ABSTRACT
In recent years several systems have been developed to integrate the management of physical and digital documents and artefacts. These systems, which often rely on technologies such as RFID, generally detect the location of a digitally tagged item within a collection, with varying degrees of location sensitivity, ranging from a room to a smaller container such as a filing cabinet or briefcase. Despite their obvious value, such systems are not capable of detecting the precise location and ordering of individual items within the managed collection of items. In this paper we present the second generation of our earlier prototype system, called SOPHYA, which utilises a wired technology to allow management and retrieval of documents and artefacts within ordered collections.

Author Keywords
Ordered document management, smart filing system, physical documents, physical artefacts, physical interfaces, tangible interfaces

ACM Classification Keywords
H.5.2 Information interfaces and presentation: User Interfaces. Physical user interfaces

General Terms
Design, Management

INTRODUCTION
For the past few decades paper and digital documents have co-existed in most modern office environments, with each type of document playing a different role, and providing a different range of benefits. For instance paper documents provide physical affordance and are often kept for historical or legal reasons, whereas digital documents are easier to store and retrieve, while also being easier to share across distance via electronic means.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. TEI 2010, January 24–27, 2010, Cambridge, Massachusetts, USA. Copyright 2010 ACM 978-1-60558-841-4/10/01...$10.00.

Despite their co-existence, however, the systems used for management and retrieval of paper and digital documents have in the past been separate, with human users being responsible for keeping the links between them when integration of some kind is required.

In recent years demonstrative systems have been developed to assist with this process of integration between physical and digital document and artefact management systems. The primary reason for development of such systems is the realisation of the fact that paper and digital documents are likely to co-exist for the foreseeable future, and therefore more automated document management systems are needed.

The development of such systems has also been made possible by recent availability of cost-effective trackable electronic technology such as RFID tags that can even now be incorporated into the paper that is used for printing documents. Other wired technology also exists that can provide more advanced functionality than just the possibility of location detection provided by RFID tags.

However, almost all these physical document and artefact management systems only provide mechanisms for tracking physical artefacts within office environments, and depending on the complexity of the deployed system, artefacts can be detected and located within areas ranging from a room to smaller areas such as a filing cabinet. We have previously developed one such system [8] which uses wired electronic components to track documents placed in physical containers (e.g. filing cabinets, bookshelves, document trays, etc.).

Although these systems bridge the gap between the management of physical and digital artefacts, they suffer from the problem of not being able to distinguish the exact relative location of an item in relation to other items within a collection, in terms of their order. As such they can’t be used effectively in an environment where the order of stored items is important. Examples of such environments include for instance libraries where it is crucial to clearly know the order of books on a physical shelf.

In this paper we introduce a system which has been developed for the purpose of digitally managing physical artefacts, such as books and documents, where identifying the order of collection items is necessary for their storage, visualisation and retrieval.
RELATED WORK

Physical document (and other artefact) management systems require some mechanism for tracking documents. Since the development of RFID technology it has become an obvious solution to this problem, and a basis for several systems developed in recent years.

Arregui et al. [1] have developed one such system that uses RFID tags attached to documents to give the documents a digital presence. RFID readers at various locations around the office scan the tags near them, allowing the system to potentially perform tasks such as locating and tracking documents. O’Neill et al. [9] extend this technology and demonstrate how it could be deployed in a patent office.

Hark et al. [5] describe a system for managing physical documents in which RFID tags are printed directly onto the document using e-ink technology. The tagged documents can then be integrated with a digital document management system.

The work mentioned above only focuses on locating documents in a general area and not on discovering accurate position or order information. Research toward more accurate location systems relevant to locating documents has been more focused on determining the position of physical artefacts for instance on a shelf or table.

Decker et al. [4] have developed a ‘Smart Shelf’ which is able to roughly determine the position of objects placed on it. An example of its use is tracking of products on supermarket shelves. Items are tagged with RFID transponders, which are read by an array of RFID antennae underneath the shelf.

Hinske [6] has extended this array of RFID antennae concept using multiple tags on artefacts to determine their position and orientation on a tabletop. To achieve greater accuracy when determining artefact positions Hinske and Langheinrich [7] mounted the RFID reader on a moving arm underneath the table.

Raskar et al. [10] developed RFIG tags, active RFID tags augmented with photo-sensing capability, used in conjunction with a handheld projector. When the projector is aimed at the tags the system can determine their relative location and then specific visualisations can be projected onto them. A proposed potential application for this system is visualising incorrectly shelved books in a library.

Although these systems allow detection of individual items precisely, none of them are readily suitable for determining the position of an individual artefact from amongst many in confined areas. For instance determining the position of a specific folder in a filing cabinet containing many such folders requires a high level of accuracy due to the size and shape of the folders. Furthermore, the extra bulk and moving parts of systems using moving antennae would make such systems unsuitable and rather costly when many antennae have to be fitted into archives containing a large number of items, and covering large spaces.

SOPHYA

We have previously developed a system called SOPHYA [8] which allows tracking, storage and retrieval of paper documents digitally. Unlike other systems discussed earlier, SOPHYA utilises wired communication rather than RFID for tracking physical documents. In a wired system power can be supplied to electronic components attached to physical documents, thus allowing for more functionality to be added to these augmented objects (e.g. folders) that would not otherwise be possible with a passive RFID-based system, or would require batteries as with an active RFID-based system.

A modular architecture was adopted for SOPHYA to facilitate easy integration with existing conventional digital document management software. Figure 1 shows an overview of the architecture of SOPHYA, and how SOPHYA can be connected to compatible digital document management software. The architecture is split into five layers: three that belong to SOPHYA (the hardware, firmware and middleware) and two that are application specific (the digital document management system server, and the clients).

The hardware component of the architecture manages the containers—so named because they are typically some form of document container (e.g. folder, binder, archival box, etc.)—rather than the individual documents themselves. Containers are placed in physical storage locations such as filing cabinets, shelves, etc. Both the container and physical storage location are augmented with electronic circuitry. Each container is given a unique ID, and optionally some user interface components such as LEDs. The physical storage locations are able to communicate with the containers to read their IDs and control their user interfaces.

Figure 1. Overview of the architecture of SOPHYA. The Sophya layers are responsible for detecting the presence of containers (e.g. folders) in physical storage locations (e.g. filing cabinets) and presenting this information to the software.
The middleware is responsible for aggregating data coming from the various physical storage locations and presenting this to the application specific document management system server software. The middleware maintains a simple database that keeps track of information specific to the containers (e.g. container IDs, location history, etc.).

Within the internal layers of SOPHYA there is no concept of digital documents. Containers and locations each have a unique ID and this is all that is known at these layers. However, the information about these unique IDs and their relationship to each other in terms of their physical storage and location can be be queried by the application specific software which would be responsible for mapping such information to digital content.

This separation of the physical management of documents (by SOPHYA's hardware, firmware and middleware) from the digital management of content associated with those documents allows for a wide range of client applications to be developed which can link with conventional digital document management systems. Furthermore, the separation also allows for the internal components of SOPHYA to be modified and updated without requiring the application specific software to be modified (and vice versa).

Although this version of SOPHYA overcomes some of the problems associated with the RFID-based systems described earlier, it still suffers from the problem of not being able to distinguish the order of containers within a physical storage location. So for instance, although an application software connected to SOPHYA can query the contents of a filing cabinet, and find out where a document folder is, it cannot find out from SOPHYA what folders are before or after a given folder. This is clearly unsatisfactory in cases such as a virtual library visualisation application where the user may wish to find out what books are contained within a particular physical library shelf, and how those books are ordered. Similarly, even though a document management application can query SOPHYA to find out in whose in-tray a particular folder is, it would not be able to see what folders are before or after that folder for processing purposes.

In the rest of this paper we describe an alternative version of SOPHYA hardware, based on a different design, which allows detection of container order within physical storage locations and thus opens up a number of new tangible interaction possibilities.

**DETERMINING ARTEFACTS’ ORDER WITH SOPHYA**

As previously mentioned, we have in the past taken a wired approach in designing SOPHYA, as opposed to using a wireless technology like RFID. Our new alternative design of SOPHYA also takes a wired approach, as this gives us the required accuracy to determine the order of artefacts managed by SOPHYA. Although we have considered how this alternative design of SOPHYA could be applied in a number of cases where determining the order of artefacts is necessary, we have so far focused our design on a system for physical document management purposes. This decision was made so that we could carry out a comparative user evaluation of this prototype against the other prototype we have previously developed which does not support ordering of documents [8].

In this physical document management environment the primary function of SOPHYA is to allow the position of folders in a collection (e.g. in a filing cabinet, on a shelf, etc.) to be determined. In such collections, folders are always arranged along a single axis, and therefore our design is only concerned with determining the position of artefacts along this single axis.

As with previous versions of SOPHYA, folders and shelves are augmented with electronic circuitry. This gives each of the folders a unique ID, and gives the shelves the ability to read the IDs of the folders. When folders are placed on a shelf, conductive pads on the bottom of the folder make contact with pads on the top of the shelf. This provides an electrical connection through which the system can query the folder’s unique ID.

We have chosen the the Dallas-Maxim 1-wire protocol [3] for communication between folders and shelves. With this protocol it is possible for a shelf to communicate with a folder using a single wire for power and data, and a common ground connection. Thus, folders require only two points of electrical contact with the cabinet, which is provided by the aforementioned conductive pads.

![Figure 2. Design of the contacts for communication between artefacts (folders) and storage locations (shelves). Folders have two conductive pads, one for common ground which contacts the conductive surface of the cabinet, and one for power/data which contacts pads in the contact array.](image)

Figure 2 shows a conceptual diagram of our filing cabinet design. We divide the shelves’ contacts up into contact array modules (discussed in the next section) to allow for installation on shelves or in filing cabinets of various sizes. As the ground connection is shared between all artefacts, the shelf itself would be conductive and form the ground plane. The
data/power pads are insulated from the ground plane by a non-conductive layer and run in a row along the length of the cabinet.

With this design each of the data pads can be polled separately to determine whether there is a folder on it, and if so what the ID of the folder is. This provides the software with an ordered list of folders on the shelf and, if the software knows the geometry of the shelf’s conductive pads, it can determine the actual position of the folders within the cabinet.

Contact Array Layout
Our design groups the shelf contacts into contact arrays, simple modules which can be installed wherever an augmented physical storage location is required (e.g. on a shelf, in a filing cabinet, etc.) The purpose of this is to simplify the retrofitting of existing storage spaces. It also makes the design more modular, and thus improves scalability, as is discussed in a later section.

![Contact Array Layout Diagram](image)

(a) Location contact array (viewed from top)  (b) Artefact contacts (viewed from underneath)

Figure 3. Layout of contact pads on location contact arrays and artefacts for our current SOPHYA prototype.

An important consideration when designing contact arrays is determining the optimal size and location of the contact pads that form the basis of this new version of the SOPHYA’s hardware. Figure 3(a) is a top down view showing the layout of the contact pads on the storage location surface (e.g. a shelf). Figure 3(b) shows the layout of contacts on the underside of an artefact.

As shown in Figure 2 and Figure 3(a), we have positioned the conductive pads used for determining the locations of the folders at the back of the shelf/filing cabinet in our prototype. This position was chosen as folders may vary in length but they will generally all be aligned at the back of the cabinet.

The offset of the location conductive pads from the back of the surface (A) prevents shorts with the back wall and allows room for circuitry. The length of the conductive pads (B) defines the distance the folders can move forward and back within the cabinet and still be detected. A small width of non-conductive surface (C) prevents the artefact contacts from shorting to ground if not seated fully on a conductive pad.

To ensure a good connection between the artefacts and location pads the width of the artefact pads (H) needs to be greater than the gap between the location pads (E). To prevent multiple artefacts from sitting on the same location pad the distance between artefact pads at the closest possible packing (i.e. the width of the artefact (I) minus the width of the artefact pad (H)) must be greater than the width of the location pads (D).

IMPLEMENTATION OF SOPHYA FOR DOCUMENT MANAGEMENT
This section describes the implementation of our new prototype document management system using the design described in the previous section.

Contacts
Our preliminary test prototype implemented the location contact arrays using thin (0.2mm) printed circuit boards (PCBs). The initial test design, as shown in Figure 4(a), incorporated a ground contact pad as well. In this design the distance between the data pads and ground pad was fairly small to allow for folders of various lengths by keeping the two contacts close together at one end of the artefact. Each of these contact array circuit boards was 96mm wide, consisting of 16 contacts, each 5mm wide, 20mm deep and 1mm apart.

![Initial Test Design Diagram](image)

(a) Location contact array  (b) Artefact contacts (top and bottom)

Figure 4. Our initial test design incorporated a ground pad into the location contact array (top pad) and had the two artefact contacts close together. We have now redesigned this, placing the contacts at each end of the artefact.

Our test prototype also implemented the artefact contacts using the same style of thin PCB. The initial design of these PCBs is shown in Figure 4(b). The bottom of the PCB had two exposed areas of copper which acted as contacts. On the reverse these connected to the appropriate pins of a DS2401 1-wire silicon serial number IC [2]. The contact board was glued to the bottom of the folder.

During our testing of this initial design we discovered that its contact placements were not optimal, resulting in unreliable communication due to lack of connection between the folders’ pads and the storage location’s contact arrays. Although pushing the folders onto the cabinet’s contacts (e.g by adding magnets to the folders) resulted in somewhat more reliable communication, the improvement was not sufficient. We also tried adding a lump of solder to each folder’s contact to give it some depth. Although this improved the connection, it also resulted in the middle contact acting as a fulcrum upon which the folder would rock.
As a result of these tests we finally decided to move the two folder contact pads to the opposite ends of the folder. The data contact pad of the folder could then connect to the storage location’s contact array, while the ground contact pad would connect to the metal shelf of the filing cabinet itself, which acted as the ground connection. By using the whole of the shelf as the ground contact rather than a single strip, it would still be possible to support folders of varying length.

Figure 5. Current folder design, with contact pads placed at either end to prevent it from rocking.

Figure 5 shows the current implementation of the folder contact pads. In this version the contacts are at each end of the folder, resulting in it sitting in more stable position on the filing cabinet shelf and providing a more reliable level of communication.

The implemented cabinet is depicted in Figure 6. We found that the base of the shelf of the filing cabinet did not provide a reliable conductive surface so we applied aluminium tape to it which solved this problem.

Figure 6. Proof of concept filing cabinet location implementation, using two location elements to provide contacts for the folders.

Control and Communication
We use the previously mentioned Dallas-Maxim 1-wire protocol for communication between the folders and shelves. In our current revision of the prototype the firmware is implemented in an FPGA\(^1\). Within each FPGA a number of 1-wire bus masters are instantiated. These continuously poll the contacts to detect the IDs of any present artefacts.

The detected IDs are passed via USB to an embedded system running the middleware. The current prototype does very little processing at the firmware level simply passing a continually updated table of devices detected at each location. The middleware takes this raw table, applies algorithms to make the detection of events more reliable, and converts it into an ordered list of artefacts that can then be queried by application specific software.

Testing
Our initial testing focused on making sure that folders can be detected when placed in the cabinet, determining how reliable their detection is, and discovering any issues that may need to be addressed in further development. Due to the nature of the system, such things could be tested without requiring a full scale prototype. Thus we have used a small setup occupying part of a filing cabinet shelf. We talk more about how this small prototype could be scaled up to a more useful size in the next section.

Another problem with our current prototype is that if the folders fall over their presence is no longer detected because of the lack of connection with the contact arrays. Folders falling over is only an issue when the filing cabinet is not full, and could be overcome using a bookend.

ADDRESSING SCALABILITY
In terms of our design an important issue to consider is the issue of scalability. Although our current prototype covers only a small area, we can demonstrate that it is possible to gracefully scale this design for the purpose of developing much larger functional systems.

The first issue to be dealt with relates to the hardware requirements of the system. Each contact in a location requires its own I/O pin on a device such as an FPGA or microcontroller. With a contact width of 5mm and spacing of 1mm, as used in our current prototype, 167 contacts would be required per metre of surface covered. One approach to minimise this number would be to use larger contacts, thus sacrificing accuracy and requiring wider artefacts. However, while finding a good balance between contact size and accuracy is part of the solution, it is not going to completely solve the problem of scalability; thus scalability needs to be addressed at the architectural level.

\(^1\)Field Programmable Gate Array—a type of reconfigurable logic device
Scalable Architecture

To address the issue of scalability and allow our design to be used in a variety of situations we have adopted a modular control architecture. Figure 7 shows the components of this architecture.

![Control architecture of SOPHYA used for providing artefact order information. This maintains the same basic structure as the architecture of our previous prototype, but with an extra layer of middleware to perform more processing in parallel.](image)

Each location is split up into a number of location elements. Each location element comprises a contact array and control firmware. The contact array is the physical area where the artefacts contact the location, and is made up of a number of contacts. Each location element includes some form of control firmware (e.g., running on a microcontroller) which is responsible for polling these pads for artefacts and presenting this to the higher layer processing middleware.

The middleware is arranged so that a number of location elements report to a processing middleware which takes the raw information provided by the firmware of all connected location elements and converts into a form that is accessible by application specific software such as digital document management systems. A central middleware interfaces this application specific software to the various processing middlewares.

To demonstrate the scalability of our design, let us assume that, for instance, each location element comprises 32 contacts and the local firmware provides updates to the processing middleware twice per second. This would require a communication speed of 4Kbit/s \(^2\) (plus communication overhead). The location elements could be connected using a multi-drop bus such as I\(^2\)C. An I\(^2\)C bus running at 100kbit/s should support at least 10 location elements, which accounts for 320 contacts or almost 2 metres of surface. A single embedded system, running processing middleware, could master a number of these I\(^2\)C buses.

\[^{2}\text{64bit (the size of the ID)} \times 32 \times 2 = 4,096\]

In terms of processing this is a totally parallelisable problem. While we may want to determine the relative order of artefacts on shelves we can consider all shelves independently. Thus the processing in large archives can be split up and performed by separate processing middleware, implemented in separate embedded systems distributed around the archive. A central middleware would then simply direct queries from the software to the relevant system.

**SEPARATION OF DISPLAY FROM ARTEFACTS**

As shown in Figure 8 our earlier prototype document management systems [11, 8] incorporated a simple means of interaction output on the folders in the form of LEDs. The output LEDs were used in these systems to display the location of a single folder or multiple folders as the result of users’ interaction with the computer application that accessed the system. These interactions could for instance include a user searching for a particular document contained in the collection, in response to which the system would turn on the LED of the folder that contained the physical document.

![Interactive elements of our earlier prototypes.](image)

(a) Vertical file prototype—interaction through an LED and button.

(b) Lateral file prototype—interaction through three LEDs

In these previous systems the inclusion of some form of output (e.g., LED, buzzer, etc.) on the electronic components attached to the artefacts was necessary, as it was impossible to show their exact location without having an output element on the artefact itself. In our latest version of SOPHYA, however, it is not necessary to have an output device on the containers (e.g., folders). The separation of the output component from the container electronics not only decreases the complexity of the electronic components and its associated production cost, but more importantly allows for more advanced output devices to be attached to the system.

**Possible display options**

The type and complexity of the output display mechanism incorporated into the design of a SOPHYA system will be dependant on its application within a particular environment.
As we have so far focused on the use of SOPHYA for document management purposes, in this section we provided demonstrative examples of different types of output technology that can be incorporated into a filing cabinet. However, these types of devices can also be incorporated into other physical settings (e.g. library shelves, etc.) and used for management of other types of artefact (e.g. books, etc.).

(a) Simple LED interface: LEDs at the front of the cabinet light to indicate the position of a requested folder. These LEDs could also be placed outside the filing cabinet to indicate the position of a file within the cabinet.

(b) Graphic LCD interface: allows for more advanced visualisation of information.

Figure 9. Examples of the visual interfaces possible with the new version of SOPHYA.

One of the simplest, and cheapest, forms of output is the use of a series of LEDs, approximately one per pad in the contact array, running along the front of the shelf, as shown in Figure 9(a). This would provide an almost exact replica of our earlier versions of SOPHYA, where individual folders could be identified as the result of users’ interaction with the client application (e.g. when folders are searched for, etc.) and their positions shown by the system lighting the LEDs corresponding to the pads upon which they sit. The advantage here is that the system could take into account the width of individual folders (by knowing how many pads it is sitting on) and light appropriate number of LEDs corresponding to the actual space the folder is occupying in the archives.

This concept could also be extended by using multi-coloured LEDs or multiple LEDs per location pads. The client application system could then display a range of information using a combination of LEDs and colours. For instance: the results of searches performed by different people could be displayed in different colours when several people are using the system simultaneously: the position of folders that have been removed and not returned to the archives could be shown; the position of misplaced items could be displayed; or when folders need to be returned back to the filing cabinet the system could show their position.

In cases where it is necessary to show more complex graphical or textual information about the content of the folders (or other artefacts), more advanced types of display output can be incorporated into the physical storage location. Figure 9(b) shows a demonstrative prototype of the latest version of SOPHYA, where an LCD is attached to a filing cabinet shelf. In this system a series of LCDs running along the front of the filing cabinet could be used to not only show simple information (e.g. the position of folders when they are searched for, what folders are missing, etc.) but also a range of more detailed information about the folders (e.g. a list of documents contained in each folder, etc.).

The SOPHYA hardware could also be easily extended to provide a combination of input mechanisms (e.g. buttons, touch sensitive pads, etc.) at the storage location along with a range of output displays, so that more advanced user interaction capabilities can be implemented. Figure 10 shows one such concept where a touch sensitive strip running along the front of the filing cabinet shelf is combined with a single LCD at the top. This particular implementation would for instance allow the user to tangibly browse through the physical archives while displaying related digital information on the screen.

Additionally, such displays do not need to be located at the archives. As the position information is available digitally, it would be possible to remotely browse the archives as they are physically arranged. For instance, we are currently in the process of developing a virtual library system that would allow users to remotely browse libraries and dynamically see the contents of physical shelves and how books are arranged on them.

EVALUATION

We have carried out a series of interviews with people in charge of document collections in nine offices of different types and sizes, ranging from a small lawyer’s office to a large newspaper’s advertising agency. These interviews have provided us with a detailed analysis of the kind of physical and digital document management systems used in these types of conventional offices. Development of SOPHYA
has been guided by the requirements that have been identified through these interviews. We are currently in the process of designing a comprehensive study of the functionality of SOPHYA, which would lead to further refinement of its functionality and its future deployment in real-work environments.

CONCLUSION

In this paper we have presented an alternative design to our previously developed SOPHYA system. The latest design allows for the management of an ordered collection of physical artefacts, where the system can detect the exact position and ordering of items within its managed collection.

We have developed a prototype version of SOPHYA based on this new design, which can be used for the management and retrieval of document folders. We have also described how such a system could be scaled up for use in real-world archives.

The new design has several advantages over the previous version. In particular the separation of the display output component from the artefacts, and the possibility of detecting the exact position of artefacts within the collections, makes it possible to develop more advanced tangible interfaces to physical collections. We are currently investigating several ideas for such interfaces, particularly for searching and browsing physical libraries.

REFERENCES


