User-Centered Design of a Visual Data Mapping Tool

Abdallah Namoun, Usman Wajid, Nikolay Mehandjiev, Ali Owrak
Manchester Business School
Manchester, M15 6PB, UK
firstname.lastname@mbs.ac.uk

ABSTRACT
Understanding the meaning and the type of a business document received by a company is important in order to determine an appropriate response. We have developed a visual tool allowing ordinary users to express mappings between arriving documents and their elements on one side and the different document types on the other. The tool is used to set up and continuously update an automatic semantic analysis mechanism which determines the document type from a set of information items contained in the document, thus allowing automatic processing associated with the types to be applied to the arriving document instances. The activities performed by end users within the visual data mapping tool are quite complex and require user-centric design to ensure tool is useful and usable. In this paper we describe the user-centric process informing the design of the tool. We conducted two workshops to discover the mental models of our users regarding the information content in their documents and their requirements towards the design of the tool. Subsequently, these findings were used to design and implement the visual data mapping tool. The resulting system was evaluated by target end users who proficiently demonstrated the usability of the developed concepts and features.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Evaluation/methodology, Graphical user interfaces (GUI), User-centered design.

General Terms
Design, Documentation, Human Factors, Languages.

Keywords
Visual data mapping tool, end user development, observational studies, user-centered design, business documents.

1. INTRODUCTION
Processing business documents such as purchase orders and invoices is an everyday activity for millions of companies of all sizes. Efficient and effective support for the receiver of such a document, to understand its meaning, determine its type, and provide a correct response can have a significant impact in terms of both time saved and improved quality of business. Such a task is usually performed manually resulting in inefficient consumption of resources. This motivated us to design a visual data mapping tool centered around the perspectives of business users using simple and yet powerful visual techniques which we elicited from a set of user studies.

Previous tools focused on enabling users to either group documents and visually explore their relationships [30] or align ontologies [31]. However, these tools do not allow users to annotate and manipulate documents nor do they consider user preferences which make them unsuitable for business document alignment. Our visual data mapping tool aims to help non-technical users to map elements of incoming documents to document types, thus contributing towards an automatic semantic analysis mechanism. The semantic analysis mechanism is designed to ensure local relevance and avoid reliance on standard specifications. For example, consider an enterprise which sends an 'order' to a partner. We do not look at the explicit document label "order" since this may not make sense in the context of our company. Instead we classify the document within our local category of "purchase request" documents, using the semantics of the information items contained in the document, such as item codes, item descriptions, customer account number, shipping address, etc. This allows us to initiate relevant business processes.

This semantic analysis mechanism [29] was developed within the context of the EC-funded project COMMIUS[1]. The mechanism is automatic and largely transparent to its users, yet it is designed to benefit from explicit user feedback correcting wrongly classified documents, or creating new document types where necessary, according to local practices and needs of the user. To support such feedback, we have developed our Visual Data Mapping Tool (VDMT). The tool is used to setup the document analysis and also steps in when a document is not recognized correctly. It aims to support users who are not programmers, providing a visual interface which allows them to identify information items supporting the correct recognition of the document, in addition to mapping a set of information items onto a document type from a company-specific document taxonomy. The operations supported by the tool are quite complex and require careful user-centric design of its interface and interaction metaphors, the process of which we describe in this paper.

Our methodology, described in Section 3, is aligned with the approach of empirical studies of programmers or psychology of programming [14], since we aim to discover the mental models of our target user group. This will include the examination of the information content in user documents and the activities of recognizing a document as belonging to a certain type and hence requiring a certain action. We perform this by conducting a "pen and paper" exercise where users were asked to identify relevant

[1]http:// www.commius.eu

information items within a set of business documents "on paper". Affinity diagramming [16] was then applied to align the information items, allowing their discovery and organization into a seed taxonomy of business documents relevant to our target sample of SMEs. These two exercises helped to shape the information model and representation underpinning the tool. To formulate additional requirements towards the interaction style and functionality supported by the tool, we also asked the users to work with a very early version of a text tagging editor.

In the second stage we used the findings to design and implement the visual data mapping tool, as described in Section 4 of this paper. The results from a summative evaluation of the tool by a set of target users, reported in Section 5, confirm the suitability of the tool and the success of our user-centered design approach.

2. RELATED WORK

2.1 End User Development and Natural Programming

End user development is an emerging field, gaining considerable attention in the research community during the last decade due to the increased proliferation of technology in our daily activities. This field is particularly concerned with providing methods and specialized tools for end users, especially those with no technical expertise, to empower them to develop software applications that can help them perform their tasks more efficiently. The focus of research has therefore shifted from producing software applications that are usable to making software applications that are modifiable and developable [17]. Following this direction, we aim to empower business users to effectively and easily categorize their documents by providing them with the necessary tools. Consequently, users will be informed with a view of identifying the appropriate process steps.

Programming tasks usually comprise designing, writing, debugging, and maintaining source code and require expertise in various areas (e.g. formal logic) which is beyond the knowledge and capacity of non-programmers. One way to make programming and software development more *natural* and *accessible* to a wider audience is to make programming activities mimic the tasks for which these ordinary users seek software support, and also use representations which are aligned with these tasks, and hopefully even familiar to the users [24]. The research thread concerned with this objective is called Natural Programming [18]. Its philosophy revolves around the development of languages and environments that are nearer to the way people work and think, and thus the empowerment of users to express their ideas in the same way they think about them. In this respect, it is concerned with the study and understanding of users' mental model and plan, working practices, and ways to solving programmable problems. The resulting knowledge is then used to formulate and shape the design of natural development languages and tools.

Many research studies in the area of End User Programming [24] have been carried out with the aim of unraveling the way non-programmers solve programming tasks and accordingly build tools that suit both their thought and work practices. For instance [21] investigated the natural expressions and structure children use to solve problems using only pen and paper in two user studies focusing on the programming of a Pac-Man game and database system. The outcome of these studies was used to develop the HANDS programming environment. Similarly [22] and [23] conducted user studies to understand how non-programmers localize errors and help eliminate them when programming spreadsheets and mashups respectively.

To be able to design authoring tools that assist business users in their everyday job, we need to understand how these users think and go about the problem of handling, annotating, and classifying business documents. Of particular importance are the questions: how should the information concepts of the mapping tool be organized and structured?, how do business users identify important information items within business documents?, and, how to they map core concepts to information items? To answer such questions we conducted empirical user studies to identify the needs of our target audience and investigate their techniques when completing their tasks.

However, most of the existing research efforts focus on the technical technologies and solutions as presented in the next section, and to the best of our knowledge we are the first to ground the development of a visual mapping tool in a notation-style investigation of how ordinary users annotate, map information items, and develop business document types.

2.2 Existing Tools for Facilitating Enterprise Interoperability

A number of tools and techniques have been created to allow the smooth exchange of information and coordination of activities between business partners, an area of activity known as Enterprise Interoperability. The most widely used approach is to define an interoperability standard such as ebXML [9] or RosettaNet, and a number of tools have been developed to help companies work with these standards. For example, *ebMail* [10] is a platform-neutral GUI system which claims to help users with minimal knowledge of ebXML to engage in trading activities. The user interface is email-client-like and aids in lowering the learning curve. Underlying the GUI, the system makes use of open ebXML standards to communicate with business partners. Since business messages are composed and read in GUI form, ebMail is useful to SMEs who do not need backend integration. For the ebXML Messaging Service, ebMail makes use of another project called *Hermes* [11].

Another tool used for semantic interoperability is *NEXUSe2e server* which provides XML based messaging using ebXML to integrate business processes that span multiple companies. Furthermore, *BEA WebLogic Integration* [12] empowers trading partners to exchange business messages using any of these business protocols: XOCP, RosettaNet, cXML, and ebXML messages. It also supports the creation and execution of workflows that model ebXML business messages. Finally, Sybase's BPI Integrator [13] (a.k.a. Web Service Integrator) is an enterprise solution addressing the complex requirements for multi-entity collaboration across a company's value chain. It seamlessly integrates business processes and enterprise data between trading partners based on leading edge standards for business-to-business collaboration such as Web Services.

The use of the centralized standard approach in COMMIUS is precluded by the envisioned application context of open marketplaces, where new participants can arrive at any time, and are able to join trading activities and exchange documents with other companies with a minimum of preparatory effort. Most suitable solutions to interoperability involve the use of complex technologies which are not easy to use by ordinary users and are thus unsuitable for SMEs [6].

Instead we leverage Core Components standard [7] to underpin interoperability in our approach by defining information units rather than complete documents. This is a standard above the representational level of XML, the Resource Description Framework (RDF) and the Web Ontology Language (OWL). The latter is used by a number of tools for ontology editing, storage and querying, including several protocols and solutions for the annotation of business documents, e.g. Annotea [1], RDF annotation [2] protocol or Magpie, which represent semi-automatic annotation based on natural language processing [3].

A variety of graphical tools have been created in order to allow users to map fields between two documents for instance: [26] developed a data transformation tool that relies on a visual form copying metaphor to meet the needs of business analysts who are not programmers, and evaluated the tool using the cognitive dimensions framework. Likewise, [25] described VisAXSM, a data mapping system which uses automated schema analysis agents and user interaction to propose potential data mapping correspondences. The authors of [28] examined three mapping visual languages, e.g. View Mapping Language, Rimu Visual Mapper, and Form-based Mapper, and evaluated their cognitive effectiveness. These candidate mappings are then reviewed and acted upon by a user as required. There is no published record of user studies associated with any of these tools.

In contrast, we investigate how our target users perceive a set of agreed core components, i.e. information "atoms", and their structures, and present how user views informed the requirements of a visual data mapping tool which steps in and handles unrecognized documents and also manages the local structure of document types.

3. USER REQUIREMENT ELICITATION

3.1 Motivation

Our initial idea was to use the information units and structures defined in a widely used standard for inter-organizational interoperability, such as UN/CEFACT [7]. We therefore selected a set of the UN/CEFACT's core components based on their suitability for the set of user stories and case studies on COMMIUS. These core components or information "atoms" were to serve as the main recognition concepts for incoming documents. Because of their central role, users also needed to comprehend them when a document was misclassified by the automatic system and manually map it to the set of local document types using our VDMT, or when users needed to manage the local document types (again through VDMT).

We thus set up a pilot study aiming to validate the degree of comprehension of the core components subset by 12 business users working for different European SMEs. However, during the study and from the initial results it became clear that the core components standard uses an information-based structure making it incomprehensible to our end users. So we re-designed the study to elicit which units of information were sought by our end users when they recognize a document, and then translate these units into a user-centric classification of information items. We refer to these units of information as (*user-defined) core concepts* throughout this paper. The resulting classification was later compared and mapped to a subset of the standard core components. During this process it became evident that some user-specified information items did not match with standard core components. These items were treated as new core components, and were included in the core components repository used by the VDMT. This procedure will be followed incase new core components are identified by the users.

Following the initial pilot study, we carried out two further user studies in order to answer three interesting questions:

Q1. What is an effective and natural way for grouping and presenting the user-defined concepts in the VDMT to business users? Alternatively to imposing some pre-defined core components and structure on business users, we wanted our users to define and agree on their core concepts as well as its underlying structure using the affinity diagramming method [16].

Q2. How do business users annotate business documents? We wanted to explore the techniques business users employ to identify important information items in a business document and the way they map relevant core concepts to these items using merely a pen and paper.

Q3. Which problems do business users encounter when annotating business documents and what are the corresponding improvements that can be made to the visual mapping tool? We aimed to extract users' problems from the annotation task and propose design suggestions to overcome these problems in the business authoring tool.

In the remainder of this section we report on the user study and its results before proceeding to describe how these results informed the design of the VDMT.

3.2 Research Methodology

We adopted the iterative user-centered approach which involves working with users to guide the development of the VDMT [16]. The design of our tool progressed through three differing stages: ***a pilot study*** to gauge users' understanding and acceptance of the standard core components [7] (reported above); ***a second user study comprising of two workshops*** to create a new set of information items (i.e. user-defined core concepts), and the organization of these core concepts to unravel the process by which documents are annotated and information is mapped; and finally ***a usability study*** to assess users' overall performance and satisfaction with the data mapping tool.

Our formative user studies comprised a series of three workshops: one pilot (as explained in section 3.1) and two "actual" workshops with representative end users. The aim of the study changed as a result of our findings from the pilot workshop, and moved from validating our choices of information items to discovering what information items would be identified by our target user group. We targeted users who have a very good knowledge of business documents such as quotes and sale invoices but do not have strong computing skills. We asked the participants in the study to tag business documents, aiming to gain understanding of the natural ways by which people annotate documents and use these as key requirements to developing an effective visual mapping tool. The two workshops after the pilot included 12 users in total. These users were office work specialists working for different business enterprises, selected for their primary role of processing business documents. Each focus group session was started by explaining the objectives and procedure of the study. We then invited participants to complete the subsequent tasks:

Task 1: specify information components on *two* differing paper-based business documents that were related to COMMIUS using a

pen, and copy the collected information items on sticky notes for the final affinity diagramming session (described in Task 4). We systematically chose two business documents which differed in structure, content and information items (Figures 1 and 2) in order to generate richer results and cover a wider spectrum of business documents.

Task 2: specify information components on another part of *two* business documents - again different to the documents used in task one - using our prototype visual mapping tool. Users were briefly introduced and trained to use the prototype tool which contained mainly two panes: a pane for listing core concepts and a pane for viewing business documents. The aim of this task was to inspect the usability of the prototype tool and provide early design recommendations.

Task 3: complete a short semi-structured questionnaire following task one and two about their experience, opinions, and the problems they faced when annotating business documents using pen and paper and using our prototype mapping tool.

Task 4: cluster different information items gathered on Post-It notes from the annotation tasks (1 & 2) into groups of related information items and create relationships between them; we call this stage the affinity diagramming session [16]. The aim is to create a user-oriented "taxonomy of core components" that will form the foundation of the VDMT in the future. Affinity diagramming is a participatory technique for sorting and grouping large amount of information into structured and meaningful groups. It is usually carried out in workshops where all participants discuss and agree on how the data should be organized.

It is worth noting that the participants performed task 1, 2 and 3 independently. We observed our participants whilst performing the above tasks and encouraged them to verbalize their thoughts following the think aloud protocol [19] while annotating the documents either on paper or using our tool in order to get insight into their mental models and the way in which they specified and categorized important information. Solutions produced were recorded for follow-up analysis.

3.3 Results

A human factors' expert reviewed the annotation and data mapping solutions provided by the focus group participants and analyzed the recorded material with particular focus on the techniques used, enabling the important information and problems encountered during the process to be identified.

First, the expert studied the sketches on paper to specify the visual notations and techniques used to highlight and categorise information items. Secondly, the expert reviewed the relevant videos to identify common cognitive problems that arose during the annotation process. Figures 1 and 2 show samples of the sketches produced by two of our participants illustrating relevant pieces of information in two differing business documents.

Users specified information items on documents using various techniques such as: circling, underlining, numbering, pointing via arrows, and adding text descriptions (i.e. semantic annotations). It is worthwhile to note that numbers were used to refer to various information groups/components. In general the main categories were numbered X whilst their subcategories were numbered X.Y, where X and Y represent two integers.

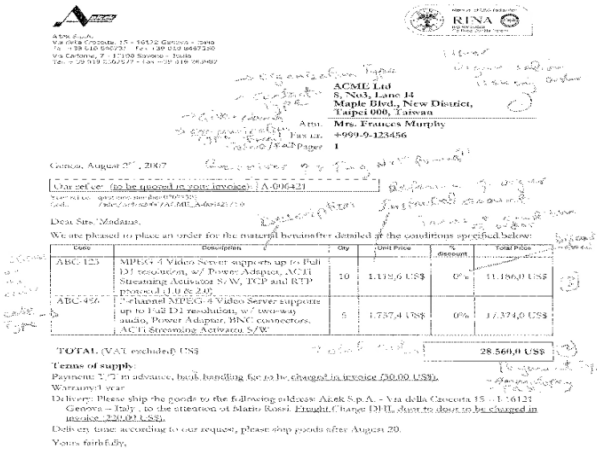

Figure 1. User annotations on pieces of information in a business document A

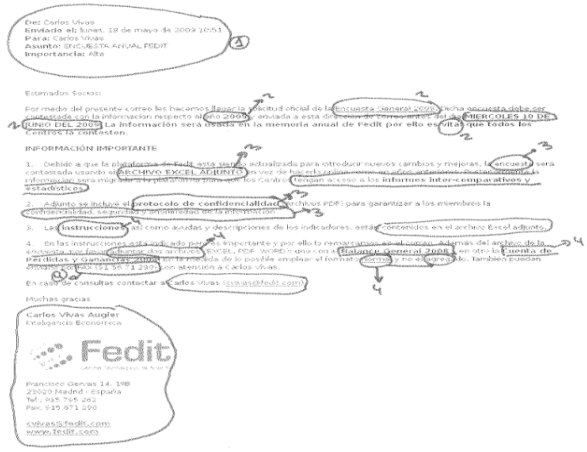

Figure 2. User annotations on pieces of information in a business document B

3.3.1 Annotation Problems

The following issues identified by the human factors expert during the annotation of paper-based documents are listed below:

Information decomposition/fragmentation: users were unclear how to annotate large elements such as tables which consist of multiple rows and columns. E.g. does a table represent one unique element or does every row, column, or cell represent a separate element? The categorization here depends on users' knowledge and interpretation of business documents, thus users' differences govern how information is annotated and how many categories are created.

Annotating relevant information: users were sometimes unable to make sound decisions about which information was relevant and useful to annotate and which information was irrelevant to annotate. This is particularly apparent in the case of documents that are lengthy and focus on various ideas. Once the relevant text was identified, deciding which information item should be included also proved to be problematic.

Concepts creation: creating meaningful and general concepts was also difficult in certain occasions. In addition, users were unsure about the prerequisites to creating a concept (i.e. what constitutes

a concept, the minimum requirements for a particular piece of data to be annotated with a concept).

Content confusion: some users expressed their frustration over parts of the documents content. However, this weak understanding could be justified by their poor knowledge of some very specialized business terms or the context of these documents.

3.3.2 Usability Problems

In the next step, the expert inspected the videos of user interaction with a prototype visual mapping tool and extracted the usability issues users encountered while annotating the document using the tool. The list of extracted issues includes:

Users were able to highlight information items that belong to different concepts using the same color. After the annotation of several information items, users were confused in regard to which highlighted text belongs to which core concept.

The current list of concepts is not detailed enough. On certain occasions, users felt that certain concepts were still general and could be further decomposed into fine-grained concepts.

Some information items were ambiguous to the users, which made selection and categorization of items a difficult task.

Some users were unable to recognize the difference between information items, especially if their names were similar, and this aggravated the concept selection problem.

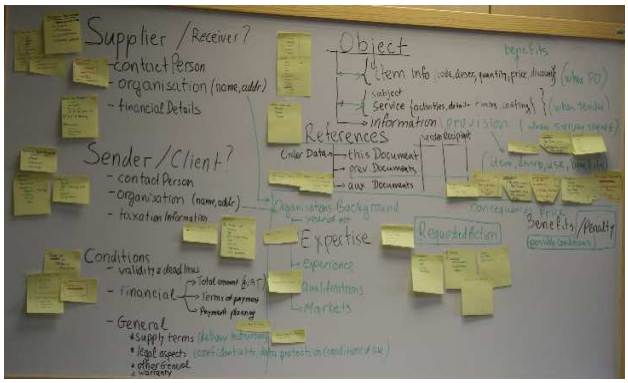

Figure 3. User-centric classification of user-defined core concepts resulting from the affinity diagramming exercise

Complex and nested sentences are harder to associate concepts to than nouns and simple sentences.

In some cases, users could not find the desired concept. Some users stated that they are uncomfortable with the current categorization offered by the system. Here it should be stressed that the tool was seeded by a tag taxonomy which has since been revised to take into account the information taxonomy created by our end users, as shown in Figure 3.

On some occasions, the system failed to annotate particular elements which frustrated the users, i.e. the system did not provide feedback to the users that their annotation process was successful.

3.4 Immediate Design Recommendations for the Visual Data Mapping Tool

Based on the problems detected during the initial set of user studies, we propose various measures to reduce the gravity of these problems and improve user annotation of business documents. These recommendations not only form the foundations and early requirements for our visual mapping tool but also serve as general design advice for similar user-focused tools used for processing business documents. The recommendations concluded from the user studies are as follows:

Self-explanatory concepts: the core concepts shown in the visual mapping tool should be self-explanatory. Moreover, each concept should contain a fairly comprehensive description which summarizes its characteristics and meaning; these descriptions could be included within a glossary that should be easily accessed when necessary.

Efficient categorization of concepts: to create an efficient list of concepts we propose to organize semantically-related concepts alphabetically, thus improving user browsing experience. Since it is not possible for a list of concepts to contain all concepts that might be required by users due to the diversity of business documents, we propose to enable users to add their own concepts or personalize existing concepts as needed. Some of the general concepts could be further decomposed into more specialized concepts to help users select the most appropriate concept.

System feedback about the annotation process: to help users during the annotation process, the system should provide continuous feedback about users' annotation activities. Successful activities as well as errors and failures should be clearly communicated to end users via error messages and failures should be clearly communicated to end users via error messages and explanations as to what went wrong.

4. DESIGN OF THE VISUAL DATA MAPPING TOOL

Based on the end-user studies we formulated the requirements for the design of the VDMT that will assist the user in mapping document instances to the taxonomy of information items or users specified core concepts. The most critical design principles were derived from what emerged as the decomposition or fragmentation problem. More specifically the tool should annotate graphically any piece of presented/rendered information. In this respect, based on the recommendations from the user studies the tool was refined to enable users to perform annotation in a two-step procedure which includes the sequential steps of highlighting the text and selecting the correlated elements from the core concepts structure. Attention was also paid to the highlighting process, so that guiding users would enable them to make the most appropriate annotation, taking into account the level of granularity of the annotation area.

While structuring the user-defined core concepts representation, the design guidelines specified that the presentation of these concepts should be hierarchical but with restricted levels (maximum three levels) to avoid time-consuming browsing. Another critical design principle was to allow users to annotate the document by selecting only the basic elements of the concepts and not the aggregate concepts. Finally, the functionality was implemented to enable users to create their own document types based on core concepts that are not mapped in the document.

Based on the user feedback and the interpreted requirements the VDMT conceptual architecture was formulated. Centered on this architecture, the conceptual level of VDMT tool utilizes four primary sources, which are implemented in four different panes

visualizing different kind of information (as depicted in Figure 4): (1) the core concepts' structure for showing the business concepts – left pane, (2) the documents instances that the user desires to annotate – centre/main pane, (3) a correlation between the existing document types and their comprising core concepts – right pane, and (4) a specification of new document types by composing existing document instances and a selection of core concepts representing necessary pieces of information in a user friendly manner (i.e. via drag and drop) – bottom pane.

The core concepts' structure is rendered in a tree-view style, as shown in the left pane in Figure 4. Only child nodes can be used for annotation since they represent basic elements / core concepts. Users can load and open a (newly arrived or previously stored) XML file that contains a document, e.g. email, in the centre pane of the tool. After loading a document the user can select and highlight a document area (containing specific information) and select/deselect among the hierarchical tree-concepts which are representative of the highlighted area. Users can also edit/remove existing annotations from the Document Instance. Following the annotation process of a business document, the user can store his newly classified document which will show up in the right pane where they can also browse existing Document Instances. These Document Instances can be opened in the central pane and customized as required.

Finally, Document Types can be easily created with the help of the newly-introduced bottom pane. Document Type creation uses a Document Instance as the "starting template" which contains assigned core concepts. These are the components already included in the Document Type. If the user is happy with those concepts, then he can proceed and save the Document Instance as a "New Document Type" which will appear in the right pane. However, if the user prefers to use additional core concepts for the Document Type, then they can choose concepts from the left pane and drop them in the bottom pane using a simple Drag & Drop feature. By clicking on "Save New Document Type" menu option the Document Type will be saved locally.

The VDMT consults an external classification module that maintains the correlations between DocumentTypes (e.g. Purchase Order, Quotation) with core concepts (e.g. TradelineItem, Address). Whenever a new annotation takes place the VDMT tool consults the classification module that internally classifies the document instance using techniques that employ Formal Concept Analysis [20]. An overview of the classification process is provided in [7]. The results of the classification process are instantly shown in the right pane of the tool.

When the user finishes the annotation process, two possible pathways are available. The first is followed when the user is not satisfied by the inferred classification of the document instance. In this case, the user himself classifies the document (e.g. as purchase request, invoice etc) and propagates this classification to the external DocumentType/core concepts correlation file. The second pathway is followed when the user is satisfied by the automatic classification. In this case the end user is allowed to enrich the document by adding annotations in the document that do not correspond to a highlighted text.

The overall tool is realized using Web 2.0 technologies in order to leverage the application's responsiveness and the end-user's interactivity. Among many Web 2.0 frameworks (e.g. jQuery) we selected YAHOO YUI® for the prototype development. Further functionality and features which have been implemented in this tool in addition to the above ones are as follows:

Document annotations: We have developed an ad-hoc *system-driven* annotation modality with which the user can highlight a text and on the fly a pop-up window prompts the recommended core concepts.

Synchronisation: More dynamism and immediate communication was included between all four panes. For instance, when clicking on a Document Instance and showing it in the centre pane, the Document Type similarities are loaded accordingly as well as the core concepts already contained in the Document Instance.

Integration: The VDMT can take the services provided by other tools (developed within COMMIUS) in order to perform the classification tasks and Document Type manipulation.

Compliance: We have made the VDMT compliant with the OSGi Framework [27] used by other tools in COMMIUS.

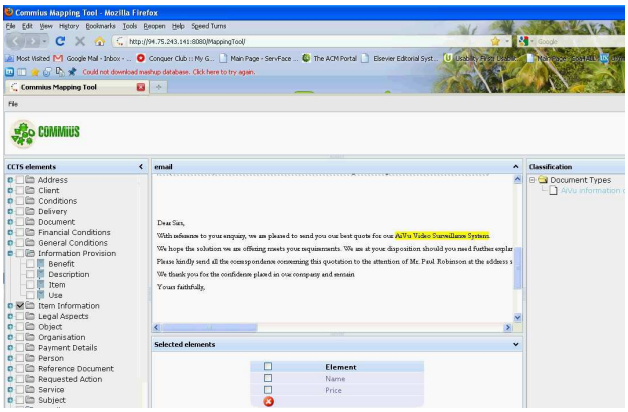

Figure 4. Screenshot of Visual Data Mapping Tool

5. USABILITY STUDY

A total of 8 business students (5 males and 3 females, mean age 25) took part in a 1 hour summative usability testing of the Visual Data Mapping Tool. The background questionnaire indicated that our participants come from a business-related background with an acceptable business experience of 3.37/5 (where 1= not expert, 5= expert). In this study, participants (1) were trained to use the tool and had the features of the tools explained to them for a period of 10 minutes, (2) completed 4 typical visual data mapping tasks, and (3) rated the usability aspects (e.g. ease of use, ease of learning) and reported their opinion about the tool in a post-test interview. The visual mapping tasks included: exploring the available core concepts and their categories, annotating three information items in a business document with three pre-defined concepts, associating three core concepts to the relevant information in a business document, and finally enriching the document by adding two core concepts that they considered relevant to the perceived type of document. The business document chosen for this study was a realistic and detailed quotation for video surveillance equipments.

The performance data showed that all participants were able to successfully complete the visual data mapping tasks (i.e. responses' correctness was 100%). In regard to task completion time, participants spent the longest time making connections (mean= 457.0 sec) between core concepts and information items and exploring available core concepts (mean= 382.66 sec).

Qualitatively, participants appreciated the simplistic structure of the tool and valued the ease by which documents can be annotated. Participants also commented on the positive impact the tool has on their efficiency and productivity at their place of work. However, they expressed concerns in regard to the length of the list of concepts (core components) and the learning curve associated with it.

Average usability scores of the Visual Data Mapping Tool were favorable. All questions were rated on a 7-point Likert scale where 1 signifies strongly disagree, 7 signifies strongly agree, and 4 signifies neutral. Users perceived the Visual Mapping Tool as easy to use (mean= 5.75, std= 1.67) and easy to learn (mean =6.25, std= 1.16) agreeing with the positive themes emerging from the qualitative data from the post-study interview. Overall, users were satisfied with the Visual Mapping Tool (mean= 5.50, std= 1.41). However, users indicated uncertainty as to whether the tool provides sufficient help information (mean= 4.12, std= 1.80) and whether it recovers from undesired mistakes easily and quickly (mean= 3.87, std= 1.45). Indeed users complained that sometimes it was quite difficult to understand the meaning of some technical concepts as they are not experts and the system did not provide sufficient descriptions explaining them. As such, we iterate the need to supply definitions (e.g. tooltips) for each core component along with realistic examples.

Table 1. Average usability rating of the Visual Data Mapping Tool (std: standard deviation)

Usability question	Mean	Std
It was simple to use this mapping tool	5.75	1.67
I can effectively complete my work using this mapping tool	5.5	1.19
I can complete my work quickly using this mapping tool	5.25	1.38
I can efficiently complete my work using this mapping tool	5.50	1.07
I feel comfortable using this mapping tool	5.87	1.12
It was easy to learn to use this mapping tool	6.25	1.16
I believe I became productive quickly using this mapping tool	5.75	1.58
The mapping tool gives error messages that clearly tell me how to fix problems	4.25	1.83
Whenever I make a mistake using this mapping tool, I recover easily and quickly	3.87	1.45
The information (such as online help, on-page messages, and other documentation) provided with this mapping tool is clear	4.12	1.80
It is easy to find the information I need	4.37	1.40
The information provided by the mapping tool is easy to understand	5.12	1.24
The information is effective in helping me complete the tasks and scenarios	5.12	1.64
The organisation of concepts on the mapping tool is clear	4.75	1.28
I can find the relevant concepts easily	4.37	0.91
The mapping tool contains the necessary concepts I need	5.62	1.30
I like using the interface of this mapping tool	5.75	1.58
This mapping tool has all the functions and capabilities I expect to have	5.25	1.48
Overall, I am satisfied with this visual data mapping tool	***5.50***	***1.41***

6. CONCLUSION AND FUTURE WORK

This paper presents the case for the design and development of a Visual Data Mapping Tool following the iterative user-centered design methodology. The tool facilitates semantic interoperability between collaborating enterprises and aims to enable business users (or SMEs) to map the structure of their existing documents to elements of core ontology of business concepts. Another important feature of the tool is to provide support for the classification of existing documents and their re-classification based on the domain concepts and relevant information items.

The current functionally offered by the tool has been directly affected by the design recommendations that were motivated by the findings of the user studies. Essentially, we conducted three rounds of user studies to inform and refine the design. A pilot study assessed user acceptance of existing core components and their categorization. A second study, taking the form of individual observational study and affinity diagramming followed. This enabled understanding as to how business users naturally map data elements and come up with easy to understand and user-oriented core components. The resulting knowledge was used to design a Visual Data Mapping Tool closer to people's inner thinking and to the ways they classify business documents. Finally, a usability study was performed to evaluate the look and feel of the Visual Mapping Tool and satisfaction of our end users.

The analysis of the end user studies show that the tool has the potential to simplify the process of mapping information items contained in different business documents, and to make this process accessible to end users who are not IT specialists or programmers, thus fitting to the profile of users in SMEs.

The importance of making VDMT accessible to users without IT expertise has been identified as one of the main conclusions from our user studies. In this respect, we have taken first important steps towards reversing the direction of information identification. That is, instead of presenting users with standards-informed information items, we elicited their perception of natural information items and relationships between them, and designed a tool to facilitate the mapping between the natural representations and the standards-based core components which is then used for information processing.

Herein we presented a novel user view on information items included in business documents and their relationships in the information taxonomy. We then described how the user view provides valuable input to the design of a VDMT, which will facilitate and simplify the user interactions with the (business documents-based) information taxonomy. In light of the conducted user studies, we define future directions and propose tentative design guidelines for developers of similar business visual data mapping tools as enumerated below:

Visual techniques to enhance visual mapping: the usability findings showed that users had problems linking the annotated information items to their corresponding concepts, and as suggested by end users, pieces of text which belong to differing core concepts should be highlighted using different colors. Assigning different colors to differing concepts will enhance users' visual mapping and thus enable them to understand and develop business document types quickly and efficiently.

Concept description and search engine to facilitate search tasks: finding the appropriate concept for a particular information item is one of the most persistent issues in annotating documents. This

issue can be encountered by providing definitions supplemented with actual examples of use, and a search engine for finding suitable concepts. The descriptions need to be tuned to the user language and avoid technical jargons.

Automatic suggestion for finding relevant concepts: browsing long lists to find concepts of interest could be a long and cumbersome process. Therefore, mapping tools should implement search fields that use automatic suggestion in order to quickly match users' concept target. It is also desirable to incorporate an intelligent semantic mechanism by which, upon user selection of a piece of text in the Document Instance, the system automatically proposes potentially relevant concepts.

System Connectors: part of the work to be carried out in the future involves lifting information from legacy systems into document instances dynamically created by the VDMT.

7. REFERENCES

[1] http://www.w3.org/2001/Annotea

[2] http://ilrt.org.discovery/2001/04/annotations/

[3] Uren, V., *et al.* 2005. Semantic Annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics: Science, Services and Agents on the WWW*. 4, 1, 14-28.

[4] Kalfolglou, Y. and Schorlemmer, M. 2003. Ontology mapping: the state of the art. *The Knowledge Engineering Review*. 18, 1, 1-31.

[5] For example, see the activities of the KRAFT project. http://www.csd.abdn.ac.uk/research/kraft

[6] Piddington, C. 2005. SME interoperability in the global economy: A discussion paper. In *proceedings of the 16th IFAC World Congress*. 16, 1.

[7] UN/CEFACT Core Components Technical Specification: http://www.unece.org/cefact/ebxml/CCTS_V2-01_Final.pdf

[8] Joseph, D. and Marín, C. A. 2010. A Study on Aligning Documents Using the Circle of Interest Technique. In *5th International Conference on Software and Data Technologies, ICSOFT 2010*.

[9] ebXML Technical Architecture Specification v1.04: http://www.ebxml.org/specs/ebTA.pdf

[10] ebMail: http://www.freebxml.org/ebmail.htm

[11] Hermes: http://www.freebxml.org/msh.htm

[12] Introducing BEA Weblogic Integration 8.1. White Paper. http://whitepapers.zdnet.com/abstract.aspx?docid=60681

[13] Sybase Web Service Integration - Product Manual: http://manuals.sybase.com/onlinebooks/group-wsi/wsig0250e/wsi25ug/@Generic__BookTextView/308;hf=0;pt=587

[14] Weinberg, G. M. 1998. *The Psychology of Computer Programming* (silver anniversary ed.). Dorset House Publishing Co., Inc., New York, NY.

[15] Beyer, H. and Holtzblatt, K. 1998. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann.

[16] Courage, C. and Baxter, K. 2005. *Understanding your users. A practical guide to user requirements: Methods, tools, and techniques*. Morgan Kaufmann., SF, CA.

[17] Lieberman, H., Paternò, F., Klann, M., and Wulf, V. 2006. End-User development: An emerging paradigm. In *End User Development*, Henry Lieberman, Fabio Paternò, and Volker Wulf, Ed. 9, 1, 1–8. Springer Netherlands, Dordrecht, 2006.

[18] Myers, B.A., Pane, J.F., and Ko, A. 2004. Natural Programming Languages and Environments. In *Communications of the ACM*, special issue on End-User Development (Sept, 2004). 47, 9, 47-52.

[19] Lewis, C. 1982. *Using the Thinking Aloud Method. In Cognitive Interface Design*. Technical Report, IBM RC-9265.

[20] Priss, U. 2005. Formal Concept Analysis in Information Science. In *Annual Review of Information Sciences and Technology*. 40, 2005.

[21] Pane, J.F. 2002. *A Programming System for Children that is Designed for Usability*. Ph.D thesis. CMU-CS-02-127., Carnegie Mellon University, Computer Science Department, Pittsburgh, PA.

[22] Ruthruff, J., Burnett, M., and Rothermel, G. 2005. An Empirical Study of Fault Localization for End-User Programmers. In *International Conference on Software Engineering*, (St. Louis, MO, 2005).

[23] Cao, J., Rector, K., Park, T.H., Fleming, S.D., S.D. Burnett, S.D., and Wiedenbeck, S.A. 2010. Debugging Perspective on End-User Mashup Programming. In *IEEE Symposium on Visual Languages and Human-Centric Computing* (Madrid, Spain, 2010).

[24] Nardi, B. A. 1993. *A small matter of programming: perspectives on end user computing*. MIT Press., Cambridge, MA.

[25] Bossung, S., Stoeckle, H., Grundy, J., Amor, R., and Hosking, J. 2004. Automated Data Mapping Specification via Schema Heuristics and User Interaction. In *19th IEEE International Conference on Automated Software Engineering* (ASE'04), 208-217.

[26] Li, Y., Grundy, J., Amor, R., and Hosking, J. 2002. A Data Mapping Specification Environment using a Concrete Business Form-Based Metaphor. In *proceedings of 2002 IEEE Symposia on Human-Centric Computing Languages and Environments* (Arlington, Sep 2002).

[27] OSGI: http://www.osgi.org/About/Technology#Framework

[28] Grundy, J. C., Hosking, J. G., Amor, R. W., Mugridge, W. B., and Li, Y. 2004. Domain-specific visual languages for specifying and generating data mapping systems. *Journal of Visual Languages & Computing*. 15, 3-4, 243-263.

[29] Wajid, U. and Marin, C. A. 2009. Enhancing Enterprise Collaboration Using Protocols for Semantic Understanding. In *proceedings of the 18th IEEE Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*.

[30] Paulovich, F. and Minghim, R. 2006. Text Map Explorer: a Tool to Create and Explore Document Maps. In *proceedings of IEEE 10th Information Visualization Conference,* 245-251.

[31] Choi, N., Song, I., and Han, H. 2006. A Survey on Ontology Mapping. *In ACM SIGMOD Record* (Sept, 2006). 35, 3, 34-41.