PaperWeb: Paper-triggered Web Interactions

Sriganesh Madhvanath, Geetha Manjunath, Suryaprakash Kompalli, Serene Banerjee, Sitaram Ramachandrum, Srinivasu Godavari

HP Laboratories
HPL-2012-237

Keyword(s):
Web tasks; mobile cameras; document image processing

Abstract:
While mobile phones have penetrated deep into tier 2 and 3 cities in India and similar emerging economies, adoption of mobile web content and web services is likely to require the creation of large numbers of relevant applications and services with usable interfaces and interaction paradigms. This paper describes PaperWeb, our effort to enable mobile phone users to use the web for day to day transactions such as paying bills, buying tickets, or fixing appointments, using familiar objects such as paper artifacts. The PaperWeb system addresses two aspects: (i) creation of useful PaperWeb interactions - without programming - by moderately tech-savvy users, and (ii) the use of these interactions by tech-naive users. We describe a user study that motivated PaperWeb, and the design of the PaperWeb system and its technology components. We conclude the paper with a discussion of what has been achieved to date and future directions.
PaperWeb: Paper-triggered Web Interactions

ABSTRACT
While mobile phones have penetrated deep into tier 2 and 3 cities in India and similar emerging economies, adoption of mobile web content and web services is likely to require the creation of large numbers of relevant applications and services with usable interfaces and interaction paradigms. This paper describes PaperWeb, our effort to enable mobile phone users to use the web for day to day transactions such as paying bills, buying tickets, or fixing appointments, using familiar objects such as paper artifacts. The PaperWeb system addresses two aspects: (i) creation of useful PaperWeb interactions - without programming - by moderately tech-savvy users, and (ii) the use of these interactions by tech-naive users. We describe a user study that motivated PaperWeb, and the design of the PaperWeb system and its technology components. We conclude the paper with a discussion of what has been achieved to date and future directions.

Categories and Subject Descriptors
H5.2. Information interfaces and presentation: User Interfaces.

General Terms
Design, Human Factors.

Keywords
Web tasks, mobile cameras, document image processing

1. INTRODUCTION
The web is one of the greatest inventions of the 21st century and its end-user value just in terms of productivity – tasks such as booking tickets, looking up information - is being reaped by millions of people around the world who can use a web browser. However despite the rapid adoption of mobile phones in emerging markets (approximately 15 million subscriptions being added each month in India alone), large numbers of individuals and small businesses in these geographies remain unable to harness the power of the web, because using a browser is often too complex a task, requiring considerable understanding of web page addressing and navigation and other concepts that we take for granted.

This was validated in a recent ethnographic user study of middle-income households in the Tier 2 town of Bhopal, where we discovered that despite a number of services moving online and these families being able to afford computers and smartphones, routine tasks such as paying bills or booking tickets still require standing in lines and considerable expenditure of time and money.

In this paper we describe PaperWeb, a system that allows users to take pictures using their mobile phone of a paper artifact and perform an associated web based task. For example clicking a picture of a utility bill enables the web task of paying the bill for that particular month. We discuss the authoring of such experiences by moderately net-savvy users using simple click and browse actions, which we believe can lead to a long tail of such experiences. PaperWeb combines document image processing technologies with web task automation to enable access to web services using a familiar interface.

There have been a number of attempts over the years at relating physical objects to digital information, from Digital Desk [2] to Sixth Sense [3]. There are also specific mobile applications available today for reading barcodes from books and other products and fetching online reviews and product information. Google Goggles [4] enables generalized web search from popular images, but does not address triggering specific tasks from personally relevant documents (the long tail). We believe that our work is the first to focus on the productivity needs of non tech-savvy users in emerging economies, and to propose a general authoring-based solution for enabling a wide variety of web tasks based on paper (and other image) triggers.

2. THE PAPERWEB SYSTEM
The PaperWeb system enables online experiences to be authored once by users with the ability to use a web browser and a PC (but no programming skills), and used by non-tech savvy users using camera-enabled mobile phones, as well as PCs with a connected camera or scanner for document capture.

Authoring: Authoring of PaperWeb interactions is performed using a desktop PC based authoring tool and involves three steps. The first step is the authoring of task-specific web widgets called tasklets using a “programming-by-doing” metaphor [1]. In simple terms, it enables recording the sequence of web browsing actions required to carry out the online task (e.g. paying the utility bill) and automatically packages it into a web widget that executes in the cloud when triggered. Any user inputs required by the task (e.g. bill date, amount) and specified as “unbound” become parameters and are prompted for when the web task is subsequently executed. These web tasks are published to an online web task repository. They may be tagged as private (for the author’s use only) or public (may be used by others).

The second step of authoring is document template authoring. This involves using a single scanned sample of the paper artifact that might be a natural trigger for the web task (e.g. utility bill), and marking all fields of interest (e.g. acct number, bill date, bill amount). The document template is then published to a cloud-based document template repository.
The final step of authoring involves associating a web task (private or public) from the web task repository with a document template from the document template repository, and (optionally) specifying the mapping between the web task parameters and fields of interest in the template. It should be noted that a web task may be mapped to more than one document template (and potentially by different users), and vice versa.

Usage: As shown in Figure 1, the end user (who in general is different from the author) uses a client application called TaskGenie on his mobile device to access and execute web tasks. Using this client, the user can search the task repository for tasks of interest using keywords. Alternatively, she can capture a physical document (e.g., the familiar utility bill that arrived in the mail) using the mobile camera and retrieve tasks associated with that document. Behind the scenes, the image is transferred to the cloud, and the input document identified from all document templates registered with the repository. The user then selects one of the retrieved tasks to execute. This causes the task parameter values to be extracted from the corresponding fields on the document, obviating the need for manual entry of these values. This step in turn involves aligning the input image with the identified document template, extracting field sub-images, and interpreting them using an OCR engine. The user can correct the automatically extracted parameter values. Thereafter the web task executes in the cloud and returns the results to the user. For the end user, taking a picture of the familiar paper bill thus allows her to pay the bill with a single click.

3. COMPONENTS AND TECHNOLOGIES

The key components of the PaperWeb system are (i) the desktop authoring tool that supports creation and association of web tasks and document templates, (ii) the cloud-based web task repository, (iii) the cloud-based web task portal, (iv) the cloud-based document template repository and image processing services for document identification, field extraction and OCR (v) TaskGenie client applications for different mobile and desktop platforms.

The PaperWeb system integrates several technologies to improve the robustness of the authored interactions. For instance, tasklets have been made more robust with respect to web site changes through multi-way modeling of websites, detection of semantic errors and proactive/reactive ways of identifying and characterizing web changes. The system also integrates support for "live" (long-running) tasklets with a "semantics-based" cache for handling disruptions in connectivity.

A robust algorithm based on compressed sensing-based is used for identifying documents using a single document sample for training, and in the presence of variable illumination, skew and perspective distortion, cropping and blur. Similarly, the document field extraction algorithm is designed to extract fields of interest in a camera-captured document image in the presence of variations in lighting, skew, perspective distortions etc.

4. DISCUSSION AND NEXT STEPS

We have prototyped the PaperWeb system using a PC-based authoring tool, TaskGenie clients for multiple mobile and desktop environments including Windows, Android, Windows Mobile and WebOS, and the cloud components hosted on Amazon Web Services. We have authored multiple web tasks such as train ticket booking, mobile bill payment and loyalty points balance checking and mapped them to the corresponding ticket, bill and loyalty card templates. We have tried and tested our prototype with multiple imaging devices such as mobile cameras, desktop scanners and document cameras.

Document identification for poor quality camera-captured images has proved to be 90% accurate in initial tests using a repository of ten document templates. The accuracy of document field extraction was found to be 96% for common image distortions. OCR accuracy is greatly dependent on the availability of autofocus in the camera and a flash for low light conditions. The overall system is adequate for a small set of personal interactions, however in order to support sharing of document templates and authored PaperWeb interactions with other users, it will be important to improve the robustness of these algorithms, build in error recovery especially for poorly captured images, and scale both robustness and performance to deal with hundreds of document templates.

Algorithms for dealing with documents that do not follow a fixed template (e.g. restaurant bills which have the same fields of interest but no fixed format) need to be incorporated. The OCR engine that we are using (Tesseract) is unable to deal with handwritten content and patterned backgrounds (e.g. text on a credit card or loyalty card). Also of interest are use cases relevant to small and medium enterprises, and those involving image triggers other than documents (e.g. movie posters, street signs).

Key next steps include field testing with end users to understand usage issues once deployed in the real world, and exploring service-oriented business models for the rollout of such services.

5. ACKNOWLEDGMENTS

Omitted for blind review.

6. REFERENCES


Figure 1: Usage of TaskGenie client application on mobile phone: (i) User takes a picture of her railway ticket (ii) document is automatically identified as a railway ticket, previously authored and mapped web tasks (booking status inquiry, cancel ticket, etc) are retrieved (iii) on selecting the booking status task, the required web task parameters are automatically extracted from the document; (iv) the booking status from the railways website is displayed.