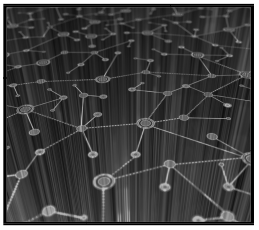

Cognitive Computing and Big Data Analytics

Judith Hurwitz
Marcia Kaufman
Adrian Bowles

WILEY

Cognitive Computing and Big Data Analytics

Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Published simultaneously in Canada

ISBN: 978-1-118-89662-4

ISBN: 978-1-118-89678-5 (ebk)

ISBN: 978-1-118-89663-1 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Library of Congress Control Number: 2014951020

Executive Editor: Carol Long
Project Editor: Tom Dinse
Technical Editors: Mike Kowolensko, James Kobielus, Al Nugent
Production Manager: Kathleen Wisor
Copy Editor: Apostrophe Editing Services
Manager of Content Development & Assembly: Mary Beth Wakefield
Marketing Director: David Mayhew
Marketing Manager: Carrie Sherrill
Professional Technology & Strategy Director: Barry Pruett
Business Manager: Amy Knies
Associate Publisher: Jim Minatel
Project Coordinator, Cover: Patrick Redmond
Proofreader: Jen Larsen, Word One
Indexer: Johnna VanHoose Dinse
Cover Designer: Wiley
Cover Image: © iStock.com/Andrey Prokhorov

We would like to dedicate this book to the power of collaboration. We would like to thank the rest of the team at Hurwitz & Associates for their guidance and support: Dan Kirsch, Vikki Kolbe, and Tricia Gilligan.

—The Authors

To my husband Warren and my two children, Sara and David. I also dedicate this book to my parents Elaine and David Shapiro.

—Judith Hurwitz

To my husband Matt and my children, Sara and Emily for their support through this writing process.

—Marcia Kaufman

To Jeanne, Andrew, Chris, and James, whose unfailing love and support allowed me to disappear long enough to write.

—Adrian Bowles

About the Technical Editors

Al Nugent is a managing partner at Palladian Partners, LLC. He is an experienced technology leader and industry veteran of more than three decades. At Palladian Partners, he leads the organization's technology assessment and strategy practices. Al has served as executive vice president, chief technology officer, senior vice president, and general manager of the Enterprise Systems Management business unit at CA Technologies. Previously, he was senior vice president and CTO at Novell, Inc., and has held CTO positions at BellSouth and Xerox. He is an independent member of the Board of Directors for Telogis and Adaptive Computing, and is an advisor to several early/mid-stage technology and healthcare startups. He is a co-author of *Big Data For Dummies* (John Wiley & Sons, 2013).

James Kobielus is a big data evangelist at IBM and a senior program director of product marketing and Big Data analytics solutions. He is an industry veteran, a popular speaker, social media participant, and a thought leader in big data, Hadoop, enterprise data warehousing, advanced analytics, business intelligence, data management, and next best action technologies.

Dr. Michael D. Kowolenko is currently an industrial fellow at the Center for Innovation Management Studies (CIMS) based at the N.C. State Poole College of Management. His research is focused on the interface of technology and business decision making. Prior to joining CIMS, he was a senior vice president at Wyeth Biotech Technical Operations and Product Supply (TO&PS), providing strategic and operations leadership perspective to ongoing integrated and cross-functional global business decisions.

About the Authors

Judith S. Hurwitz is president and CEO of Hurwitz & Associates, LLC, a research and consulting firm focused on emerging technology including Big Data, cognitive computing, cloud computing, service management, software development, and security and governance. She is a technology strategist, thought leader, and author. A pioneer in anticipating technology innovation and adoption, she has served as a trusted advisor to many industry leaders over the years. Judith has helped these companies make the transition to a new business model focused on the business value of emerging platforms. She was the founder of CycleBridge, a life science software consulting firm, and Hurwitz Group, a research and consulting firm. She has worked in various corporations including Apollo Computer and John Hancock. Judith has written extensively about all aspects of enterprise and distributed software. In 2011, she authored *Smart or Lucky? How Technology Leaders Turn Chance into Success* (Jossey Bass, 2011).

Judith is a co-author on six *For Dummies* books, including *Big Data For Dummies*, *Hybrid Cloud For Dummies*, *Cloud Computing For Dummies*, *Service Management For Dummies*, and *Service Oriented Architecture For Dummies*, 1st and 2nd Editions (all John Wiley & Sons).

Judith holds B.S. and M.S. degrees from Boston University. She serves on several advisory boards of emerging companies. She is a member of Boston University's Alumni Council. She was named a distinguished alumnus at Boston University's College of Arts & Sciences in 2005. She is also a recipient of the 2005 Massachusetts Technology Leadership Council award.

Marcia A. Kaufman is COO and principle analyst at Hurwitz & Associates, LLC, a research and consulting firm focused on emerging technology including Big Data, cognitive computing, cloud computing, service management, software development, and security and governance. She has authored major studies on advanced analytics and has written extensively on cloud infrastructure, Big Data, and security. Marcia has more than 20 years of experience in business strategy, industry research, distributed software, software quality, information management, and analytics. Marcia has worked within the financial services, manufacturing, and services industries. During her tenure at Data Resources Inc. (DRI), she developed econometric industry models and forecasts. She holds an A.B. degree from Connecticut College in mathematics and economics and an M.B.A. degree from Boston University.

Marcia is a co-author on six retail *For Dummies* books including *Big Data For Dummies*, *Hybrid Cloud For Dummies*, *Cloud Computing For Dummies*, *Service Management For Dummies*, and *Service Oriented Architecture For Dummies*, 1st and 2nd Edition (all John Wiley & Sons).

Dr. Adrian Bowles is the founder of STORM Insights, Inc., a research and advisory firm providing services for buyers, sellers, and investors in emerging technology markets. Previously, Adrian founded the Governance, Risk Management & Compliance Roundtable for the Object Management Group, the IT Compliance Institute with 101 Communications, and Atelier Research. He has held executive positions at Ovum (Datamonitor), Giga Information Group, New Science Associates, and Yourdon, Inc. Adrian's focus on cognitive computing and analytics naturally follows his graduate studies. (His first natural language simulation application was published in the proceedings of the International Symposium on Cybernetics and Software.) Adrian also held academic appointments in computer science at Drexel University and SUNY-Binghamton, and adjunct faculty positions in the business schools at NYU and Boston College. Adrian earned his B.A. degree in psychology and M.S. degree in computer science from SUNY-Binghamton, and his Ph.D. degree in computer science from Northwestern University.

Acknowledgments

Writing a book on a topic as complex as cognitive computing required a tremendous amount of research. Our team read hundreds of technical articles and books on various aspects of technology underpinning of the field. In addition, we were fortunate to reach out to many experts who generously spent time with us. We wanted to include a range of perspectives. So, we have many people to thank. We are sure that we have left out individuals who we met at conferences and provided insightful discussions on topics that influenced this book. We would also like to acknowledge the partnership and collaboration among the three of us that allowed this book to be written. We would also like to thank our editors at Wiley, including Carol Long and Tom Dinse. We appreciate the insights and assistance from our three technical editors, Al Nugent, James Kobielus, and Mike Kowolenko.

The following people gave generously of their time and insights: Dr. Manny Aparicio; Avron Barr, Aldo Ventures; Jeff Cohen, Welltok; Dr. Umesh Dayal, Hitachi Data Systems; Stephen DeAngelis, Enterra; Rich Y. Edwards, IBM; Jeff Eisen, IBM; Tim Estes, Digital Reasoning; Sara Gardner, Hitachi Data Systems; Murtaza Ghadyali, Reflexis; Stephen Gold, IBM; Manish Goyal, IBM; John Gunn, Memorial Sloan Kettering Cancer Center; Sue Feldman, Synthexis; Dr. Fern Halper, TDWI; Dr. Kris Hammond, Narrative Science; Ed Harbor, IBM; Marten den Haring, Digital Reasoning; Dr. C. Martin Harris, Cleveland Clinic; Dr. Larry Harris; Dr. Erica Hauver, Hitachi Data Systems; Jeff Hawkins, Numenta and The Redwood Center for Theoretical Neuroscience; Rob High, IBM; Holly T. Hilbrands, IBM; Dr. Paul Hofmann, Space-Time Insight; Amir Husain, Sparkcognition, Inc.; Terry Jones, WayBlazer; Vikki Kolbe, Hurwitz & Associates; Michael Karasick, IBM; Niraj Katwala, Healthline Networks, Inc.; Dr. John Kelly, IBM; Natsuko Kikutake, Hitachi Consulting Co., LTD; Daniel Kirsch, Hurwitz & Associates; Jeff

Margolis, Welltok; D.J. McCloskey, IBM; Alex Niznik, Pfizer; Vincent Padua, IBM; Tapan Patel, SAS Institute; Santiago Quesada, Repsol; Kimberly Reheiser, IBM; Michael Rhoden, IBM; Shay Sabhikhi, Cognitive Scale; Matt Sanchez, Cognitive Scale; Chandran Saravana, SAP; Manoj Saxena, Saxena Foundation; Dr. Candy Sidner, Worchester Polytechnic Institute; Dean Stephens, Healthline Networks, Inc.; Sridhar Sudarsan, IBM; David E. Sweenor, Dell; Wayne Thompson, SAS Institute; Joe Turk, Cleveland Clinic; and Dave Wilson, Hitachi Data Systems.

—Judith Hurwitz

—Marcia Kaufman

—Adrian Bowles

Contents

Introduction

With huge advancements in technology in the last 30 years, the ability to gain insights and actions from data hasn't changed much. In general, applications are still designed to perform predetermined functions or automate business processes, so their designers must plan for every usage scenario and code the logic accordingly. They don't adapt to changes in the data or learn from their experiences. Computers are faster and cheaper, but not much smarter. Of course, people are not much smarter than they were 30 years ago either. That is about to change, for humans and machines. A new generation of an information system is emerging that departs from the old model of computing as process automation to provide a collaborative platform for discovery. The first wave of these systems is already augmenting human cognition in a variety of fields. Acting as partners or collaborators for their human users, these systems may derive meaning from volumes of natural language text and generate and evaluate hypotheses in seconds based on analysis of more data than a person could absorb in a lifetime. That is the promise of cognitive computing.

Human Intelligence + Machine Intelligence

Traditional applications are good at automating well-defined processes. From inventory management to weather forecasting, when speed is the critical factor in success and the processes are known in advance, the traditional approach of defining requirements, coding the logic, and running an application is adequate. That approach fails, however, when we need to dynamically find and leverage obscure relationships between data elements, especially in areas in which the volume or complexity of the data increases rapidly. Change, uncertainty, and complexity are the enemies of traditional systems.

Cognitive computing—based on software and hardware that learns without reprogramming and automates cognitive tasks—presents an appealing new model or paradigm for application development. Instead of automating the way we already conduct business, we begin by thinking about how to augment the best of what the human brain can do with new application capabilities. We start with processes for ingesting data from inside and outside the enterprise, and add functions to identify and evaluate patterns and complex relationships in large and sometimes unstructured data sets, such as natural language text in journals, books, and social media, or images and sounds. The result is a system that can support human reasoning by evaluating data in context and presenting relevant findings along with the evidence that justifies the answers. This approach makes users more efficient—like a traditional application—but it also makes them more effective because parts of the reasoning and learning processes have been automated and assigned to a tireless, fast collaborator.

Like the fundamentals of traditional computing, the concepts behind smart machines are not new. Even before the emergence of digital computers, engineers and scientists speculated about the development of learning machines that could mimic human problem solving and communications skills. Although some of the concepts underlying the foundation technologies—including machine intelligence, computational linguistics, artificial intelligence, neural networks, and expert systems—have been used in conventional solutions for a decade or more, we have seen only the beginning. The new era of intelligent computing is driven by the confluence of a number of factors:

- The growth in the amount of data created by systems, intelligent devices, sensors, videos, and such
- The decrease in the price of computer storage and computing capabilities
- The increasing sophistication of technology that can analyze complex data as fast as it is produced
- The in-depth research from emerging companies across the globe that are investigating and challenging long-held beliefs about what the collaboration of humans and machines can achieve

Putting the Pieces Together

When you combine Big Data technology and the changing economics of computing with the need for business and industry to be smarter, you have the beginning of fundamental change. There are many names for this paradigm shift: machine learning, cognitive computing, artificial intelligence, knowledge management, and learning machines. But whatever you call it, this change is actually the integration of the best of human knowledge about the world with

the awesome power of emerging computational systems to interpret massive amounts of a variety of types of data at an unprecedented rate of speed. But it is not enough to interpret or analyze data. Emerging solutions for cognitive computing must gather huge amounts of data about a specific topic, interact with subject matter experts, and learn the context and language of that subject. This new cognitive era is in its infancy, but we are writing this book because of the significant and immediate market potential for these systems. Cognitive computing is not magic. It is a practical approach to supporting human problem solving with learning machines that will change markets and industries.

The Book's Focus

This book takes a deep look at the elements of cognitive computing and how it is used to solve problems. It also looks at the human efforts involved in evolving a system that has enough context to interpret complex data and processes in areas such as healthcare, manufacturing, transportation, retail, and financial services. These systems are designed as collaboration between machines and humans. The book examines various projects designed to help make decision making more systematic. How do expertly trained and highly experienced professionals leverage data, prior knowledge, and associations to make informed decisions? Sometimes, these decisions are the right ones because of the depth of knowledge. Other times, however, the decisions are incorrect because the knowledgeable individuals also bring their assumptions and biases into decision making. Many organizations that are implementing their first cognitive systems are looking for techniques that leverage deep experience combined with mechanization of complex Big Data analytics. Although this industry is young, there is much that can be learned from these pioneering cognitive computing engagements.

Overview of the Book and Technology

The authors of this book, Judith Hurwitz, Marcia Kaufman, and Adrian Bowles are veterans of the computer industry. All of us are opinionated and independent industry analysts and consultants who take an integrated perspective on the relationship between different technologies and how they can transform businesses and industries. We have approached the writing of this book as a true collaboration. Each of us brings different experience from developing software to evaluating emerging technologies, to conducting in-depth research on important technology innovations.

Like many emerging technologies, cognitive computing is not easy. First, cognitive computing represents a new way of creating applications to support business and research goals. Second, it is a combination of many different

technologies that have matured enough to become commercially viable. So, you may notice that most of the technologies detailed in the book have their roots in research and products that have been around for years or even decades. Some technologies or methods such as machine learning algorithms and natural language processing (NLP) have been seen in artificial intelligence applications for many decades. Other technologies such as advanced analytics have evolved and grown more sophisticated over time. Dramatic changes in deployment models such as cloud computing and distributed computing technology have provided the power and economies of scale to bring computing power to levels that were impossible only a decade ago.

This book doesn't attempt to replace the many excellent technical books on individual topics such as machine learning, NLP, advanced analytics, neural networks, Internet of Things, distributed computing and cloud computing. Actually, we think it is wise to use this book to give you an understanding of how the pieces fit together to then gain more depth by exploring each topic in detail.

How This Book Is Organized

This book covers the fundamentals and underlying technologies that are important to creating cognitive system. It also covers the business drivers for cognitive computing and some of the industries that are early adopters of cognitive computing. The final chapter in the book provides a look into the future.

- **Chapter 1: "The Foundation of Cognitive Computing."** This chapter provides perspective on the evolution to cognitive computing from artificial intelligence to machine learning.
- **Chapter 2: "Design Principles for Cognitive Systems**." This chapter provides you with an understanding of what the architecture of cognitive computing is and how the pieces fit together.
- **Chapter 3: "Natural Language Processing in Support of a Cognitive System."** This chapter explains how a cognitive system uses natural language processing techniques and how these techniques create understanding.
- **Chapter 4: "The Relationship Between Big Data and Cognitive Computing."** Big data is one of the pillars of a cognitive system. This chapter demonstrates the Big Data technologies and approaches that are fundamental to a cognitive system.
- **Chapter 5: "Representing Knowledge in Taxonomies and Ontologies."** To create a cognitive system there needs to be organizational structures for the content. This chapter examines how ontologies provide meaning to unstructured content.

- **Chapter 6: "Applying Advanced Analytics to Cognitive Computing."** To assess meaning of both structured and unstructured content requires the use of a wide range of analytical techniques and tools. This chapter provides insights into what is needed.
- **Chapter 7: "The Role of Cloud and Distributed Computing in Cognitive Computing."** Without the ability to distribute computing capability and resources, it would be difficult to scale a cognitive system. This chapter explains the connection between Big Data, cloud services, and distributed analytic services.
- **Chapter 8: "The Business Implications of Cognitive Computing."** Why would a business need to create a cognitive computing environment? This chapter explains the circumstances in which an organization or business would benefit from cognitive computing.
- **Chapter 9: "IBM's Watson as a Cognitive System."** IBM began building a cognitive system by initiating a "grand challenge." The grand challenge was designed to see if it could take on the best Jeopardy! players in the world. The success of this experiment led to IBM creating a cognitive platform called Watson.
- **Chapter 10: "The Process of Building a Cognitive Application."** What does it take for an organization to create its own cognitive system? This chapter provides an overview of what the process looks like and what organizations need to consider.
- **Chapter 11: "Building a Cognitive Healthcare Application."** Each cognitive application will be different depending on the domain. Healthcare is the first area that was selected to create cognitive solutions. This chapter looks at the types of solutions that are being created.
- **Chapter 12: "Smarter Cities: Cognitive Computing in Government."** Using cognitive computing to help streamline support services in large cities has huge potential. This chapter looks at some of the initial efforts and what technologies come into play to support metropolitan areas.
- **Chapter 13: "Emerging Cognitive Computing Areas."** Many different markets and industries can be helped through a cognitive computing approach. This chapter demonstrates which markets can benefit.
- **Chapter 14: "Future Applications for Cognitive Computing."** It is clear that we are early in the evolution of cognitive computing. The coming decade will bring many new software and hardware innovations to stretch the limits of what is possible.