CREATING INTELLIGENT CONTENT WITH LIGHTWEIGHT DITA

Carlos Evia

ATTW
ROUTLEDGE

CREATING INTELLIGENT CONTENT WITH LIGHTWEIGHT DITA

Creating Intelligent Content with Lightweight DITA documents the evolution of the Darwin Information Typing Architecture (DITA) – a widely used open standard for structuring technical content. DITA has grown in popularity and features since its origins as an internal grammar for structuring technical documentation at IBM. This book introduces Lightweight DITA (LwDITA, which should be read as "Lightweight DITA") as a proposed version of the DITA standard that reduces its dependence on complex Extensible Markup Language (XML) structures and simplifies its authoring experience. This volume aims to reconcile discrepancies and similarities in methods for authoring content in industry and academia and does so by reporting on DITA's evolution through the lens of computational thinking, which has been connected in scholarship and media to initiatives for learning to code and programming.

Evia's core argument is that if technical communicators are trained with principles of rhetorical problem solving and computational thinking, they can create structured content in lightweight workflows with XML, HTML5, and Markdown designed to reduce the learning curve associated with DITA and similar authoring methodologies. At the same time, this book has the goal of making concepts of structured authoring and intelligent content easier to learn and teach in humanities-based writing and communication programs. This book is intended for practitioners and students interested in structured authoring or the DITA standard.

Carlos Evia is an associate professor in the Department of Communication at Virginia Tech. He teaches courses in professional communication and content management strategies. Carlos is also a voting member of the DITA technical committee and co-chair of the Lightweight DITA subcommittee at the Organization for the Advancement of Structured Information Standards.

ATTW Book Series in Technical and Professional Communication
Tharon Howard, Series Editor

Citizenship and Advocacy in Technical Communication
Godwin Y. Agboka and Natalia Mateeva

Communicating Project Management
Benjamin Lauren

Lean Technical Communication: Toward Sustainable Program Innovation
Meredith A. Johnson, W. Michele Simmons, and Patricia A. Sullivan

Scientific and Medical Communications: A Guide for Effective Practice
Scott A. Mogull

Plain Language and Ethical Action: A Dialogic Approach to Technical Content in the 21st Century
Russell Willerton

Rhetoric in the Flesh: Trained Vision, Technical Expertise, and the Gross Anatomy Lab
T. Kenny Fountain

Social Media in Disaster Response: How Experience Can Build for Participation
Liza Potts

For additional information on this series please visit www.routledge.com/ATTW-Series-in-Technical-and-Professional-Communication/book-series/ATTW, and for information on other Routledge titles visit www.routledge.com.

CREATING INTELLIGENT CONTENT WITH LIGHTWEIGHT DITA

Carlos Evia

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2019
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
Names: Evia, Carlos, 1971- author.
Title: Creating intelligent content with lightweight DITA / Carlos Evia.
Description: [New York, New York] : Routledge, [2019] | Series: ATTW book series in technical and professional communication | Includes bibliographical references.
Identifiers: LCCN 2018034997| ISBN 9780815393825 (hardback) | ISBN 9780815393818 (paperback) | ISBN 9781351187510 (ebook)
Subjects: LCSH: Darwin Information Typing Architecture. | Technical writing—Data processing. | Writing—Automation. | Information organization—Computer programs.
Classification: LCC QA76.76.H94 E97 2019 | DDC 025.00285—dc23
LC record available at https://lccn.loc.gov/2018034997

ISBN: 9780815393818 (hbk)
ISBN: 9780815393825 (pbk)
ISBN: 9781351187510 (ebk)

Typeset in Minion
by Swales & Willis Ltd, Exeter, Devon, UK

To Jane and Sofia for being everything.
A mi papá (el original Dr. Carlos Evia), a mi mamá (que no es doctora pero de todo sabe), y a mis hermanos.

CONTENTS

FIGURES

PREFACE

In the summer of 2004, I was on my way to Blacksburg, Virginia, to start my new job as assistant professor of technical communication at Virginia Tech. Fresh out of graduate school at Texas Tech University, I was still in that period in which I thought it was my job to read and absorb everything and anything that was published or presented about technical communication. My new office only had a laptop computer on an otherwise empty desk when I started reading some of the just-published proceedings from the 2004 conference (now called the Summit) of the Society for Technical Communication.

The proceedings included a presentation titled "The Future of Technical Communication According to Those Who Teach It," by David Dayton – a fellow graduate (a couple of years ahead of me) of the PhD in Technical Communication and Rhetoric at Texas Tech University. Under the title of "Plate Tectonics of T-COM," Dayton's presentation included a slide that probably changed my academic life. Dayton's slide identified "two major fault lines (that) create tension, release continuous discourse" in the field of technical communication:

- Between academics and practitioners
- Among academics: literary/discourse focused versus social-science/technology focused.

As a somewhat naïve new PhD graduate, I was unaware of those fault lines. Probably it was because I came to the academic side of technical communication from industry (I was a practicing technical writer and rather clumsy interface designer when I decided to get a PhD). Or maybe it was because I was coming from a social sciences and computing systems background in another

country. In Mexico, English departments (if they exist) focus on English as a second language and technical writing does not have a strong (any?) presence in literature or letters departments.

I remember calling my father and telling him about Dayton's presentation. I told him I was unsure about my affiliation on those divisions. As a new college professor, did I need to side more with academics? As a new professor in an English department, did I need to learn more about literary approaches and discourse?

My father's advice now sounds simple, but at the time was eye-opening and somehow unexpected. Why not do both? Why not build bridges and connect the seemingly separate camps?

Fifteen years later, this book represents my contribution to patching those fault lines. *Creating Intelligent Content with Lightweight DITA* is the work of an academic who has tried to keep an active presence in industry. It also attempts to connect discourse and writing principles to technological approaches for managing content.

Getting here has been a long trip, and this book was, at different points, a very different work. Initially, it was going to be a history of the computer manual. In that version, the book claimed that the obsolescence of the manual as the main genre of technical communication was one of the causes of the professional fault lines identified in Dayton's presentation. Some of that work survived and is present in Chapter 2. At another point, the book was going to be a guide for incorporating principles of computational thinking in technical writing courses aimed at students of computer science. That work influenced Chapters 7 and 8.

Nevertheless, my experience with the Darwin Information Typing Architecture (DITA) originally as a user, then as a professor teaching DITA in my classes, and eventually as a committee member and spec author, took the book in its final direction. Both original ideas for this book (the one about the history of the manual and the one about computational thinking in technical writing courses) somehow ended up in the same place: they proposed Lightweight DITA (LwDITA) as a solution to more than one problem affecting technical communication and content development. LwDITA represents the evolution of the computer manual and it provides a way to introduce computational thinking in writing courses.

When Taylor & Francis/Routledge circulated the prospectus and sample chapters for this book to anonymous reviewers, their feedback was on point: the book should be about Lightweight DITA. Since I had been breathing LwDITA in my teaching and research work at Virginia Tech, and in my committee contributions to OASIS since 2014, that sounded like an excellent idea. At the same time, it presented a very risky challenge: I had to write a whole book about a proposed standard for structuring information that had not been approved and will not yet be approved by the time the book is published. Therefore, this is not the ultimate LwDITA user guide, and the proposed content components and

syntax details included in this book might change as the standard goes through the rounds of committee and general public approval enforced by OASIS. The history behind LwDITA, however, will not change. The need for a process or lifecycle that relies on simple markup structures and values the work of human authors to produce intelligent content will not change either. The evolution of DITA in a path that incorporates feedback from users, practices in writing instruction and communication that represent the bulk of this book will stay the same even if the LwDITA authoring formats look slightly different when the standard is finally approved.

Written from the hypocenter of the fault line between academia and industry in technical communication, I hope this book reaches its target audiences in both camps.

Reference

Dayton, D. (2004). The future of technical communication according to those who teach it. *Paper presented at the conference of the Society for Technical Communication*. May 10, 2004.

ACKNOWLEDGMENTS

This book reflects the hard work of the Lightweight DITA subcommittee at OASIS. We have been in conference calls for many years designing, evaluating, and revising the recommendations included in this book. I want to thank all active members of the subcommittee, but I need to specifically acknowledge the contributions and support of Rahel Anne Bailie, Bill Burns, Stan Doherty, Mark Giffin, Tim Grantham, Rob Hanna, Scott Hudson, and Keith Schengili-Roberts. In the early days of the subcommittee, Don Day, John Hunt, and Amber Swope made invaluable contributions to shape the proposed standard.

Kris Eberlein and Robert Anderson are also members of the subcommittee, but their guidance, leadership, and friendship as DITA spec authors deserve a special mention.

LwDITA would only be a theoretical experiment without the work of adventurous and generous developers like Jarno Elovirta, George Bina, and Radu Coravu. Bernard Aschwanden's interest in the proposed standard should take it to the next level in adoption and promotion. Sarah O'Keefe provided challenging and relevant feedback. JoAnn Hackos and Dawn Stevens made (and still make) room for LwDITA-related sessions in events of the Center for Information-Development Management that have promoted the proposed standard beyond the work of the dedicated subcommittee at OASIS.

Thank you to James Mathewson, Susan Carpenter, John Carroll, Jenifer Schlotfeldt, and all the former and current IBMers who shared with me their experiences about the past, present, and future of DITA and LwDITA.

My mentor and friend Carolyn Rude read all prototypes of this book, from the history of the manual to the LwDITA reference guide. Her feedback and vision as a pioneer explorer and bridge-builder of the technical communication fault lines has inspired me for decades. Rebekka Andersen, Tim Lockridge,

Kelly Pender (who actually knows a lot about LwDITA but probably won't admit it), and Russell Willerton helped finding sources, reading chapters, and keeping my disciplinary identity healthy. Allison Hutchison and Kelly Scarff brought their priceless graduate student eyes to the project. Allison's feedback, full of "I get it now!" moments, was a joy to read. Thank you to the dozens of undergraduate and graduate students in my courses who have been participating in my LwDITA experiments since 2014. Bernice Hausman and Joe Eska supported my DITA habit for many years.

Manuel Pérez-Quiñones, Ed Fox, Steve Sheetz, Patrick Fan, Chris Zobel, Deborah Tatar, and Steve Harrison have been at one point or another involved in my explorations for disrupting tradition and adapting principles of computational thinking to "traditionally non-computing" fields like writing and communication.

Tharon Howard believed in this project and encouraged me to write it in its current shape. "We have a good number of proposals in the series now," he said. "But nothing like yours."

Thank you to my brother Pedro Evia (the real Chef Pedro) and his business partner Eduardo Rukos for the *Sensei Sushi* materials.

Finally, if this book needed a co-author, that would be Michael Priestley – the original *ditaguy*. As the father of DITA and LwDITA, Michael has been an innovator and visionary in the world of content development. But above all things he has been a friendly collaborator, leader, and dad-joke generator. And if this book needed a substitute author, that would be my fact-checker, cheerleader, critic, and friend Alan Houser. An undisclosed percentage of this book's royalties might (or might not) go to a fund to buy a Vespa-themed Rolex for Alan.

Thank you to my wife and constant source of motivation Dr. Jane Robertson Evia. Thank you to Sofia Evia for being an awesome bebé. . . I mean, niña grande. Here's a link to Sofia's response to the question of "what is papá's book about?" https://youtu.be/MvPuw8XqOYs.

FOREWORD

Carlos Evia's book *Creating Intelligent Content with Lightweight DITA* is a most welcome addition to the ATTW Book Series in Technical and Professional Communication (TPC) because, as its title suggests, it is a book about building and maintaining content which is appropriate for users using an exciting new standard for Extensible Markup Language (XML) called "Lightweight DITA" or LwDITA. Like all the other books in the ATTW Book Series, *Creating Intelligent Content with Lightweight DITA* is solidly based on its author's comprehensive knowledge of the literature in the field and years of teaching TPC students to create content in DITA for industry clients. Balancing between his mastery of the academy and his practical industry experience as a voting member of the DITA technical committee and as co-chair of the Lightweight DITA subcommittee at the Organization for the Advancement of Structured Information Standards, Evia's book provides TPC students and practitioners with two valuable outcomes: 1) an introduction to a development model for creating digital content adapted for users, and 2) an accessible introduction to the new LwDITA standard.

Evia's work is firmly situated in what is starting to be called the shift from the "craftsman model" of Technical Communication to the "Component Content Management model." This shift has come about as a result of the ways that new technologies like XML DITA, single-source authoring environments, and Content Management Systems have changed the ways that technical writers deliver information to audiences (or what we would now call "users"). As Carlos explains in Chapter 2, the "craftsman" model made technical communicators responsible for the development and delivery of *entire* documents. Workplace practitioners wrote complete documentation sets, recommendation reports, marketing brochures, etc. And our curricula and textbooks are still based on the

assumption that our students are being prepared as craft persons who will do this same sort of work in their careers. The Component Content Management paradigm shift—which Carlos describes in Chapter 2, which Tatiana Batova and Rebekka Andersen outline in their 2017 *IEEE Transactions* article on skills needed for Component Content Management, and which JoAnn Hackos describes in her 2015 ISO/IEC/IEEE 26531 standard on content management—has helped our field understand that in the digital era technical writers in the workplace don't "write" so much as they develop, delineate, and manage small content modules which they collect into information architectures. Today's technical communicators still have to write well, but now they also need a skill set which Carlos calls "computational thinking." They have to pull on their knowledge of content strategy, information architectures, and Rhetoric in order to produce usable documentation sets in digital environments.

The Component Content Management shift has tremendous implications for TPC pedagogies and curriculum development. As Filipp Sapienza observed in the *Journal of Technical Writing and Communication* back in 2002, it suggests that "technical communicators will probably face a day when all organizational documents are saved in XML format." As educators, we should prepare our students for that experience, but as Carlos observes in this book, the problem is that learning XML and its complicated DITA standard is really hard. Even academics familiar with XHTML and CSS can find DITA intimidating. But because the new Lightweight DITA standard Carlos and his colleagues have been developing is much less dependent on both XML and DITA, it's far more accessible to TPC students and practitioners new to coding. LwDITA is exciting because it enjoys the benefits of the semantic web and single-source authoring without the incredibly steep learning curve required for XML and DITA.

This book is an extremely timely and much needed introduction to the Component Content Management and computational thinking movement in TPC. As such, it's a pleasure to have it in the ATTW Book Series in Technical and Professional Communication.

Dr. Tharon W. Howard
Editor, ATTW Book Series in Technical
and Professional Communication
July 21, 2018