An Analysis of Cross-Document Linking Mechanisms

Ahmed A.O. Tayeh and Beat Signer
Web & Information Systems Engineering Lab
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels, Belgium
atayeh,bsigner@vub.be

ABSTRACT

Physical and digital documents do often not exist in isolation but are implicitly or explicitly linked. Previous research in Human-Computer Interaction and Personal Information Management has revealed certain user behaviour in associating information across physical and digital documents. Nevertheless, there is a lack of empirical studies on user needs and behaviour when defining these associations. In this paper, we address this lack of empirical studies and provide insights into strategies that users apply when associating information across physical and digital documents. In addition, our study reveals the limitations of current practices and we suggest improvements for associating information across documents. Last but not least, we identify a set of design implications for the development of future cross-document linking solutions.

CCS CONCEPTS

• **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **Pointing**; **User centered design**; **Empirical studies in interaction design**;

KEYWORDS

Cross-document linking; linking mechanisms; user linking behaviour; reading and writing; hypertext; annotation

1 INTRODUCTION

As already mentioned by Vannevar Bush in 1945, documents do not exist in isolation but are often related to other documents [6]. This has also been confirmed by HCI research investigating the user behaviour while reading and writing physical and digital documents. For example, Marshal [11] stated that users use highlights or underlines as anchors while reading to explicitly associate (link) information within a single document. Some other research in the domain of Personal Information Management (PIM) identified that users use digital and physical folders for organising and relating (associating) information across documents [9].

A number of solutions have been proposed in order to help users in associating information within and across documents. For instance, the majority of recent digital document formats support simple forms of mostly unidirectional linking features. Furthermore, various open hypermedia and annotation solutions have been introduced to allow users to link and annotate documents in order to create explicit associations. Moreover, some recent solutions [2, 5] enable the use of annotations for associating information across documents.

Even though there exist some general findings regarding the user behaviour when linking information within and across documents, to the best of our knowledge there are no studies aiming to understand a knowledge worker's behaviour when associating information across physical as well as digital documents. As a result of the lack of such empirical studies, the development of most existing linking systems has been solely motivated by limitations of their predecessors or new features to be supported, rather than on a clear understanding of the needs and requirements of end users [17]. We believe that only a better understanding of the user behaviour in linking and associating documents as well as the user requirements for cross-document linking will lead to more advanced and usable linking solutions. In this paper, we therefore provide a detailed investigation of user needs and behaviour when associating information across physical and digital documents. Our approach is based on a multi-case design approach consisting of an online survey and interviews with some of the participants of the online survey in order to investigate their cross-document linking behaviour as well as their feedback about existing solutions. The insights that we gained from the presented user study further enabled us to formulate a number of design implications for future cross-document link services.

We begin in Section 2 by highlighting previous research results. In Section 3, we discuss the methodology used for our study, whilst the results of the study are presented in Section 4. In Section 5 we critically discuss the study results and provide a number of design implications for future cross-document link services, before providing some concluding remarks in Section 6.

2 BACKGROUND

There have been a number of studies in different domains revealing some interesting findings regarding the user behaviour when associating information within and across digital and physical documents. In the digital world, some studies have stated that system folders, which are used to organise documents, are a way of associating "entire" documents [9]. Besides the use of folders, previous studies have shown that users make use of annotations in order to create associations between parts of documents. In the physical world, the filing, piling and mixing organisational strategies [10, 19] are used to organise relevant documents.

Previous studies in HCI that aimed to understand and analyse a user's reading and writing activities have stated that users perform a *cross-document referencing* task in order to integrate and associate information from one or multiple documents. According to Adler et al. [1], this cross-document referencing task forms a major part of the reading and writing activity. The work by O'Hara et al. [13, 14] revealed that readers of digital or physical documents tend to use different kinds of annotations (e.g. marginal notes or underlines) to define references between documents. In her work 'Toward an Ecology of Hypertext Annotation' [11], Marshall stated that some user annotations in physical books are an emulation of *hypertext patterns*, such as creating and referring to anchors within a book. For example, readers used highlights or underlines as anchors in a printed book in order to refer to them from another part of the same book. They also tend to use annotations to explicitly make references to entire chapters or sections within the same book.

The literature within the field of Personal Information Management is mainly focusing on how people organise and retrieve information artefacts. Despite the fact that Whittaker et al. [20] have noted the lack of empirical research in PIM, there are some interesting findings that are relevant in the context of our research. Jones et al. [9] stated that folders are used to summarise, organise and associate information that is relevant for a specific user task (e.g. a planned project or some course material). Another study by Boardmann et al. [4] revealed that some users use a consistent folder naming convention for relating resources to each other. According to the same study, users create so-called *overlapping folders* (folders with the same name) by using different tools (e.g. Outlook and the file system) in order to organise resources that are related to the same production activity. A user might, for instance, create a folder named 'marriage' in an email client to store all emails concerning their marriage plan and at the same time create a system folder with the same name for storing other related documents.

We believe that the existing findings regarding the user behaviour in associating information across documents are not sufficient for building future linking solutions. In the presented user study we therefore thoroughly investigated the user behaviour in linking and associating information across physical and digital documents.

3 METHODOLOGY

We have chosen a qualitative approach using a multi-case design to explore different users' cross-document information association behaviour [21]. Our study relies on a mixed methods approach consisting of an online survey combined with interviews with participants of the online survey. The participants of our study were informed that the collected data would be used for scientific research as well as in scientific publications. Furthermore, they were ensured that their data would be treated confidentially and be fully anonymised if used in any publication.

3.1 Data Collection

3.1.1 Online Survey. For the purpose of this study, we designed an online questionnaire (for details about the questions see [17]) which focuses on investigating whether users associate information in both, the physical and digital media, which mechanisms they use to associate information, why they create these associations and whether they are happy with the mechanisms that they currently use to define these associations. Our online survey allowed us to collect data from a much larger number of participants than it would have been feasible through alternative methods such as observations or think aloud experiments [21].

In order to identify the reasons for associating information as well as potential problems, we first needed to find out whether users had ever felt the need to link or associate information in a particular way (further referred to as a scenario). A negative reply to this question implies that it is rather unlikely that a user has ever engaged in this type of linking activity, whereas a positive reply indicates that they are likely to have engaged in associating information or experienced an inability to do so. In the case of a negative reply, the survey therefore immediately moved on to the next scenario. In case of a positive answer, participants were asked for more information about their behaviour in associating information or any difficulties preventing them from creating associations.

The survey contained both open-ended questions as well as quantitative questions using a 5-point Likert scale. The open-ended questions enabled participants to freely describe their previous activities in associating information within and across documents. It is worth mentioning that some of the survey questions investigated the information association mechanisms that have already been identified by previous research, including the use of digital folders and annotations. The survey was conducted using the LimeSurvey[1] online tool and contained four groups of questions:

(1) Questions related to a participant's demographic information such as gender, country of residence, education and age;
(2) Questions related to a participant's behaviour in associating information in the physical space. These questions differentiated between two main scenarios:
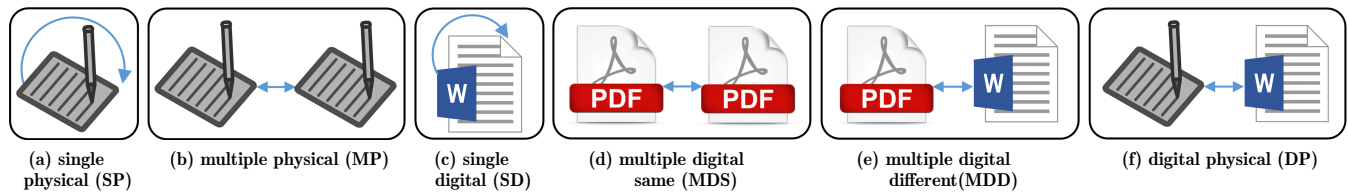
---

[1]https://www.limesurvey.org

Figure 1: Different scenarios of linking and associating information

- Associating information within a **s**ingle **p**hysical document (**SP**) (e.g. between two different sections or chapters of a printed book)
- Associating information across **m**ultiple **p**hysical documents (**MP**). An annotation in a printed document that declares the existence of a relationship to (parts of) another printed document is an example for this scenario.

(3) Questions about a participant's behaviour in associating information within and across digital documents. The questions differentiated between three main scenarios:
  - Associating information within a **s**ingle **d**igital document (**SD**);
  - Associating information across **m**ultiple **d**igital documents of the **s**ame document type (**MDS**);
  - Associating information across **m**ultiple **d**igital documents of **d**ifferent document types (**MDD**).

(4) Questions about a participant's behaviour in associating information across **d**igital and **p**hysical documents (**DP**).

Note that the scenario names and their abbreviations (illustrated in Figure 1) are extensively used in the remainder of this paper. At the end of the survey, the participants were invited to upload any screenshots or images of associations they have created in the past and we received 15 images in total. Finally, the participants were also invited to provide their email address if they were willing to potentially extend their participation in the form of an interview. As an incentive, participants who provided their email address could win a 30 Euro Amazon voucher.

3.1.2 Interviews. In addition to the online survey, we interviewed some of the survey participants in order to further investigate their answers and gain detailed insights about their information association behaviour. We used unstructured open questions to obtain descriptions of interesting mechanisms for associating information or clarifications of vague answers given during the survey. The interviews were performed either face-to-face or via Skype, lasted 15 minutes on average and were recorded. In addition to the recordings, the researchers also collected some notes during the interviews. The interviews and notes were later transcribed in a Word document in order to facilitate the data analysis.

3.2 Population

The link to our online survey has been distributed internationally via email to researchers from various research fields.

Given the focus of our study, we chose to recruit participants from a population of researchers as they represent a group of *knowledge workers* who frequently use documents and can be expected to engage in document linking. It is worth mentioning that other user groups such as secretaries could also be considered as knowledge workers since they frequently use documents. However, we believe that researchers are more engaged in associating information especially when reading and writing scientific articles. In total, 238 people completed the survey. Our sample includes Master's students ($n = 23$), PhD students ($n = 169$) and researchers holding a PhD degree ($n = 46$). The 238 participants consisted of 82 female and 156 male participants, ageing from 21 to 60 years. While 97 participants have provided us their email addresses, we only selected 12 participants for a follow-up interview based on their answers. Note that given the population, the results of our study might be generalised for the community of researchers but not necessarily for other knowledge workers.

3.3 Data Analysis

The collected quantitative data was analysed using descriptive statistics, while the qualitative data was analysed using an informal coding. During the qualitative analysis, the written comments of participants, the notes of the interviewer as well as the received images and screenshots were all taken into account. First, the qualitative data was carefully checked in search of common association mechanisms used by the participants. Based on this assessment, a list of association mechanisms was compiled for every scenario presented to the participants (i.e. SP, MP, SD, MDS, MDD and DP). Second, we identified the characteristics and limitations of each mechanism and calculated how many participants applied that particular mechanism.

4 RESULTS

Our study shows that *most knowledge workers are either occasionally or frequently associating information across documents* as highlighted in Figure 2. Users are associating information during both, reading and writing activities, which confirms the findings presented in [1, 9, 13]. As shown in Table 1, users apply different linking mechanisms to associate information in the different scenarios. Please note that in Table 1, association mechanisms that have been already identified in previous research are written in normal font, whereas the mechanisms that we have discovered are emphasised in

Mechanism		SP	MP	SD	MDS	MDD	DP
Annotations & highlights	B1	24.2%	28.7%	35.8%	30.2%	28.9%	21.7%
	B2	30.5%	29.4%	27.6%	20.1%	19.3%	24.8%
	U1	31%	22.6%	35%	45.6%	48.2%	25.5%
	U2	31%	42.4%	34.3%	33.5%	34.4%	35.6%
Line & arrow drawings		11.8%	—	—	—	—	—
Common symbols		2.2%	—	—	—	—	—
Separate documents		2.2%	7.5%	3.3%	12.7%	12.4%	4.6%
Physical folders		—	3.4%	—	—	—	—
Post-it notes		1.1%	1.3%	—	—	—	0.7%
Physical counterparts		—	—	2.8%	3.3%	3.4%	3.8%
Copy & paste		—	—	2.2%	1.3%	1.3%	—
Digital folders		—	—	—	63%	66%	—
External applications		—	—	—	3.3%	2.7%	—
Writing physical parts into a digital document		—	—	—	—	—	2.3%
Scanning physical documents		—	—	—	—	—	3.1%

Table 1: Linking mechanisms used in the different information association scenarios

italics. Further, the percentages presented in Table 1 are relative to the number of participants who associate information in the given scenario as illustrated in Table 2.

4.1 Linking Information in Physical Documents

As shown in Table 2, many participants (74.3%) indicated that they associate information in a single physical document, with 11 participants (6.2%) doing this very frequently (see Figure 2). In the case of MP, 61.3% of the participants indicated that they have faced situations where they had to associate information across two or more physical documents, with only 6 of them (4.1%) doing this very frequently. This difference in terms of frequency might be caused by various factors. As discussed later, some users apply simple and trivial mechanisms to associate information in SP that cannot be applied in MP (e.g. drawing a simple line between the associated parts). Furthermore, sometimes the closeness of the associated parts in SP (e.g. information on the same page) helps users to easily associate information.
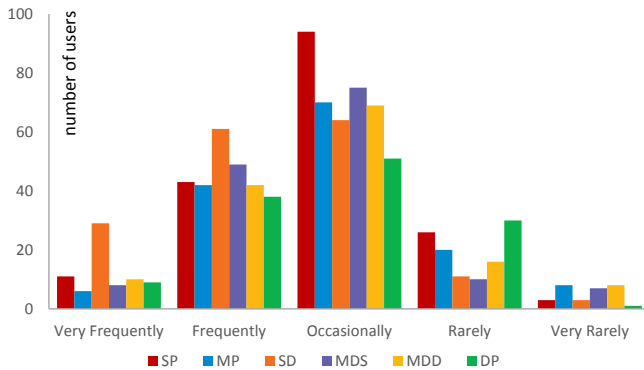

Figure 2: Linking frequency in the different scenarios

Our investigation revealed that participants are applying different mechanisms to create associations between pieces of information in both physical scenarios. Most of the participants use annotations (e.g. comments, arrows or highlights) in order to explicitly associate information. This seems to confirm the findings presented in [11, 13, 14]. The use of annotations as well as highlights enables participants to *establish associations at any level of granularity* since participants are able to highlight and annotate any part of a document such as a chapter, a section, a paragraph or a single word. Annotations and highlights also yield to different types of *bidirectional* and *unidirectional* associations.

Scenario	Size of the subsample relative to the total number of participants
SP	74.3%
MP	61.3%
SD	70.5%
MDS	62.6%
MDD	60.9%
DP	54.2%

Table 2: Relative size of the subsample of participants who associate information in a particular scenario

Our study identified two main types of bidirectional associations resulting from using annotations and highlights as illustrated in Figure 3. For the first type (B1), the bidirectionality is established by highlighting the different parts and writing annotations with references to each other next to these parts. For the second type (B2), participants only write annotations next to each part without any highlighting. In a bidirectional association, participants write down references to all the linked parts in order to help them later in information refinding tasks. The B1 associations mimic the creation of hypertext anchors since participants explicitly highlight the exact text anchors to be linked with each other. This confirms Marshall's findings [11] that participants create anchors in a printed book in order to refer to them from another part in the same document.

In the SP scenario, B1 associations are used by 24.2% of the participants, whereas 28.7% use them in the MP scenario.
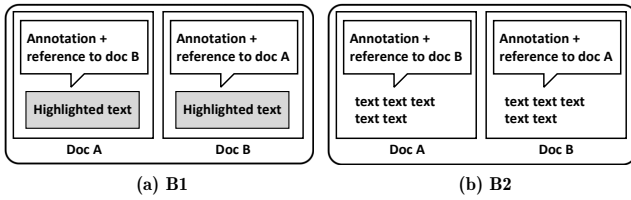
Figure 3: Two main types of bidirectional associations using annotations and highlights

The B2 associations are used by 30.5% of the participants in the SP scenario and by 29.4% in the MP scenario. We also discovered other interesting bidirectional association patterns. One important pattern emerges in a single document scenario where the parts are close to each other (e.g. on the same page). In this case, 11.8% of the participants indicated that they would normally draw arrows or lines between the associated parts. Moreover, four participants (2.2%) indicated that they use common symbols or numbers in order to match the different pieces in the single document scenario. The use of line drawings and common symbols as association mechanisms illustrates the flexibility in the SP scenario.

We further discovered that many participants are creating *unidirectional associations* between the intended pieces of information. Two main types of unidirectional associations resulting from using annotations and highlights were identified. As illustrated in Figure 4, in the first type (U1), the unidirectionality is established by highlighting the different parts and by writing annotations next to one of the parts. In the second type (U2), annotations are written next to one of the parts without any highlighting. The U1 associations are used by 31% of the participants in scenario SP and by 22.6% of the participants in scenario MP. U2 associations are used by 31% of the participants in the SP scenario and by 42.4% in the MP scenario. The collected data from the survey and the interviews shows that participants prefer to have bidirectional associations due to the advantages in re-finding the linked parts.
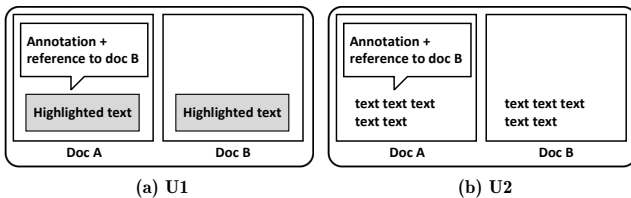

Figure 4: Two main types of unidirectional associations using annotations and highlights

Another interesting finding is that numerous participants tend to use a *separate document* in order to explicitly create associations, in particular in the MP scenario. In the SP scenario, 2.2% of the participants indicated that they use a

separate document (digital or physical) in order to note, summarise or copy and paste the different parts. On the other hand, in the MP scenario, 7.5% of the participants are using a separate document to explicitly indicate the association.

Other association mechanisms applied by participants include the use of physical folders, ring binders as well as paper stacks in order to associate information across multiple physical documents (MP). These mechanisms result in the filing and piling organisational strategies that are well known in the PIM literature [10]. Finally, the use of post-it notes was mentioned for associating information in both, the SP and MP scenarios.

4.2 Linking Information in Digital Documents

In general, most participants are applying numerous mechanisms to associate information in the different digital scenarios. From the participants who associate information, 7.7% in the SD scenario, 3.3% in the MDS scenario and 3.4% in the MDD scenario have indicated their inability to create associations between the intended parts. In general, the lack of a suitable linking tool as well as the necessary effort to create associations between the intended parts are the main reasons that prevented participants from creating these associations.
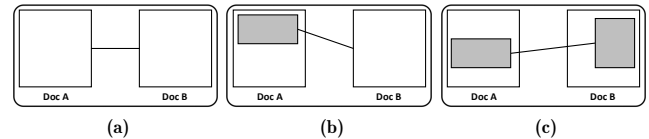

Figure 5: Associations between digital documents are created to link entire documents (a), parts of a document with an entire document (b) or different parts of the documents (c)

As illustrated in Figure 5, associations across different digital documents are created at *different levels of granularity* to link entire documents (a), parts of a document with another document (b) or different parts of two documents (c). In the context of the MDS scenario, 75.1% of the participants who associate information create associations between the different parts of documents (c). In the same scenario, 57% of the participants associate information between parts of a document and another document (b) and 48.3% of the participants create associations between entire documents (a). In the MDD scenario, 46.2% of the participants associate information between entire documents (a), 80% between parts of different documents (c) and 61.3% between parts of a document with another document (b).

4.2.1 Single Digital Document. In the SD scenario, 70.5% of the participants indicated that they have faced situations where they associated information within a single document. While 82.3% of those participants use annotations as well as highlights to explicitly create the associations, another 10% of the participants have adopted various other interesting mechanisms to create the associations between the different parts.

As mentioned earlier, the remaining 7.7% of the participants are not able to create the intended associations.

The annotation features provided by the different document viewers and applications such as Foxit Reader[2] have enabled the majority of the participants to create the associations using annotations and highlights not only in the SD scenario but also in the other digital MDS and MDD scenarios. This is an interesting finding since it contradicts to some extent some findings of a previous study carried out by O'Hara et al. [13] which compared participants' reading activities of physical and digital documents. In their study, the majority of participants did not prefer to annotate digital documents. According to O'Hara et al., the difficulty of annotating documents and the inflexibility of interaction techniques via mouse and keyboard were the main reasons that prevented participants from annotating digital documents. The discrepancy between our findings and these earlier study findings might be explained in different ways. For example, in the study of O'Hara et al., participants were obliged to use a specific document viewer (Microsoft Word 6.0) for reading their documents and some participants might not have been familiar with the offered annotation features.

By using the annotations and highlights mechanism to create associations, participants tend to create bidirectional and unidirectional associations between the different parts. Out of the 82.3% of the participants who use annotations and highlights, 35.8% create bidirectional associations of type B1 and 27.6% of type B2. Unidirectional associations of type U2 are established by 34.3%, whereas U1 associations are used by 35% of the participants.

As mentioned before, other interesting association mechanisms are used. Various users (3.3%) tend to use separate documents to summarise or copy and paste the associated parts. According to numerous participants, *"the new document that was used to describe the different associations served as a database of related information or as a starting point to remember some content of the linked documents at a later stage"*. Five participants (2.8%) mentioned that they prefer to read printed versions of the documents and associate information in physical counterparts. Two of them have explicitly indicated the flexibility in annotating physical documents compared to digital documents and four participants (2.2%) mentioned that they *"copy one part and paste it next to the other"*.

4.2.2 Different Documents of the Same Document Type.
In this scenario, 62.6% of the participants have adopted various mechanisms to create associations between pieces of information in different documents of the same document type. These mechanisms include storing the linked documents in the same digital folder, using annotations and using separate documents for describing the intended associations.

Participants are mainly using the digital folder mechanism (63%) in order to create associations between entire documents. This confirms the finding of Jones et al. [9] stating that participants use folders to organise related information.

It is worth mentioning that previous research has shown that traditional digital hierarchical folder structures are ineffective and cognitively demanding [3].

As in the previous scenarios, unidirectional and bidirectional associations are resulting from the use of annotations and highlights. Unidirectional associations of type U1 are created by 45.6% of the participants while the ones of type U2 are applied by 33.5%. Bidirectional associations of type B1 are established by 30.2% and of type B2 by 20.1% of the participants.

Some other mechanisms are applied by 22.6% of the participants. The use of a separate document to associate information is adopted by quite a large percentage of the participants with 12.7% of the participants using separate documents to summarise, describe or copy and paste the parts.

The use of the physical counterpart mechanism has been applied by five participants (3.3%). They preferred to associate information in the printed version of the intended documents. *"I prefer to work on printed material which I can also easily sort physically in addition to storing the files in the same folders"*, one participant justified their strategy. Three other participants (2%) rely on their memories to remember associations but according to them, over time they tend to forget their associations.

Five of the participants (3.3%) have indicated the use of some external applications (e.g. Evernote[3] and Zotero[4]) to create associations between entire documents of the same digital type. Bibliography reference managers such as Zotero are used by participants to make associations between entire documents in order to facilitate the creation of citations and bibliography while writing scientific reports. Most of the interviewees (12) did not indicate the use of bibliography reference managers in the online survey. Nevertheless, in the interviews, all of them confirmed the use of these systems during scientific writing activities. Further, most of the interviewees did not consider the bibliography reference managers as association and linking tools.

4.2.3 Different Documents of Different Types.
In this scenario, 60.9% of the participants associate information across two or more documents. As in the previous scenario, storing associated documents in the same folder is the dominant mechanism adopted by 66% of the study participants.

By using annotations and highlights, unidirectional associations are established more often than bidirectional associations. The U1 associations are created by 48.2% of the participants whereas U2 associations are established by 34.4% of the participants. 28.9% of the participants create B1 associations whereas 19.3% establish B2 associations.

Another 20.8% of the participants are applying various other mechanisms similar to the ones presented earlier for the MDS scenario. They form a subset the participants who have mentioned the use of various association mechanisms other than the folder and annotations mechanisms in the previous MDS scenario.

---

[2]https://www.foxitsoftware.com/pdf-reader/

[3]https://evernote.com

[4]https://www.zotero.org

Mechanism	Directionality		Granularity	
	Unidirectional	Bidirectional	Entire document	Any level
Annotations & highlights	✓	✓	✓	✓
Line & arrow drawings	✗	✓	✗	✓
Common symbols	✓	✓	✓	✓
Separate documents	✓	✓	✓	✓
Physical folders	✗	✗	✓	✗
Post-it notes	✓	✓	✗	✓
Physical counterparts	✓	✓	✓	✓
Copy & paste	✓	✓	✗	✓
Digital folders	✗	✗	✓	✗
External applications	✗	✗	✓	✗
Writing physical parts into a digital document	✓	✗	✗	✓
Scanning physical documents	✗	✗	✓	✓

Table 3: Characteristics of the resulting associations

4.3 Linking Physical and Digital Documents

More than half of the participants (54.2%) indicated that they have associated information across digital and physical documents. The collected data reveals that participants are using different mechanisms to associate information across digital and physical documents. By using annotations and highlights, they tend to create either unidirectional or bidirectional associations. Out of the 54.2% of the participants who associate information in the DP scenario, 21.7% tend to create bidirectional associations of type B1. Furthermore, bidirectional associations of type B2 are created by 24.8% of the participants. Unidirectional associations of type U1 are created by 25.5% and of type U2 by 35.6% of the participants.

Another 16.3% of the participants mentioned the use of a wide range of mechanisms to associate the intended parts. Some of them (3.8%) indicated that they tend to print the digital document in order to have flexibility in annotating and archiving. Other users (2.3%) prefer to write the information of the physical part into the digital document and 4.6% indicated the use of the separate document mechanism to associate information across digital and physical media. While one of them uses a separate physical document, the rest use separate digital documents. Four other participants (3.1%) mentioned that they scan the physical document or take a photo of it with their smartphones. The scanned document or photo is then stored in the same digital folder as the digital document. It is obvious that those four participants are also using the same folder strategy in combination with the scanning mechanism. Finally, one participant (0.7%) mentioned the use of post-it notes on physical documents for creating associations.

4.4 Details of Information Linking Mechanisms

Most of the association mechanisms identified by our study are a result of different work practices. Moreover, each of them produces different types of associations (e.g. unidirectional or bidirectional). Furthermore, participants have different levels of satisfaction regarding their adopted association mechanisms.

4.4.1 Characteristics of the Links. We have already discussed some characteristics (bidirectionality and unidirectionality) of associations resulting from the use of annotations and highlights in the different scenarios. Two important aspects should be kept in mind while investigating the nature of an association; *the granularity of the associated parts* (e.g. fragments or entire documents) as well as *the attached references to the associated parts*. A bidirectional association enables a user to navigate from a source to a target document and vice versa. On the other hand, a unidirectional association only allows them to navigate from a source to a target document. Finally, some associations, such as the ones created by the folder mechanism, do not imply any traversal. A summary of the characteristics of the resulting associations is provided in Table 3.

The use of the annotations and highlights mechanism enables users to create all types of associations at any level of granularity. Annotations written next to two linked parts in a bidirectional association should contain references to each other in order to enable a user to easily navigate between the linked documents. If annotations in one linked part do not contain reference(s) to the other part(s), the created association is categorised as a unidirectional association. Different document viewers enable users to write annotations at any place in a document. Users can further highlight almost any part of a document and thereby associate information at the word, paragraph, section or chapter level as well as define links between entire documents.

The drawing mechanism (i.e. lines and arrows) that is used to connect parts of close-by information produces bidirectional associations. Users are able to see all endpoints of the resulting associations. The fact that "most" of the drawings are attached to a single document limits the possibilities to create associations at some levels of granularity. With the drawing mechanism users can, for example, not establish associations between two chapters of a document. As with drawings, the common symbols mechanism would normally produce bidirectional associations. In contrast to the drawing mechanism, common symbols support the linking at any level of granularity. Users are able to associate different chapters,

sections or even entire documents with the common symbols as long as they attach the right references to the linked parts.

The use of separate physical or digital documents for associating information enables the creation of unidirectional and bidirectional associations at any level of granularity. Some users write the correct references to all the linked documents which results in a bidirectional association. On the other hand, a unidirectional association is established by not writing a reference to one of the linked parts. Note that some users create associations between entire documents by summarising all the linked documents or by only writing down the titles of the linked documents.

The use of the system folder as well as bibliography reference managers enables users to create undirected associations between entire documents. The traversal between documents is not defined for both of these mechanisms. A system folder visualises its documents in a way (e.g. a list) that enables users to navigate to any document. Most bibliography reference managers do not enable the navigation between documents but from a document to the corresponding system folder.

The associations resulting from physical archives (folders) as well as the scanning mechanism have more or less the same characteristics as the associations resulting from the use of system folders. The scanning of documents implies that the linked documents are stored in the same folder, whereas the system folders are an emulation of the physical archives. The use of the printed versions of digital documents mechanism to associate information implies the use of annotations or physical archives for associating information. Thereby, the associations have similar characteristics as the associations resulting from using annotations or physical archives. Last but not least, the associations resulting from the copy and paste, the writing of physical parts into digital documents and the post-it notes mechanisms depend on the recording of references to the associated documents as well as the granularity of the information part that has been noted down or copied.

4.4.2 User Satisfaction with the Used Linking Mechanisms.
Figure 6 summarises the study participants' satisfaction with their used mechanisms in the three different digital scenarios as well as across digital and physical documents. Note that the percentages presented in Figure 6 are relative to the number of participants associating information in the corresponding scenario (see Table 2). It is clear that many participants (57.7%) are satisfied with the mechanisms they use for associating information in a single document (SD). In contrast to the SD scenario, the majority of participants are not happy or uncertain about the way they create associations in the other three scenarios. For the MDS scenario, 31.6% of the participants have indicated that they are not satisfied with the used mechanisms and 32.2% are uncertain about the way they create associations. In the MDD scenario, 40% and 28.3% are not satisfied or uncertain, respectively. Finally, in the DP scenario, 59.7% are not satisfied with the used mechanisms.
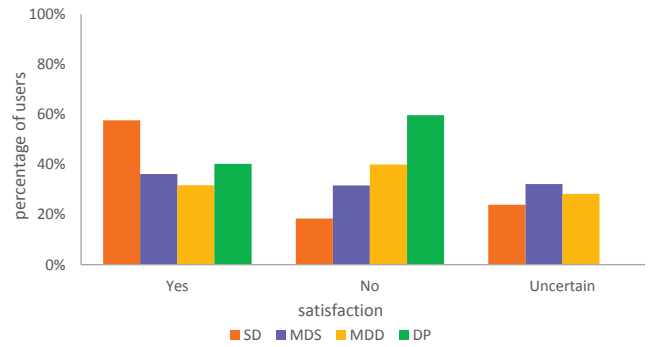

Figure 6: User satisfaction when associating information in digital documents and across digital and physical documents

Participants have given us numerous reasons for not being satisfied with their currently used association mechanisms. Some participants, in particular those reporting difficulties to create associations in different scenarios, have given us general complaints such as: *"I cannot create it"* and *"It is not intuitive how to make the associations"*.

In the context of the different digital scenarios, most participants who use annotations and highlights to associate information complained about the *"wasted time in creating and refinding an association"* between documents. Many of the participants who adopted the organisation of related documents within system folders have given numerous reasons for not being satisfied with this mechanism. Some of their complaints were: *"common folder is not the most accurate"* and *"the structure [of the folders] is unclear, it is not organised well enough, causing confusion and waste of time"*. Last but not least, participants who use separate documents as well as the copy and paste mechanism were mainly concerned about the wasted time.

In the DP scenario, participants were complaining about the time wasted in creating associations as well as the retrieval and tracking of the linked documents. We think that most of the used association mechanisms in the DP scenario require an effort from users in creating annotations in the two different media (digital and physical) or scanning and archiving documents.

4.4.3 The Need for a Linking Tool.
After asking the survey participants about potential issues with their used mechanisms, they were questioned whether there is a need for a tool or a mechanism to facilitate the linking (associating) of information in the different scenarios. Most participants (179) indicated a need for a tool that easily supports the creation of associations between parts of digital documents as well as the navigation between these document parts.

In the digital scenarios, 25.1% out of the 179 participants shared their thoughts and suggestions for a suitable tool for associating information. 35 participants emphasised that a tool must support the creation of hyperlinks. One of them mentioned: *"It comes natural to me that such a system would resemble the mechanism of the Web"*. Some participants were

concerned about the usability of any future linking tool: *"It should be extremely quick and easy"*.

Some participants were concerned about the inclusion of essential features in any future linking tool and two of them about the integration of the linking mechanism in their workflow. One participant mentioned: *"It would be dreamable if the linking mechanism would be available in the native document viewing program, and not that we are forced to use another application to link documents"*. Another participant who complained about the searching and refinding of associations stated: *"[...] allowing me to link and then search for the associations"*.

In the DP scenario, 20% out of the 59.7% who are not satisfied with the used mechanisms have shared their thoughts and suggestions for a better linking mechanism across digital and physical documents. One of the participants who is familiar with the recent Anoto[5] digital pen and paper technology suggested to exploit this technology in realising future cross-media linking solutions. Other participants suggested the use of QR codes and RFID tags. Another participant suggested a totally different idea of *"automatically scanning physical documents and do text recognition on them. Then collect keywords from them to organise them on the computer together with the digital files"*.

4.4.4 User Work Practices. Most of the interviewees have mentioned the use of multiple documents while reading and writing, not only in the physical space but also in the digital space. The use of multiple screens for multi-document viewing and associating information is a very common work practice for most of the interview participants. These findings are consistent with previous research findings [13]. According to the study participants, the simultaneous use of multiple documents enables them to easily annotate related information or to summarise it in a separate document.

The annotation features of different document viewers are an important manner to establish associations by using annotations and highlights. Nevertheless, as raised by different participants, while users are able to create different kinds of annotations within their documents, they are not offered the necessary functionality to search, re-find or manage their annotations. According to various participants, most of the time they are forced to open their documents and search within them to recall and find the previously created associations.

Another interesting work practice that we have discovered is the limited use of bibliography reference managers in reading activities. Participants are mainly using these applications for some scientific writing activities. We discovered that most users are not aware of the many interesting features (e.g. to relate documents to each other) offered by these applications. Indeed, as most of the interviewees have indicated, these systems are not primarily linking tools, but they can be seen as a layer on top of the system folder where users can group documents of different system folders into one unified collection.

---

[5]http://www.anoto.com

Last but not least, recent advancements in the resolution of embedded smartphone cameras enabled various participants to scan and take photos of physical documents while associating information in the DP scenario. According to two interviewees, this practise does not require much effort in taking the photo, but it needs an effort for archiving and storing the photo. According to one of them, the required effort depends on the installed smartphone applications as well as the synchronisation of the smartphone with a user's desktop or laptop computer.

5 DESIGN IMPLICATIONS

Our study revealed a number of interesting and important findings. Participants have adopted various mechanisms to associate information in digital and physical space at any level of granularity, including the document, chapter, section, paragraph or even word level.

In most of the investigated scenarios, users are mainly using annotations, folders and separate digital or physical documents to explicitly create the associations between documents. Some of the association mechanisms used in the physical space (SP and MP) are emulated with or similar to other mechanisms in the digital space (SD, MDS and MDD). The use of a physical folder mechanism (e.g. filing cabinets) in the physical space is, for example, emulated via the digital folder mechanism in digital space. The use of separate documents to associate information is a common mechanism for both, digital and physical media.

In addition, our study shows that users tend to create different types of associations (i.e. bidirectional, unidirectional and undirected associations) between associated parts. Nevertheless, regardless of the nature of the created associations, most users are not satisfied with their used mechanisms. In the following, we outline a number of design implications (D1-D5) for future linking solutions, that we have derived from the collected data as well as from issues and suggestions mentioned by the study participants.

D1: Granularity of the associations: As mentioned earlier, the associations between documents exist at any level of granularity and a user should therefore be able to establish "hyperlinks" at any level of granularity. The fact that document formats have different logical structures and document models (e.g. tree or linear structures) [8] should not create any barrier in supporting this feature in a future linking solution and an application should rather support different fragment identifiers (selectors or anchors) for different document formats.

D2: Bidirectional associations: Many users prefer bidirectional associations over unidirectional associations. A future association and linking tool should not only enable the creation of bidirectional associations, but also allow users to seamlessly navigate these associations in both directions. We believe that users who are used to create unidirectional or undirected associations will not complain about the support of bidirectional hyperlinks after experiencing their advantages. Unidirectional hyperlinks are often criticised since

they cannot be navigated in both directions [16, 18]. An update or delete operation on one endpoint of a unidirectional hyperlink normally produces a broken or dangling hyperlink. Bidirectional hyperlinks are less exposed to such inconsistencies when updating or deleting endpoints since an application can ensure the consistency of both endpoints.

D3: Side-by-side documents: The use of multi-document viewing enables users to easily associate information across documents. It is worth mentioning that many systems have been built based on the side-by-side reading and annotation of documents [7]. A future linking tool should enable the side-by-side visualisation of documents to offer users some flexibility in creating "hyperlinks" between the visualised documents by using simple drag-and-drop interactions.

D4: Linking across digital and physical: It is not sufficient to build applications that support the associating of information in digital media only but we should think about possibilities to support the information integration and association across physical and digital media [12, 15] based on digital pen and paper or other technologies. A future linking application should overcome the limitations of existing cross-media applications by supporting the seamless integration of printed materials and arbitrary digital media.

D5: Management of the associations Giving users the possibility to create associations between different documents should come along with the possibility to manage their associations. As discussed before, a future linking solution should have an integral management component that helps users in managing and searching their associations. An optimal link management component should go beyond the simple search mechanism for associations. Instead, users should be able to filter and sort their associations based on various dimensions, including the types of the associated documents, the context of associations or the time when they were created.

6 CONCLUSION

We have presented the first study mainly investigating the user behaviour when associating information within as well as across digital and physical documents. Our study revealed twelve different linking mechanisms and their characteristics, which are adopted by participants when associating information in digital and physical documents. Furthermore, our study demonstrated that there is a need for an efficient and suitable information and cross-document linking solution in order to help users in creating and finding their associations. Based on the study participants' feedback about the limitations of the used linking mechanisms, their suggestions for better solutions as well as our interpretation of the collected data, we finally proposed a number of design implications for future cross-document linking solutions.

ACKNOWLEDGEMENTS

REFERENCES

[1] Annette Adler, Anuj Gujar, Beverly L. Harrison, Kenton O'Hara, and Abigail Sellen. 1998. A Diary Study of Work-Related Reading: Design Implications for Digital Reading Devices. In *Proceedings of CHI 1998*. Los Angeles, USA. https://doi.org/10.1145/274644.274679

[2] Maristella Agosti and Nicola Ferro. 2005. A System Architecture as a Support to a Flexible Annotation Service. In *Proceedings of the 6th DELOS Workshop*. Cagliari, Italy. https://doi.org/10.1007/11549819_10

[3] Deborah K. Barreau. 1995. Context as a Factor in Personal Information Management Systems. *Journal of the American Society for Information Science* 46, 5 (1995). https://doi.org/10.1002/(SICI)1097-4571(199506)46:5<327::AID-ASI4>3.0.CO;2-C

[4] Richard Boardmann, Robert Spence, and Angela M. Sasse. 2003. Too Many Hierarchies? The Daily Struggle for Control of the Workspace. In *Proceedings of HCI 2003*. Crete, Greece.

[5] Paolo Bottoni, Roberta Civica, Stefano Levialdi, Laura Orso, Emanuele Panizzi, and Rosa Trinchese. 2004. MADCOW: A Multimedia Digital Annotation System. In *Proceedings of AVI 2004*. Gallipoli, Italy. https://doi.org/10.1145/989863.989870

[6] Vannevar Bush. 1945. As We May Think. *Atlantic Monthly* 176, 1 (1945).

[7] Nicholas Chen, Francois Guimbretiere, and Abigail J. Sellen. 2013. Graduate Student Use of a Multi-Slate Reading System. In *Proceedings of CHI 2013*. Paris, France. https://doi.org/10.1145/2470654.2466237

[8] Richard Furuta. 1989. *Structured Documents*. Cambridge University Press, Chapter Concepts and Models for Structured Documents.

[9] William Jones, Ammy J. Phuwanartnurak, Rajdeep Gill, and Harry Bruce. 2005. Don't Take My Folders Away! Organizing Personal Information to Get Things Done. In *Proceeding of CHI 2005*. Portland, USA. https://doi.org/10.1145/1056808.1056952

[10] Thomas W. Malone. 1983. How People Organize Their Desks? Implications for the Design of Office Information Systems. *ACM Transactions on Office Information Systems* 1, 1 (1983). https://doi.org/10.1145/357423.357430

[11] Catherine C. Marshall. 1998. Toward an Ecology of Hypertext Annotation. In *Proceedings of Hypertext 1998*. Pittsburgh, USA. https://doi.org/10.1145/276627.276632

[12] Moira C. Norrie and Beat Signer. 2003. Web-Based Integration of Printed and Digital Information. In *Proceedings of the Di-Web CAiSE 2002 Workshop*. 200–219. https://doi.org/10.1007/3-540-36556-7_17

[13] Kenton O'Hara and Abigail Sellen. 1997. A Comparison of Reading Paper and On-Line Documents. In *Proceedings of CHI 1997*. Atlanta, USA. https://doi.org/10.1145/258549.258787

[14] Kenton O'Hara, Fiona Smith, William Newman, and Abigail Sellen. 1998. Student Readers' Use of Library Documents: Implications for Library Technologies. In *Proceedings of CHI 1998*. Los Angeles, USA. https://doi.org/10.1145/274644.274678

[15] Beat Signer. 2008. *Fundamental Concepts for Interactive Paper and Cross-Media Information Spaces*. Books on Demand GmbH.

[16] Beat Signer. 2010. What is Wrong with Digital Documents? A Conceptual Model for Structural Cross Content Composition and Resuse. In *Proceedings of ER 2010*. Vancouver, Canada. https://doi.org/10.1007/978-3-642-16373-9_28

[17] Ahmed A.O Tayeh. 2016. *A Dynamically Extensible Cross-Document Link Service*. Ph.D. Dissertation. Vrije Universiteit Brussel.

[18] Ahmed A.O Tayeh and Beat Signer. 2014. Open Cross-Document Linking and Browsing based on A Visual Plug-in Architecture. In *Proceedings of WISE 2014*. Thessaloniki, Greece. https://doi.org/10.1007/978-3-319-11746-1_17

[19] Sandra Trullemans and Beat Signer. 2014. From User Needs to Opportunities in Personal Information Management: A Case Study on Organisational Strategies in Cross-Media Information Spaces. In *Proceedings of DL 2014*. London, UK. https://doi.org/10.1145/JCDL.2014.6970154

[20] Steve Whittaker, Loren Terveen, and Bonnie A. Nardi. 2000. Let's Stop Pushing the Envelope and Start Addressing it: A Reference Task Agenda for HCI. *Human Computer Interaction* 15, 2 (2000). https://doi.org/10.1207/S15327051HCI1523_2

[21] Robert K. Yin. 2009. *Case Study Research: Desing and Methods*. Sage Publications.