User Interfaces Supporting Casual Data-Centric Interactions on the Web

by

David F. Huynh

S.M. Computer Science and Engineering, Massachusetts Institute of Technology (2003)
B.A.Sc. Computer Engineering, University of Waterloo (2001)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2007

Author . . . .
Department of Electrical Engineering and Computer Science
August 27, 2007

Certified by . . . .
David R. Karger
Professor of Computer Science and Engineering
Thesis Supervisor

Certified by . . . .
Robert C. Miller
Professor of Computer Science and Engineering
Thesis Supervisor

Accepted by . . . .
Authur C. Smith
Chairman, Department Committee on Graduate Students

User Interfaces Supporting Casual Data-Centric Interactions on the Web

by

David F. Huynh

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2007, in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science and Engineering

Abstract

Today's Web is full of structured data, but much of it is transmitted in natural language text or binary images that are not conducive to further machine processing by the time it reaches the user's web browser. Consequently, casual users—those without programming skills—are limited to whatever features that web sites offer. Encountering a few dozens of addresses of public schools listed in a table on one web site and a few dozens of private schools on another web site, a casual user would have to painstakingly copy and paste each and every address into an online map service, copy and paste the schools' names, to get a unified view of where the schools are relative to her home. Any *more* sophisticated operations on data encountered on the Web—such as re-plotting the results of a scientific experiment found online just because the user wants to test a different theory—would be tremendously difficult.

Conversely, to publish structured data to the Web, a casual user settles for static data files or HTML pages that offer none of the features provided by commercial sites such as searching, filtering, maps, timelines, etc., or even as basic a feature as sorting. To offer a rich experience on her site, the casual user must single-handedly build a three-tier web application that normally takes a team of engineers several months.

This thesis explores user interfaces for casual users—those without programming skills—to extract and reuse data from today's Web as well as publish data into the Web in richly browsable and reusable form. By assuming that casual users most often deal with small and simple data sets, declarative syntaxes and direct manipulation techniques can be supported for tasks previously done only with programming in experts' tools.

User studies indicated that tools built with such declarative syntaxes and direct manipulation techniques could be used by casual users. Moreover, the data publishing tool built from this research has been used by actual users on the Web for many purposes, from presenting educational materials in classroom to listing products for very small businesses.

Thesis Supervisors: David R. Karger and Robert C. Miller
Titles: Professors of Computer Science and Engineering

to my parents
who are my constant source of courage

Acknowledgements

I would like to thank my thesis advisors, David R. Karger and Robert C. Miller, for guiding me when I was lost and for trusting me to explore freely on my own when I found my direction. I could not have had better supervision for every step of this arduous journey, or more gratification and pride at the end.

The SIMILE project members have also been instrumental. I thank Stefano Mazzocchi, Ryan Lee, Andrew Plotkin, Ben Hyde, Richard Rodgers, V. Alex Brennen, Eric Miller, and MacKenzie Smith for their continuing encouragement and support in various forms as well as their perspectives, insights, and wisdom that bring practicality to my research. Without SIMILE, my research would not have had the same impact and reach.

The Haystack group and User Interface Design group members have provided tremendously insightful feedback on my research, and many of them have been my travel companions throughout this long journey. Many thanks go to Vineet Sinha, Jaime Teevan, Karun Bakshi, Nick Matsakis, Harr Chen, Yuan Shen, Michael Bernstein, Adam Marcus, Sacha Zyto, Max Goldman, Greg Little, and Jones Yu.

I have also enjoyed many exchanges with users of my software published through SIMILE. I thank Johan Sundström, Josh Aresty, and Keith Alexander for their thoughts as well as their code patches, not to mention their enthusiasm in singing to their friends more praises of my work than it deserves. There can be no more satisfaction to a tool builder than to see his tools used for real.

Many people have participated in my user studies. I am grateful for their time and their insightful feedback. I also would like to thank those who have helped recruite subjects for my studies: William Reilly, Robert Wolfe, Ann Whiteside, and Rebecca Lubas.

Over the years, I have enjoyed stimulating conversations with many other people regarding my research. I thank Paolo Ciccarese, Steven Drucker, Mira Dontcheva, Eric Neumann, Daniel Tunkelang, Kingsley Idehen, Ivan Herman, Michael Bergman, Jon Crump, Danny Ayers, Vinay Mohta, Ian Jacobs, Sandro Hawke, Wing Yung, Lee Feigenbaum, Ben Szekely, Emmanuel Pietriga, Chris Bizer, Justin Boyan, Glen McDonald, Michael Bolin, Alex Faaborg, Ora Lassila, Deepali Khushraj, Ralph Swick, Daniel Weitzner, Ann Bassetti, Chao Wang, Gregory Marton, Boris Katz, and Steve Garland.

continued on next page

continued from previous page

My friends have made the journey much more fun. I thank Sayan Mitra, Vineet Sinha, Han-Pang Chiu, Harold Fox, Agnes Ng, Rui Fan, Karun Bakshi, Stefano Mazzocchi, Ryan Lee, Paolo Ciccarese, Heidi Pan, Max van Kleek, Gregory Marton, Michael Bernstein, Olya Veselova, Brooke Cowan, and Jimmy Lin. Special thanks go to 李真 for keeping my company everyday throughout.

Whereas my colleagues asked many deep questions, my family asked a very simple one, "Are you done yet?!" It was deceptively simple: while I could often provide clear responses to my colleagues, I always had to mumble some answer to my family. But now the reply is a resounding "Yes." I've come to the end of my seemingly quixotic journey, and I'm grateful to my family for being so caring and anxious to always ask, "Are you done yet?" but also for being so understanding to never question why.

Whatever you do will be insignificant, but it is very important that you do it.
— Mahatma Gandhi —

CONTENTS