither at being a fully exhaustive survey on annotations nor at comparing all the existing annotation systems along some predefined set of features. The reader interested in these issues can refer to [22, 31, 52, 56]

In order to achieve our goal, we proceed in a twofold way: firstly, Section 1 presents a thorough study of the annotation from an historical point of view; secondly, Section 2 analyzes the present perspectives about annotations. The rationale for studying both the historical and the present viewpoints about annotations is that our current notion of annotation, even in a digital world, is strongly influenced by the long history of the annotation. Thus, looking at both aspects can provide us with a better understanding of annotations. After this step, Section 4 discusses how annotations can be used in order to search for relevant documents in response to a user

query. Section 5 gathers up the observations made in the previous sections in order to highlight some key features about annotations that have to be taken into account. Finally, Section ?? draws some conclusions.

1 Groundwork on Annotations

A basic step in approaching the problem of annotations is to define the meaning of the different terms that come into play and to investigate their historical usage over the course of time. All of this is necessary in order to gather information as to delineate the contours of the problem.

Thus, in this section, we have conducted a research about annotations that has a literary approach. We firmly believe that studying the terms, their meaning, their etymology, and the way they have been used can provide us with a solid groundwork on which we can build the subsequent research. When we talk about annotations, we deal with a concept that has been stratified for a long period of time in our culture, and literary research is the most effective way to benefit from the pre-existing knowledge of our cultural heritage [3]. The outcomes of this research are key points concerning the features of the annotation, that we should take into account when we develop a model for the annotation and when we design a system capable of providing annotation functionalities.

This kind of research is complementary to user studies that are conducted in order to gather user requirements. Furthermore, it completes user studies with a knowledge that users are not often able to express, because often they overlook what they have naturally absorbed from their cultural heritage. In some sense, we are conducting a user study, where our user is the history of the annotation and we ask what features of the annotation are relevant for us.

We have adopted the following methodology: first, we look up the meaning of the term at hand in the dictionary; then, we investigate its etymology and its historical usage; finally, we gather the information provided by the two previous steps so that we can emphasize some key points about the annotation.

1.1 The Term Annotation

Both [32, p. 57] and [35, p. 198] define the word *annotation*, firstly, as the *act of annotating* and, secondly, as *a note added in explanation especially of some literary work*. [35, p. 198] further observes that the word *annotation* can also be used in sentences with a passive sense, as thing worthy of annotation, with the meaning of *note-worthy* and *worth remembering*. The word *annotation* is closely related to the verb *annotate* that, in turn, means *to supply (a written work, such as an ancient text) with critical or explanatory notes* [32, p. 57] and *to note down, to write down, to record something* [35, p. 198].

Both [23, p. 107] and [32, p. 57] trace the etymology of annotation back to the Latin word *annōtātiō*, that simply means *annotation, note* [21, p. 189]. The Latin word *annōtātiō*, in turn, derives from the Latin verb *annōtāre*, which means *to annotate* and *to observe in writing* [21, p. 190]. Finally, the Latin verb *annōtāre* comes from the Latin word *nōta*, that means *note, mark* [21, p. 1823], plus the intensifying prefix *ād*, which in compound words means *to approach, to tend* and thus *to add* [21, pp. 37–40]. Both [21, p. 1823] and [23, p. 1047] agree that the Latin word *nōta* has an uncertain etymology.

This brief discussion highlights some interesting points about the annotation. Firstly, the annotation is not only an object or something that is passive, but it also contains the notion of activity, as its first meaning “act of annotating”. In this sense, the annotation calls for an active involvement by the subject who is engaged in the act of supplying explanatory matter or keeping record of something. Furthermore, the annotation covers, in its second meaning, also the purpose of this active involvement, that is to produce an intellectual work in order to add an explanation to some literary work, as an example. This idea is further supported by the meaning of the verb *to annotate*, which broadens the spectrum of the word *annotation* towards keeping record of something. Therefore, on the whole, the annotation requires an active involvement in order to produce an intellectual work that has to be recorded. These facets of the annotation are present also in the etymology of the word *annotation*: the Latin verb *annōtāre* means *to make written observations, comments or remarks in the durable and recordable written form*; on the other hand, the Latin word *nōta* recalls the *note* or the *mark* put to remember or highlight something. Finally, the outcomes of the act of annotating are also taken into account: indeed, both the annotated object, in a passive sense, and the content of the annotation, in an active sense, become noteworthy and worth recording.

1.2 Terms Related to Annotation

The range of our investigations can be widened in order to take into consideration also synonyms and terms related to the word *annotation*.

[63, p. 14] provides the following synonyms of the term *annotation*: *comment, commentary, elucidation, explanation, footnote, gloss, interpretation, and note*. We can also add to this list the word *jotting*, which is a very brief annotation [32, p. 789], and the word *scholium*, which is a particular kind of annotation [38, p. 158].

Table 1 provides the definitions for the different terms listed above. Note that these definitions often refer to printed documents or texts, since they are taken from an English dictionary [32]. On the other hand, we should consider that their validity is not limited only to printed

documents, but it can also be applied and extended to information resources in a digital context.

As it can be noticed from table 1, these words are often defined by terms used to describe other words in the list, or they refer to the same notion of explaining, expounding, interpreting, clarifying, recording something. In this way, they reveal how closely related they are.

In conclusion, the terms listed in table 1 support, refine and enforce what has been observed above about the word annotation, introducing further kinds of annotation which cover different needs and tasks, such as the gloss, the postil, the note, the jotting, and so on.

Now, we can move a step further and investigate in more detail the terms gloss, scholium, and postil in order to understand the rich semantics of the annotation and how it has evolved with the passing of time and its current consequences.

1.3 The Term Gloss

As reported in [32, p. 620] and [23, p. 673], the word gloss derives from the ancient Greek word γλῶσσα (glōssa), that means *tongue, language, idiom, spoken word, foreign or obsolete word* [60, p. 393].

As reported in [36, pp. 652–653], at the time of the ancient Greeks, the term gloss meant an obscure, archaic, dialect, or rare locution that required an additional explanation. These locutions were object of study by grammarians or object of research by scholarly poets, especially the Alexandrine poets, who embellished their compositions with these terms. Then, gloss meant the explanations themselves of such locutions, either collected in wide-ranging lexicons or as interlinear notes placed on top of the words to explain. This was a methodology of study and a lexicographical practice that dates back to very ancient times (there were glosses to Homer already in the V century B.C.) and that was fully developed by the grammarians of the Alexandrine age. During the Bizantine age and the Middle Ages, the term gloss meant an interlinear or marginal note to a biblical or juridical codex. For the biblical codices, the gloss was a very short paraphrase to explain a passage of the Bible, sometimes together with a mention to its allegorical interpretation. On the other hand, for the juridical codices, the glosses were explanatory annotations, that constituted a thorough commentary to the text.

The gloss was a practice that flourished especially in the juridical context, as reported by [34, pp. 427–429]. During the Roman Empire, one of the usual literary forms of the Roman jurisprudence was the comment to the works of former jurists, so that it is often possible to distinguish the annotated text from the annotation to the text; furthermore, the glosses were sometimes physically separated from the annotated text. However, the most famous use of this kind of method of study is due to the Bolognese school: indeed, the word gloss denoted the way of studying the Justinian Code practised in

Bologna, which began in the II century A.D. The Bolognese gloss passed from a simpler form to a more complex one, that is it passed from simple interlinear notes to a real theoretical treatment of the subject. The glossarist reveals the contradictions (*contrāriētātēs*¹) of the Justinian books, raises doubts (*dūbītātīōnēs* or *dūbiētātēs*), that often give rise to controversies (*dissēnsiōnēs*). The contradictions often find an explanation (*sōlūtiō*) and the doubts disappear by means of an appropriate distinction (*distinctiō* or *diffērentiā*). The glossarist teaches the Justinian books and creates cases in point and examples that originate glosses pointing out the different cases (*cāsūs*); furthermore, the glossarist fixes and defines rules derived from the texts he studies, and, accordingly, creates glosses that report such rules (*rēgūlae*) and definitions (*dēfīnītīōnēs*). In conclusion, the Bolognese gloss was a way of doing research aimed at defining and elucidating the law.

This discussion about the term gloss points out some interesting facets of the annotation, that has not fully emerged in the previous observations about the term annotation. The intellectual work entailed by the gloss is of very high quality, because it is a method both of study and of research. This kind of intellectual work gives us an idea of how strong the active involvement required by the gloss is: it does not concern only the author himself, but it is also capable of involving and stimulating a wide community of people that works, studies and does research on a subject. Thus, it turns out that an annotation may comprise a public dimension, because it becomes the vehicle for carrying and transmitting ideas and knowledge to other people, or it may comprise a shared dimension, if the recipients of the annotation are less numerous. Finally, the research or study aspects, and the public or shared dimension entailed by the gloss help us to understand how durable and recordable the annotations are. Indeed, they are not only comments and remarks to a text, but also an autonomous intellectual work, that worths recording.

1.4 The Term Scholium

As reported in [32, p. 1305] and [23, p. 1479], the word scholium derives from the ancient Greek word σχολίον (schólion), that means *comment, explanation* [60, p. 1793]. The ancient Greek word σχολίον (schólion), in turn, comes from the ancient Greek word σχολή (scholḗ), that means *scholar activity* and *school* [60, p. 1793].

[33, p. 198–199] reports that the word scholium designates short annotations or explanations written by a reader in the margin of a manuscript. The distinguishing features of the scholium are the fact that they are

¹ The italicized words in brackets are the Latin technical terms used to indicate the specific technique applied in each step of the method of study. Note that they are the translation of the word which precedes them.

Word	Definition
Comment	<i>a note explaining or criticizing a passage in a text and explanatory or critical matter added to a text</i>
Commentary	<i>an explanatory series of notes or comments</i>
Elucidation	<i>making clear (something obscure or difficult)</i>
Explanation	<i>the act or process of explaining and a statement or occurrence that explains and a clarification of disputed terms or points</i>
Footnote	<i>a note printed at the bottom of a page, to which attention is drawn by means of a reference mark in the body of the text</i>
Gloss	<i>a short or expanded explanation or interpretation of a word, expression, or foreign phrase in the margin or text of a manuscript</i>
to Jot	<i>to write a brief note of</i>
Jotting	<i>something jotted down</i>
Interpretation	<i>the act or process of interpreting or explaining; elucidation and the result of interpreting; an explanation</i>
Note	<i>a brief summary or record in writing especially a jotting for future reference and a short written statement giving any kind of information and a critical comment, explanatory statement, or reference in the text of a book, often preceded by a number</i>
Notes	<i>short descriptive or summarized jottings taken down for future reference</i>
Observation	<i>the act of observing or the state of being observed and a comment or remark and the facts learned from observing</i>
Postil	<i>a commentary or marginal note, as in a Bible</i>
Record	<i>an account in permanent form, especially in writing, preserving knowledge or information about facts or events</i>
Scholium	<i>a commentary or annotation, especially on a classical text</i>

Table 1. Definition of the terms related to annotation from [32] .

anonymous and fragmentary. The scholium is anonymous because, initially, the reader writes in the margin of the manuscript his own observations or passages taken from a commentary for both personal use and scholastic needs. The next owner of the manuscript often extends the scholium or modifies it. Thus, the lack of organic unity is explained in this way. Often the scholia contain also citations by the authors from which the observations are taken; this way, they are very useful in order to reconstruct the doctrines and the works of ancient grammarians that may no longer exist.

The term scholium suggests another facet of the annotation: it may be created for personal purposes, that is the annotation may entail a private dimension, since the main recipient of the annotation is the author themselves. However, the private dimension may represent only the initial intention of the annotation, because also other people reading an annotated text can benefit from existing annotations and can modify or extend them; thus, the annotation passes from a private dimension to a shared one. Taken to the extreme this process encompasses the possibility that an annotation becomes the means to study the thought of authors that otherwise would be lost; thus, the annotation passes from a private dimension to a public one. In conclusion, private annotations are part of this spectrum of possibilities and this makes us aware of the necessity to carefully preserve private annotations, because they may become

worth recording also for different reasons from the ones that motivated their creation.

1.5 The Term Postil

As introduced in table 1, a postil is a short annotation – often a marginal or interlinear note – to a text, handwritten by a scholar or by the author himself in order to express observations, explanations, or criticisms. During the Middle Ages, the postils were a scholastic practice and they sometimes represented comments that were broader than simple notes [37, p. 1030].

Both [32, p. 1145] and [23, p. 1239] trace the etymology of postil back to the Latin terms *pōst illā (verbā textūs)* that mean *after those (words in the text)*, which often was the opening phrase of such annotation.

Thus, the word postil points out in its etymology itself one of the main aspects concerning the annotation: the annotation is the result of an intellectual work on an existing text and it follows an already existing text. Thus, the annotation comprises a temporal dimension that is often not explicit but that limits the creation of the annotation to the existence of another text. This temporal relationship between the annotation and the annotated text does not mean that the annotation cannot be considered as a stand-alone intellectual work – and some glosses and scholia are by right autonomous pieces of knowledge – but it imposes a temporal order-

ing between the existence of an annotated text and the annotation annotating it, that cannot be neglected.

2 Approaches to the Management of Annotations

Many user studies are aimed at understanding annotation practices and discovering common annotation patterns. [44] studied personal annotative practices of American college students in order to point out the form the annotations take on in the textbooks and the function of the annotations derived from their form. [44, pp. 237–238] discovered that:

First, annotations are *procedural signals*, cluing in the student to where an assignment starts, what material is important (and as we will see, unimportant), and what material might require a second (or successive readings). Second, annotations are *placemarks*; they hold the quotes that are being reserved for the paper that the student will write at the end of the term, the chemical reactions and term definitions the student must memorize for the final, the theorem that is key to the proof in the homework assignment. Third, they are an *in situ way of working problems*. Fourth, annotations record *interpretive activity*, either from another reader (e.g. a professor's explanation), or as the result of careful reading (the student has interpreted it him or herself). Fifth, and most elusively, these markings act as a *visible trace of a reader's attention*, a focus on the passing words, and a marker of all that has already been read (as if these words are now possessed). Finally, the markings may just be incidental, *reflecting the material circumstance of reading*.

[45] carries on her research work and categorizes annotations along several dimensions, that reflect the form which annotations may take on: formal versus informal annotations, explicit versus tacit annotations, annotations as writing versus annotations as reading, hyperextensive versus extensive versus intensive annotations, permanent versus transient annotations, published versus private annotations. Finally, [46], [47], and [62] investigate the relationship among private, shared and public annotations and how they can be exploited to find useful passages in the text.

It is worth noting how the findings of [44] and [45] agree with the outcomes of the study conducted in the previous section about the historical perspective on annotations. Indeed, both glosses and scholia are, to some extent, *placemarks*, an *in situ way of working problems*, they record an *interpretive activity*, and so on. Also the different dimensions of the annotation are taken into account by the historical perspective: glosses are often

more formal annotations than scholia and postils, that are usually informal; scholia can be tacit annotations due to their fragmentariness while glosses can be explicit annotations; all the kinds of annotations described in Section 1 act as a bridge between reading and writing; glosses may be considered intensive annotations, postils may be more extensive annotations and both glosses and scholia often contain references to other authors and quotations of other texts, that is a way of being hyperextensive annotations; the stratification of glosses and scholia in our cultural heritage is a clear sign of the passage from transient to permanent annotations; finally, the difference among postils, scholia, and glosses comprises the distinction between private and public annotations. On the other hand, neither [44] nor [45] explicitly points out the temporal dimension entailed by annotations and the temporal ordering between annotations and annotated objects, which has been discussed in Section 1.5 talking about the term postil.

[56] suggest a list of desirable properties for annotations: annotations should appear *in situ*, that is on the documents themselves; they should be *highly expressive*; they should be *format and platform independent*; they should be *extensible, yet composable*, that is they should allow different styles of annotation; they should be *distributed, open, and robust*, that is they may reside in a place while referring to documents in another place.

Over the years, a lot of research work concerning annotations has been done, where the main focus of this work has been on the employment of ad-hoc devices or handheld devices which enable reading appliances with annotation capabilities [49–51, 61], or on the design and development of document models and systems which support annotations [55–58, 22, 16] in specific management systems, in particular:

- in digital libraries [2, 8, 6, 9, 3, 7, 59, 30],
- in the Web [13–15, 25, 40, 52, 69, 68],
- in collaborative systems and working groups [26, 27, 53, 67], and
- in databases [12, 19, 20, 18, 65].

All of this research work has led to different viewpoints about what an annotation is [7, 5]; these different viewpoints are taken into consideration in the following.

2.1 Annotations are Metadata

They can be considered as additional data which concern an existing content, that is annotations are metadata, as they clarify in some way the properties and the semantics of the annotated content. For example, the Annotea² project developed by the World Wide Web Consortium (W3C) [40] sees annotations as metadata and interprets them as the first step in creating an infrastructure which

² <http://www.w3.org/2001/Annotea/>

will handle and associate metadata with content and will lead to the Semantic Web³.

Another example is MPEG-7 [39], formally named “Multimedia Content Description Interface”, which is a standard for annotating and describing multimedia content data. MPEG-7 supports to some degree the interpretation of the information meaning, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes support as many broad range of applications as possible. As a further example, in the context of *DataBase Management System (DBMS)* [12] sees annotations as “information about data such as provenance, comments, or other types of metadata”.

2.2 Annotations are Contents

Differently from the previous case, they are additional contents which concern an existing content [52]; indeed, they increase existing content by providing an additional layer of content that elucidates and explains the existing one.

This viewpoint about annotations entails an intrinsic dualism between annotation as content enrichment and annotation as stand-alone document [3]:

- *annotation as content enrichment*: in this view annotations are considered as mere additional content regarding an existing document and as a result they are not autonomous entities but in fact they rely on previously existing information resources as to justify their existence;
- *annotation as stand-alone document*: in this view annotations are considered as real documents and are autonomous entities that maintain some sort of connection with an existing document.

This twofold nature of the annotation is clear if we think about the process of studying a document: firstly, we can start annotating some interesting passages that require an in-depth investigation, which is an annotation as content enrichment; then we can reconsider and collect our annotations and we can use them as a starting point for a new document, covering the points we would like to explain better, all of which is an annotation as a stand-alone document. In this case the annotation process can be seen as an informal, unstructured elaboration that could lead to a rethinking of the annotated document and to the creation of a new one. Also [16] agree with this viewpoint about annotations and consider them as being reliant on the annotated objects; in this way, [16] consider annotations as content enrichment.

2.3 Annotations Constitute a Hypertext

They allow the creation of new relationships among existing contents, by means of links that connect annotations together with existing content. In this sense we can consider that existing content and annotations constitute a hypertext [6], according to the definition of hypertext provided in [1]. This hypertext can be exploited not only for providing alternative navigation and browsing capabilities, but also for offering advanced search functionalities. Furthermore, [45] considers annotations as a natural way of creating and growing hypertexts that connect information resources in a *Digital Library Management System (DLMS)* by actively engaging users. Finally, the hypertext that exists between information resources and annotations enables different annotation configurations: the first are *threads of annotations*, i.e. an annotation made in response to another annotation, and the second are *sets of annotations*, i.e. a bundle of annotations on the same information resource [3, 7].

2.4 Annotations are a Kind of Context

Annotations introduce a new content layer aimed at elucidating the meaning of underlying documents, so that annotations can make hidden facets of the annotated documents in a more explicit way. In conclusion, we can consider that annotations constitute a special kind of context, that we call *annotative context*, for the documents of a DLMS, because they provide additional content which is related to the annotated documents. This viewpoint about annotations covers a wide range of annotations, ranging from personal jottings in the margin of a page to scholarly comments made by an expert in order to explain a passage of a text. Thus, these different kinds of annotations involve different scopes for the annotation itself and, consequently, different kinds of annotative context. If we deal with a personal jotting, the recipient of the annotation is usually the author himself and so this kind of annotation involves a *private annotative context*; on the other hand, the recipients of a scholarly annotation are usually people who are not necessarily related to the author of the annotation, which thus involves a *public annotative context*; finally, a team of people can work together on a shared topic and can exchange annotations related to the topic in question: thus, in this case we have a *collaborative annotative context* [6].

[53] make use of annotations for exploiting and reconstructing different types of context. For example, annotations can be useful for reconstructing the *creation context* of historic documents, when this context is not available any more. For example, systems like *Collaboratory for Annotation Indexing and Retrieval of Digitized Historical Archive Material (COLLATE)* [17, 26, 66] or *Imaginum Patavinae Scientiae Archivum (IPSA)* [2, 9,

³ <http://www.w3.org/2001/sw/>

8] support researchers in performing this task. Furthermore, [53] exploit annotation for gathering information about the *interpretation context* of a document, so that it is possible to provide users with a framework where the interpretation of documents takes place. Finally, also the *collection context*, which provides information about documents in relation to other documents within a collection, can be enhanced by using the hypertext that exists between annotations and annotated documents [53].

2.5 Annotations are Dialog Acts

They are part of a discourse with an existing content, where not only the single statement within the discourse is taken into consideration, but also its position and type within the discourse structure. This approach is taken by [26, 27, 66] who consider annotations as a means to place a document in the collaborative discourse to which it is related.

[15, 16] rely, in some way, on this notion of annotation, because they make use of the *Rhetorical Structure Theory (RST)* [43] in order to define a set of annotation types. Note that also [66] use the RST for the same purpose.

Finally, [25] agree, to some extent, with this viewpoint about annotations. Indeed, they interpret annotations as a means that allow a “two way exchange of ideas between the authors of the documents and the documents users”.

3 Usages of Annotations

However, the viewpoints discussed in the previous section are not completely disjointed, on the contrary, they may overlap and they may be simultaneously present in some situations.

In the following sections, we will go into more detail about the current viewpoints concerning annotations and we will present some interesting cases of usage of annotations and *Information Management Systems (IMS)* with annotation capabilities in the context of digital libraries, the Web, and databases.

3.1 Digital Libraries

Digital libraries are not only the digital versions of traditional libraries and archives, but offer means which go beyond mere presentation of the content stored in digital repositories. In the following we point out this fact by discussing two definitions of digital libraries, which come from two different fields. The more computer science oriented view is expressed in the introduction in the first issue of the *International Journal on Digital Libraries*, cited by [28]:

Digital Libraries are concerned with the creation and management of information resources, the movement of information across global networks and the effective use of this information by a wide range of users.

Librarians have a different definition of *Digital Library (DL)*, as proposed by the Digital Library Federation, 1998, cited by [28]:

Digital Libraries are organisations that provide the resources, including the specialised staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities.

Annotations can be exploited in order to provide users with the distinguishing features of DL systems highlighted above. Note, however, that also archives have to be taken into account in this context, although they are not explicitly mentioned in the previous definitions. The *creation* of new information resources is supported by annotations in two ways. First, when users add annotations to existing information resources, they in turn become new information resources themselves. 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. Annotations might increase and expand the information resources managed by the digital library. In this way, they may provide *interpretations* of information resources. User communities benefit from such interpretations in that they help the understanding of the annotated resource and contain additional information about it. As an example, in the Humanities interpretation is one of the basic tasks scholars perform: systems like COLLATE [26, 66] or IPSA [2, 9, 8] support this task through annotations. Annotations support user communities in *accessing* the information resources provided by the digital library in a personalised and customized way: indeed, users can create annotations that link different documents, enabling alternative paths for browsing digital contents and thus structuring them in alternative ways.

Different layers of annotations can coexist on the same document: a private layer of annotations accessible only by the annotations author themselves, a collective layer of annotations, shared by a team of people, and finally a public layer of annotations, accessible to all the users of the digital library. In this way, user communities can benefit from different views of the information resources managed by the digital library [44, 46, 47]. A DL can encourage cooperative work practices, enabling the sharing of documents and annotations, also with the aid of special devices, such as XLibris [61]. Finally, as

suggested in [48,51], searching, reading and annotating a DL can be done together with other activities, for example working with colleagues. This may also occur in a mobile context, where merging content and wireless communication can foster ubiquitous access to DL systems, improving well established cooperative practices of work and exploiting physical and digital resources. The wireless context and the small form factor of handheld devices challenge our technical horizons for information management and access and require specialized solutions in order to overcome the constraints imposed by such kinds of devices, as reported in [4].

In the context of a DL system it is also possible to create automatic annotations, which may facilitate the user's first approach with a document. Automatic annotations can be created by using topic detection techniques in order to associate each annotation with its related topic, which constitutes the context of the annotation. In this way, a document can be re-organized and segmented into topics, whose dimension can range in many different sizes, and annotations can present a brief description of those topics. Then, by applying automatic hypertext construction techniques, similar to those presented in [10], those pairs of topics and annotations can be linked together, proposing an alternative way of navigating the content of a digital library.

Finally, [59] and [30] propose a data model for the composition and metadata management of documents in a distributed setting, such as a DL system. They allow the creation of *composite documents*, that are made up of either composite documents or *atomic documents*, that can be any piece of material uniquely identifiable. A set of annotations is associated to each composite document, where [59] and [30] interpret annotations as terms taken from a controlled vocabulary or taxonomy to which all authors adhere. They provide algorithms to automatically compute the annotations of composite documents starting from the annotations of its composing atomic documents, by means of a subsumption relation defined within the taxonomy mentioned above.

3.2 The Web

As previously introduced, the Annotea project [40] considers annotations as metadata. Annotea defines annotations as comments, notes, explanations, or other types of external remarks that can be attached to any Web document or a selected part of the document without modifying the document. Annotea uses *Resource Description Framework (RDF)*⁴ and *eXtensible Markup Language (XML)*⁵ for describing annotations as metadata and XPointer⁶ for locating the annotations in the annotated document. Annotea employs a client-server

architecture, where annotations reside in dedicated servers and a specialized browser is capable of retrieving them upon request, when visiting a Web page. [41] and [42] move a step further and employ annotations as an extension of the bookmarks in order to improve the collaboration among users: indeed, the additional data provided by annotations are exploited to describe, organize, categorize, share, and search for the bookmarks.

Moreover, the W3C Multimodal Interaction Working Group⁷ is developing a markup language called *Extensible MultiModal Annotation (EMMA)* [68]. EMMA is a markup language intended for providing semantic interpretations for a variety of inputs, such as speech, natural language text, and *Graphical User Interface (GUI)* input. The language is focused on annotating the interpretation information of single and composed inputs, and it is expected that this markup will be used primarily as a standard data interchange format between the components of a multimodal system. The general purpose of EMMA is to represent information automatically extracted from a user's input by an interpretation component. EMMA provides a simple structural syntax for the organization of interpretations and instances, and an annotative syntax derived from RDF to apply the annotation to the input data at any level.

As a further example, *Multimedia Annotation of Digital Content Over the Web (MADCOW)* is based on a client-server architecture as Annotea is. Servers are repositories of annotations to which different client can connect, while the client is a plug-in for a standard Web browser [13]. MADCOW employs *HyperText Transfer Protocol (HTTP)* in order to annotate Web resources and allows both private and public annotations. Moreover, it allows different pre-established types of annotations, such as explanation, comment, question, solution, summary, and so on; in this respect, MADCOW opts for a solution similar to the one of COLLATE, which is not Web-based but it models annotations as different types of dialog acts [26].

3.3 Databases

Annotations are used also in the context of the DBMSs and, in particular, in the case of *curated databases* and *scientific databases*. SWISS-PROT⁸ is a curated protein sequence database, which strives to provide a high level of annotation, such as the description of the function of a protein, its domains structure, and so on. In this case, the annotations are embedded in the database and merged with the annotated content. BIODAS⁹ provides a *Distributed Annotation System (DAS)*, that is a Web-based servers system for sharing lists of annotations across a certain segment of the genome. In this

⁴ <http://www.w3.org/RDF/>

⁵ <http://www.w3.org/XML/>

⁶ <http://www.w3.org/XML/Linking>

⁷ <http://www.w3.org/2002/mmi/Group/>

⁸ <http://www.expasy.org/sprot/>

⁹ <http://biomas.org/>

case, the annotation are not mixed together with the content they annotate, but they are separated from it. Annotations have types, methods and categories: the annotation type is selected from a list of types that have biological significance; the annotation method is intended to describe how the annotated feature was discovered and may include a reference to a software program; the annotation category is a broad functional category that can be used to filter, group and sort annotations. Finally, annotations may also be associated with *Web Uniform Resource Locators (URLs)* that provide additional human readable information about the annotation itself [64]. Another example is SEED, a *Peer-To-Peer (P2P)* system which aims to provide the biology community a suite of open source tools to enable distributed teams of researchers to rapidly annotate new genomes. In particular, the SEED enables researcher to create, collect, and maintain sets of gene annotations organized by group of related biological and biochemical functions across many organisms [54].

In the context of scientific databases, [18] proposes an archiving technique in order to manage and archive different versions of such kinds of databases, as time moves on. [18] exploit the hierarchical structure of scientific data in order to represent the content and the different versions of the database with a tree structure. They attach annotations to the nodes of the tree, annotations that contain time-stamp and key information about the underlying data structure. Thus, these annotations are metadata about the database itself. These annotations different from the annotations contained in the database, that are metadata about genome sequences. In conclusion, this annotated tree structure provides an additional data layer, that allows the development of efficient algorithms in order to archive and search for the different versions of the database.

[19,20] investigate the usage of annotations with respect to the *data provenance* problem, sometimes also referred to as *data lineage* or *data pedigree*, which is the description of the origins of a piece of data and the process by which it arrived in a database. [19] distinguishes between *why-provenance*, which explains why a given piece of data is in the database, and *where-provenance*, which explains where a given piece of data comes from. Data provenance is a relevant issue in the field of curated and scientific databases, such as genome databases, because in this field there are few databases that are sources of data, so that we can actually that they receive the experimental data; all the other databases are in some sense views of these source databases or of other views. The distinguishing feature of these databases is the fact that they have to be curated: in fact, they provide corrections and annotations to the original source data made by experts. It is now clear that data provenance is essential to any user interested in the accuracy and timeliness of the data. In particular, where-provenance is important for understanding

the source of errors in data and for carrying annotations through database queries, problems addressed in [20]. [12] carry on the research about where-provenance and propose and implement an extension to a relational DBMS and an extension to *Structured Query Language (SQL)*, called *propagate SQL (pSQL)*, which provides a clause for propagating annotations to tuples through queries. [12] intend annotations to be an information about data such as provenance, comments, or other types of metadata; they envisage the following applications of annotations in DBMS: tracing the provenance and flow of data, reporting errors or remarks about a piece of data, and describing the quality or the security level of a piece of data.

4 Using Annotation for Search Purposes

Despite all of the research in modelling annotations and providing annotation-enabled systems, there is much less study regarding the usage of annotations for retrieving documents.

[29] compare queries based on annotations with relevance feedback, and considers annotation-based queries as an automatic technique for query construction, since queries are automatically generated from annotated text, e.g. from highlighted text.

[26] consider annotations – specifically annotations threads – as an extension of the document they belong to, creating a discourse context, in which not only the annotation itself but also its position in the discourse and its type, are exploited for searching and retrieving documents; this approach is revised and extended upon in [27] to probabilistic datalog.

[6] address this issue in the context of data fusion [24]. In fact, annotations provide us with an additional source of evidence, which is complementary to the one already contained in the set of documents. Thus, we can exploit annotations with two final goals of retrieving more relevant documents and of ranking them better. Furthermore, the paths that connect annotations to documents in the hypertext become the vehicle for moving this further source of evidence from annotation towards the documents. Furthermore, both *Hypertext Information Retrieval (HIR)* techniques [11] and *link fusion* techniques [70] can be exploited in order to provide an advanced search strategy that involves annotations.

5 Key Features

On the whole, the line of reasoning conducted in the previous sections provides us with some key features of the annotation that we should take into account when designing systems that have to support the management of digital annotations on digital contents. Those key features are presented in the following subsections.

5.1 First-class Intellectual Work

Annotations are a valuable intellectual work, as it emerges from the discussion in Section 1 and from the user studies reported in Section 2. The spectrum of this intellectual work is very broad, because it ranges from explaining and enriching an information resource with personal observations to transmitting and sharing ideas and knowledge on a subject. In conclusion, annotations can be geared not only to the way of working of the individual and to a method of study, but also to a way of doing research.

5.2 Various Facets

Annotations comprise different viewpoints, as discussed in Section 2: they may be considered as metadata, content, hypertext, context, or dialog acts. Moreover, the boundaries between these viewpoints are not sharp and they may coexist. All of these viewpoints have to be taken into account, especially because they are tightly coupled with and are the expression of the different kinds of intellectual work that an annotation may bear.

5.3 Different Scopes

Annotations involve different scopes and different kind of annotative context: they can be private, shared or public, according to the type of intellectual work that is carried out. Moreover, the boundaries between these scopes are not fixed but they may vary and evolve as the time passess.

5.4 Active Involvement

Annotations call for an active involvement, whose degree varies according to aim of the annotation: private annotations requires the involvement of the authors, although shared or public annotations involve the participation of a whole community. Thus, annotations are suitable for improving collaboration and co-operation among users.

5.5 Temporal Dimension

Annotations implicitly entail a temporal dimension, that regulates the temporal ordering among annotations and annotated information resources. This temporal dimension needs to be taken into account and to be explicitly modeled, if we aim at designing and developing systems with annotation capabilities.

5.6 Search by Using Annotations

Annotations provide an additional source of evidence that can be effectively exploited in order to search for

documents. Furthermore, this additional source of evidence holds complex relationships with the annotated documents. These complex relationships can be made explicit by using the hypertext that connects documents and annotation. This way, the hypertext become the vehicle for moving this source of evidence from annotations towards documents, so that it is possible to develop advanced search strategies.

5.7 System Viewpoint

Annotations support a wide range of usages, as the previous discussion shows. Thus, annotations functionalities should not be embedded in any given system, but rather in a stand-alone system capable of providing annotation functionalities to other systems.

Acknowledgements

The work reported in this paper has been partially funded by the ECD project, which is a joint program between the Italian National Research Council and the Italian Ministry of Education, with regards to law 449/97-99.

The work was also partially supported by the DELOS Network of Excellence on Digital Libraries, as part of the Information Society Technologies (IST) Program of the European Commission (Contract G038-507618).

References

1. M. Agosti. An Overview of Hypertext. In Agosti and Smeaton [11], pages 27–47.
2. M. Agosti, L. Benfante, and N. Orio. IPSA: A Digital Archive of Herbals to Support Scientific Research. In T. M. T. Sembok, H. B. Zaman, H. Chen, S. R. Urs, and S. H. Myaeng, editors, *Proc. 6th International Conference on Asian Digital Libraries. Digital Libraries – Digital Libraries: Technology and Management of Indigenous Knowledge (ICADL 2003)*, pages 253–264. Lecture Notes in Computer Science (LNCS) 2911, Springer, Heidelberg, Germany, 2003.
3. M. Agosti and N. Ferro. Annotations: Enriching a Digital Library. In T. Koch and I. T. Sølvyberg, editors, *Proc. 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*, pages 88–100. Lecture Notes in Computer Science (LNCS) 2769, Springer, Heidelberg, Germany, 2003.
4. M. Agosti and N. Ferro. Chapter X: Managing the Interactions between Handheld Devices, Mobile Applications, and Users. In E. P. Lim and K. Siau, editors, *Advances in Mobile Commerce Technologies*, pages 204–233. Idea Group, Hershey, USA, 2003.
5. M. Agosti and N. Ferro. A System Architecture as a Support to a Flexible Annotation Service. In C. Türker, M. Agosti, and H.-J. Schek, editors, *Digital Library Architectures: Peer-to-Peer, Grid, and Service-Oriented, Post-proceedings of the 6th Thematic Workshop of the*

- EU Network of Excellence DELOS*, page (in print). Lecture Notes in Computer Science (LNCS) 3664, Springer, Heidelberg, Germany, 2005.
6. M. Agosti and N. Ferro. Annotations as Context for Searching Documents. In F. Crestani and I. Ruthven, editors, *Proc. 5th International Conference on Conceptions of Library and Information Science – Context: nature, impact and role*, pages 155–170. Lecture Notes in Computer Science (LNCS) 3507, Springer, Heidelberg, Germany, 2005.
 7. M. Agosti, N. Ferro, I. Frommholz, and U. Thiel. Annotations in Digital Libraries and Collaboratories – Facets, Models and Usage. In R. Heery and L. Lyon, editors, *Proc. 8th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2004)*, pages 244–255. Lecture Notes in Computer Science (LNCS) 3232, Springer, Heidelberg, Germany, 2004.
 8. M. Agosti, N. Ferro, and N. Orio. Annotating Illuminated Manuscripts: an Effective Tool for Research and Education. In M. Marilino, T. Sumner, and F. Shipman, editors, *Proc. 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005)*, pages 121–130. ACM Press, New York, USA, 2005.
 9. M. Agosti, N. Ferro, and N. Orio. Annotations as a Support to Research Users. In T. Catarci, S. Christodoulakis, and A. Del Bimbo, editors, *Proc. 7th International Workshop of the EU Network of Excellence DELOS on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib’05)*, pages 117–120. Centromedia, Viareggio, Italy, 2005.
 10. M. Agosti and M. Melucci. Information Retrieval Techniques for the Automatic Construction of Hypertext. In A. Kent and C.M. Hall, editors, *Encyclopedia of Library and Information Science*, volume 66, pages 139–172. Marcel Dekker, New York, USA, 2000.
 11. M. Agosti and A. Smeaton, editors. *Information Retrieval and Hypertext*. Kluwer Academic Publishers, Norwell (MA), USA, 1996.
 12. D. Bhagwat, L. Chiticariu, W.-C. Tan, and G. Vijayvargiya. An Annotation Management System for Relational Databases. In M. A. Nascimento, M. T. Özsu, D. Kossmann, R. J. Miller, J. A. Blakeley, and K. B. Schiefer, editors, *Proc. 30th International Conference on Very Large Data Bases (VLDB 2004)*, pages 900–911. Morgan Kaufmann, 2004.
 13. P. Bottoni, R. Civica, S. Levialdi, L. Orso, E. Panizzi, and R. Trinchese. MADCOW: a Multimedia Digital Annotation System. In M. F. Costabile, editor, *Proc. Working Conference on Advanced Visual Interfaces (AVI 2004)*, pages 55–62. ACM Press, New York, USA, 2004.
 14. P. Bottoni, R. Civica, S. Levialdi, L. Orso, E. Panizzi, and R. Trinchese. Digital Library Content Annotation with the MADCOW System. In T. Catarci, S. Christodoulakis, and A. Del Bimbo, editors, *Proc. 7th International Workshop of the EU Network of Excellence DELOS on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib’05)*, pages 111–116. Centromedia, Viareggio, Italy, 2005.
 15. P. Bottoni, R. Civica, S. Levialdi, L. Orso, E. Panizzi, and R. Trinchese. Storing and Retrieving Multimedia Web Notes. In S. Bhalla, editor, *Proc. 4th International Workshop on Databases in Networked Information Systems (DNIS 2005)*, pages 119–137. Lecture Notes in Computer Science (LNCS) 3433, Springer, Heidelberg, Germany, 2005.
 16. P. Bottoni, S. Levialdi, and P. Rizzo. An Analysis and Case Study of Digital Annotation. In N. Bianchi-Berthouze, editor, *Proc. 3rd International Workshop on Databases in Networked Information Systems (DNIS 2003)*, pages 216–230. Lecture Notes in Computer Science (LNCS) 2822, Springer, Heidelberg, Germany, 2003.
 17. H. Brocks, A. Stein, I. Frommholz, U. Thiel, and A. Dirsch-Weigand. How to Incorporate Collaborative Discourse in Cultural Digital Libraries. In S. Handschuh, R. Dieng-Kuntz, N. Collier, and S. Staab, editors, *Proc. ECAI 2002 Workshop on Semantic Authoring, Annotation & Knowledge Markup (SAAKM 2002) at the 15th European Conference on Artificial Intelligence (ECAI 2002)*, pages 21–26, 2002.
 18. P. Buneman, S. Khanna, K. Tajima, and W.-C. Tan. Archiving Scientific Data. *ACM Transactions on Database Systems (TODS)*, 29(1):2–42, March 2004.
 19. P. Buneman, S. Khanna, and W.-C. Tan. Why and Where: A Characterization of Data Provenance. In J. Van den Bussche and V. Vianu, editors, *Proc. 8th International Conference on Database Theory (ICDT 2001)*, pages 316–330. Lecture Notes in Computer Science (LNCS) 1973, Springer, Heidelberg, Germany, 2001.
 20. P. Buneman, S. Khanna, and W.-C. Tan. On Propagation of Deletions and Annotations Through Views. In S. Abiteboul, P. G. Kolaitis, and L. Popa, editors, *Proc. 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2002)*, pages 150–158. ACM Press, New York, USA, 2002.
 21. F. Calonghi. *Dizionario latino-italiano*, 3 edizione (interamente rifusa ed aggiornata del dizionario Georges – Calonghi) – 15 tiratura. Rosenberg & Sellier, Nuova Offito, Mappano, Torino, Italia, 1986.
 22. P. Constantopoulos, M. Doerr, M. Theodoridou, and M. Tzobanakis. On Information Organization in Annotation Systems. In G. Grieser and Y. Tanaka, editors, *Proc. International Workshop on Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets*, pages 189–200. Lecture Notes in Computer Science (LNCS) 3359, Springer, Heidelberg, Germany, 2004.
 23. M. Cortelazzo and P. Zolli. *DELI: Dizionario Etimologico della Lingua Italiana*, 2 edizione a cura di M. Cortelazzo e M. A. Cortelazzo. Zanichelli, Bologna, Italia, 1999.
 24. W. B. Croft. Combining Approaches to Information Retrieval. In W. B. Croft, editor, *Advances in Information Retrieval: Recent Research from the Center for Intelligent Information Retrieval*, pages 1–36. Kluwer Academic Publishers, Norwell (MA), USA, 2000.
 25. D. Fogli, G. Fresta, and P. Mussio. On Electronic Annotation and Its Implementation. In M. F. Costabile, editor, *Proc. Working Conference on Advanced Visual Interfaces (AVI 2004)*, pages 98–102. ACM Press, New York, USA, 2004.
 26. I. Frommholz, H. Brocks, U. Thiel, E. Neuhold, L. Iannone, G. Semeraro, M. Berardi, and M. Ceci. Document-Centered Collaboration for Scholars in the Humanities – The COLLATE System. In T. Koch and I. T. Sølvberg, editors, *Proc. 7th European Conference on Research and Advanced Technology for Digital Libraries*

- (*ECDL 2003*), pages 434–445. Lecture Notes in Computer Science (LNCS) 2769, Springer, Heidelberg, Germany, 2003.
27. I. Frommholz, U. Thiel, and T. Kamps. Annotation-based Document Retrieval with Four-Valued Probabilistic Datalog. In R. Baeza-Yates, Y. Maarek, T. Roelleke, and A. P. de Vries, editors, *Proc. 3rd XML and Information Retrieval Workshop and the 1st Workshop on the Integration of Information Retrieval and Databases (WIRD2004)*, pages 31–38. <http://homepages.cwi.nl/~arjen/wird04/wird04-proceedings.pdf> [last visited 2005, April 5], 2004.
 28. N. Fuhr, P. Hansen, A. Micsik, and I. Sølvberg. Digital Libraries: A Generic Classification Scheme. In P. Constantopoulos and I. T. Sølvberg, editors, *Proc. 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, pages 187–199. Lecture Notes in Computer Science (LNCS) 2163, Springer, Heidelberg, Germany, 2001.
 29. G. Golovchinsky, M. N. Price, and B. N. Schilit. From Reading to Retrieval: Freeform Ink Annotations as Queries. In F. Gey, M. Hearst, and R. Tong, editors, *Proc. 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1999)*, pages 19–25. ACM Press, New York, USA, 1999.
 30. B. Gueye, P. Rigaux, and N. Spyrtos. Taxonomy-Based Annotation of XML Documents: Application to eLearning Resources. In G. A. Vouros and T. Panayiotopoulos, editors, *Proc. 3rd Hellenic Conference on AI – Methods and Applications of Artificial Intelligence (SETN 2004)*, pages 33–42. Lecture Notes in Computer Science (LNCS) 3025, Springer, Heidelberg, Germany, 2004.
 31. S. Handschuh and S. Staab, editors. *Annotation for the Semantic Web*. IOS Press, Amsterdam, The Netherlands, 2003.
 32. P. Hanks, editor. *Collins Dictionary of the English Language*. William Collins Sons & Co.Ltd., Glasgow, UK, 1979.
 33. IEL. *Enciclopedia Italiana di scienze, lettere ed arti*, volume XXIV (SCAR–SOC). Istituto della Enciclopedia Italiana, Istituto Poligrafico dello Stato, Roma, Italia, 1950.
 34. IEL. *Enciclopedia Italiana di scienze, lettere ed arti*, volume XVII (GIAP–GS). Istituto della Enciclopedia Italiana, Istituto Poligrafico dello Stato, Roma, Italia, 1951.
 35. IEL. *Vocabolario della lingua italiana*, volume I (A–C). Istituto della Enciclopedia Italiana, Arti Grafiche Ricordi, Monotipia Olivieri, Milano, Italia, 1986.
 36. IEL. *Vocabolario della lingua italiana*, volume II (D–L). Istituto della Enciclopedia Italiana, Arti Grafiche Ricordi, Monotipia Olivieri, Milano, Italia, 1987.
 37. IEL. *Vocabolario della lingua italiana*, volume III** (Pe–R). Istituto della Enciclopedia Italiana, Arti Grafiche Ricordi, Monotipia Olivieri, Milano, Italia, 1991.
 38. IEL. *Vocabolario della lingua italiana*, volume IV (S–Z). Istituto della Enciclopedia Italiana, Arti Grafiche Ricordi, Monotipia Olivieri, Milano, Italia, 1994.
 39. ISO. MPEG Standards - Coded representation of video and audio. <http://www.iso.ch/iso/en/prods-services/popstds/mpeg.html> [last visited 2005, June 24], July 2004.
 40. J. Kahan and M.-R. Koivunen. Annotea: an open RDF infrastructure for shared Web annotations. In V. Y. Shen, N. Saito, M. R. Lyu, and M. E. Zurko, editors, *Proc. 10th International Conference on World Wide Web (WWW 2001)*, pages 623–632. ACM Press, New York, USA, 2001.
 41. M.-R. Koivunen and R. Swick. Metadata Based Annotation Infrastructure offers Flexibility and Extensibility for Collaborative Applications and Beyond. <http://www.w3.org/2001/Annotea/Papers/KCAP01/annotea.html> [last visited 2005, June 24], September 2001.
 42. M.-R. Koivunen, R. Swick, and E. Prud'hommeaux. Annotea Shared Bookmarks. <http://www.w3.org/2001/Annotea/Papers/KCAP03/annoteabm.html> [last visited 2005, June 24], October 2003.
 43. W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: A Theory of Text Organization. Technical Report ISI/RS-87-190, Information Sciences Institute, University of Southern California, Los Angeles (CA), USA, June 1987.
 44. C. C. Marshall. Annotation: from Paper Books to the Digital Library. In R. B. Allen and E. Rasmussen, editors, *Proc. 2nd ACM International Conference on Digital Libraries (DL 1997)*, pages 233–240. ACM Press, New York, USA, 1997.
 45. C. C. Marshall. Toward an Ecology of Hypertext Annotation. In R. Akscyn, editor, *Proc. 9th ACM Conference on Hypertext and Hypermedia (HT 1998): links, objects, time and space-structure in hypermedia systems*, pages 40–49. ACM Press, New York, USA, 1998.
 46. C. C. Marshall and A. J. B. Brush. From Personal to Shared Annotations. In L. Terveen and D. Wixon, editors, *Proc. Conference on Human Factors and Computing Systems (CHI 2002) – Extended Abstracts on Human Factors in Computer Systems*, pages 812–813. ACM Press, New York, USA, 2002.
 47. C. C. Marshall and A. J. B. Brush. Exploring the Relationship between Personal and Public Annotations. In H. Chen, H. Wactlar, C.-C. Chen, E.-P. Lim, and M. Christel, editors, *Proc. 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2004)*, pages 349–357. ACM Press, New York, USA, 2004.
 48. C. C. Marshall, G. Golovchinsky, and M. N. Price. Digital Libraries and Mobility. *Communications of the ACM*, 44:55–56, 2001.
 49. C. C. Marshall, M. N. Price, G. Golovchinsky, and B.N. Schilit. Introducing a Digital Library Reading Appliance into a Reading Group. In N. Rowe and E. A. Fox, editors, *Proc. 4th ACM International Conference on Digital Libraries (DL 1999)*, pages 77–84. ACM Press, New York, USA, 1999.
 50. C. C. Marshall, M. N. Price, G. Golovchinsky, and B.N. Schilit. Designing e-Books for Legal Research. In E. A. Fox and C. L. Borgman, editors, *Proc. 1st ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2001)*, pages 41–48. ACM Press, New York, USA, 2001.
 51. C. C. Marshall and C. Ruotolo. Reading-in-the-Small: A Study of Reading on Small Form Factor Devices. In W. Hersh and G. Marchionini, editors, *Proc. 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2002)*, pages 56–64. ACM Press, New York, USA, 2002.

52. K. Nagao. *Digital Content Annotation and Transcoding*. Artech House, Norwood (MA), USA, 2003.
53. E. Neuhold, C. Niederée, A. Stewart, I. Frommholz, and B. Mehta. The Role of Context for Information Mediation in Digital Libraries. In Z. Chen, H. Chen, Q. Miao, Y. Fu, E. A. Fox, and E.-P. Lim, editors, *Proc. 7th International Conference on Asian Digital Libraries. Digital Libraries: International Collaboration and Cross-Fertilization (ICADL 2004)*, pages 133–143. Lecture Notes in Computer Science (LNCS) 3334, Springer, Heidelberg, Germany, 2004.
54. R. Overbeek, T. Disz, and R. Stevens. The SEED: A Peer-To-Peer Environment for Genome Annotation. *Communication of the ACM*, 47(11):46–51, November 2004.
55. T. A. Phelps and R. Wilensky. Towards Active, Extensible, Networked Documents: Multivalent Architecture and Applications. In E. A. Fox and G. Marchionini, editors, *Proc. 1st ACM International Conference on Digital Libraries (DL 1996)*, pages 100–108. ACM Press, New York, USA, 1996.
56. T. A. Phelps and R. Wilensky. Multivalent Annotations. In C. Peters and C. Thanos, editors, *Proc. 1st European Conference on Research and Advanced Technology for Digital Libraries (ECDL 1997)*, pages 287–303. Lecture Notes in Computer Science (LNCS) 1324, Springer, Heidelberg, Germany, 1997.
57. T. A. Phelps and R. Wilensky. Multivalent Documents. *Communications of the ACM*, 43(6):83–90, June 2000.
58. T. A. Phelps and R. Wilensky. The Multivalent Browser: A Platform for New Ideas. In E. V. Munson, editor, *Proc. 2001 ACM Symposium on Document Engineering*, pages 58–67. ACM Press, New York, USA, 2001.
59. P. Rigaux and N. Spyrtos. Metadata Inference for Document Retrieval in a Distributed Repository. In M. J. Maher, editor, *Proc. 9th Asian Computing Science Conference – Advances in Computer Science (ASIAN 2004) – Higher Decision Making. Dedicated to Jean-Louis Lassez on the Occasion of His 5th Cycle Birthday*, pages 418–436. Lecture Notes in Computer Science (LNCS) 3321, Springer, Heidelberg, Germany, 2004.
60. L. Rocci. *Vocabolario greco italiano*, 34 edizione. Società Editrice Dante Alighieri, Italia, 1989.
61. B. N. Schilit, M. N. Price, and G. Golovchinsky. Digital Library Information Appliances. In I. Witten, R. Akscyn, and F. M. Shipman, editors, *Proc. 3rd ACM International Conference on Digital Libraries (DL 1998)*, pages 217–226. ACM Press, New York, USA, 1998.
62. F. Shipman, M. N. Price, C. C. Marshall, and G. Golovchinsky. Identifying Useful Passages in Documents based on Annotation Patterns. In T. Koch and I. T. Sølvberg, editors, *Proc. 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*, pages 101–112. Lecture Notes in Computer Science (LNCS) 2769, Springer, Heidelberg, Germany, 2003.
63. A. Spooner. *A Dictionary of Synonyms and Antonyms*. Oxford University Press, New York, USA, 1999.
64. L. D. Stein, S. Eddy, and R. Dowell. Distributed Sequence Annotation System (DAS) – Version 1.53. <http://www.biodas.org/documents/spec.html> [last visited 2005, June 24], March 2002.
65. W.-C. Tan. Containment of Relational Queries with Annotation Propagation. In G. Lausen and D. Suciu, editors, *Proc. 9th International Workshop on Database Programming Languages (DBPL 2003)*, pages 37–53. Lecture Notes in Computer Science (LNCS) 2921, Springer, Heidelberg, Germany, 2004.
66. U. Thiel, H. Brocks, A. Dirsch-Weigand, A. Everts, I. Frommholz, and A. Stein. Queries in Context: Access to Digitized Historic Documents in a Collaboratory for the Humanities. In M. Hemmje, C. Niederée, and T. Risse, editors, *From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments. Essays Dedicated to Erich J. Neuhold on the Occasion of His 65th Birthday*, pages 117–127. Lecture Notes in Computer Science (LNCS) 3379, Springer, Heidelberg, Germany, 2005.
67. U. Thiel, H. Brocks, I. Frommholz, A. Dirsch-Weigand, J. Keiper, A. Stein, and E. J. Neuhold. COLLATE – A collaboratory supporting research on historic European films. *International Journal on Digital Libraries*, 4(1):8–12, August 2004.
68. W3C. EMMA: Extensible MultiModal Annotation markup language – W3C Working Draft 14 December 2004. <http://www.w3.org/TR/emma/> [last visited 2005, June 24], December 2004.
69. W3C. Annotea Project. <http://www.w3.org/2001/Annotea/> [last visited 2005, June 24], March 2005.
70. W. Xi, B. Zhang, Z. Chen, Y. Lu, S. Yan, W.-Y. Ma, and E. A. Fox. Link Fusion: A Unified Link Analysis Framework for Multi-Type Interrelated Data Objects. In S. Feldman, M. Uretsky, M. Najork, and C. Wills, editors, *Proc. 13th International Conference on World Wide Web (WWW 2004)*, pages 319–327. ACM Press, New York, USA, 2004.