Synthesis Lectures on
Information Concepts, Retrieval, and Services

SYNTHESIS
COLLECTION OF TECHNOLOGY

Bernard J. Jansen · Kholoud K. Aldous ·
Joni Salminen · Hind Almerekhi ·
Soon-gyo Jung

Understanding Audiences, Customers, and Users via Analytics

An Introduction to the Employment of Web, Social, and Other Types of Digital People Data

Springer

Synthesis Lectures on Information Concepts, Retrieval, and Services

Series Editor

Gary Marchionini, School of Information and Library Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

This series publishes short books on topics pertaining to information science and applications of technology to information discovery, production, distribution, and management. Potential topics include: data models, indexing theory and algorithms, classification, information architecture, information economics, privacy and identity, scholarly communication, bibliometrics and webometrics, personal information management, human information behavior, digital libraries, archives and preservation, cultural informatics, information retrieval evaluation, data fusion, relevance feedback, recommendation systems, question answering, natural language processing for retrieval, text summarization, multimedia retrieval, multilingual retrieval, and exploratory search.

Bernard J. Jansen · Kholoud K. Aldous ·
Joni Salminen · Hind Almerekhi ·
Soon-gyo Jung

Understanding Audiences, Customers, and Users via Analytics

An Introduction to the Employment of Web, Social, and Other Types of Digital People Data

Springer

Bernard J. Jansen
Hamad bin Khalifa University
Doha, Qatar

Kholoud K. Aldous
Hamad bin Khalifa University
Doha, Qatar

Joni Salminen
University of Vaasa
Vaasa, Finland

Hind Almerekhi
Qatar National Research Fund
Doha, Qatar

Soon-gyo Jung
Hamad bin Khalifa University
Doha, Qatar

ISSN 1947-945X ISSN 1947-9468 (electronic)
Synthesis Lectures on Information Concepts, Retrieval, and Services
ISBN 978-3-031-41932-4 ISBN 978-3-031-41933-1 (eBook)
https://doi.org/10.1007/978-3-031-41933-1

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

This book to dedicated to all those who use or want to use analytics to better their understanding of their audiences, customers, or users.

Foreword

I am pleased to introduce this book on the Web, social media, and user analytics.

As the world of digital media continues to evolve rapidly, it is more important than ever to understand the powerful tools and techniques available to measure and analyze data. This book provides a comprehensive overview of the current landscape of social media, user, and Web analytics, along with critical insights and best practices for leveraging this data for long-term success.

Social media, user, and Web analytics are essential to any digital marketing strategy. They provide the information necessary to understand customer behavior, generate meaningful insights, and make data-driven decisions. With this book, readers will learn to develop effective strategies for monitoring, measuring, and optimizing their digital presence.

I believe this book will be invaluable for anyone looking to maximize their online presence. It is an essential resource for marketers, entrepreneurs, and digital professionals alike.

Thank you for picking up this book. I hope you find it to be a valuable and informative guide.

Doha, Qatar
March 2023

Bernard J. Jansen

Preface

The use of the Web, user, and social media analytics world has grown exponentially.

Companies and individuals are using the Web and social media data to inform their business decisions and better understand their audience base. This book introduces the world of the Web, people, and social media analytics and how to use the data collected to make better decisions.

This book is written for those who are somewhat new to the Web and social media analytics but interested in learning more than the basics. While designing this book, we "drank our own Kool-Aid", leveraging analytics and personas to guide the structure, content, and flow of this book. It is intended to help those who are looking to gain the skills needed to analyze data from the Web and social media and help them make informed decisions, including practitioners, teachers, and researchers.

This book provides an overview of the concepts and techniques used in the Web and social media analytics, including data gathering methods, data preparation, quality, analytical methods, and validity. It also covers the potential Web and social media analytics applications, such as marketing, customer relationships, and product design.

This book is intended as a starting point for those interested in Web and social media analytics and with some background in these areas. We hope that it will provide readers with the knowledge and skills necessary to make the most of their data.

The authors make a note here in accordance with publishing guidelines at the time of writing. Open AI's GPT-3 large language model (LLM) was used for generating work-in-progress (WIP) text for some of the sections of chapters in this book. This WIP text was then manually reviewed and edited by an expert author familiar with the topic. This process involved adding, editing, and removing parts of the text. Each chapter was further reviewed by at least one of the other authors, usually more. Therefore, the book's

text is manually verified by expert humans, albeit benefitting, in parts, from content automatically generated by an LLM.

Doha, Qatar
March 2023

Bernard J. Jansen
Kholoud K. Aldous
Joni Salminen
Hind Almerekhi
Soon-gyo Jung

Acknowledgements

Writing a book can be a difficult and complex process. It takes a lot of research and dedication over a long period of time to put all that effort into a concise manuscript. Even though it is a challenge, it can also be an incredibly rewarding experience. We would not have been able to write this book without many individuals' help, and we are deeply grateful. Thank you!

The authors thank the executive leadership team at Qatar Computing Research Institute (QCRI), a national research institute within Hamad bin Khalifa University and a member of the Qatar Foundation. Such a great place to work! Dr. Ahmed Elmagarmid is the founding Executive Director of QCRI and is a leading computer scientist researcher, academic supporter, and business executive. Dr. Ahmed has been a consistent and enthusiastic supporter of our research. This book would not have been possible without the high-quality research environment that he provides. Good to have key leaders supporting what you do! It makes great things possible!

We also thank our colleagues at QCRI, both past and present, for the great work environment that only great people can provide! Although a team, each also has individuals to acknowledge personally.

Jim Jansen dedicates his efforts to this book to Kathy Mooney, who never got the chance to pursue an education beyond 8th grade. However, she was always happy and proud to see her son do "educational" things, like writing books. Sometimes, one can trace a life-changing influence to a place. For him, that place is the Fredonia Public Library (Fredonia, Kansas, US). It was a wealth of knowledge, joy, escape, and learning. Thank you! Also, he would like to personally thank one of the co-authors, Soon-gyo Jung, for his extra contribution in designing and developing the analytics systems that motivate the content of this book!

Joni Salminen dedicates this book to the two women in his life: his wife Parl, whose support and care make it possible for him to focus completely on research; and his mom Helena, who gave him a lot of common sense. Also, he thanks Rami, Lauri, and Mekhail for being such great friends! Also he is thankful to his dad Timo, who always encouraged his brother and him to do well in school. Thanks, Dad—miss you.

Hind Almerekhi owes an immense debt of gratitude to her mother Aisha, who has been a constant source of encouragement and support. Aisha's unwavering love and care have been a driving force that has propelled Hind toward realizing her full potential. She would also like to extend her heartfelt appreciation to her older sisters, Noor and Maha, who have always been by her side, providing attentive sisterly guidance and unwavering support. She would also like to thank her youngest sister Alanoud, whose willingness to lend a helping hand whenever needed has been invaluable. Finally, Hind is deeply grateful to her older brothers, Hamad, Jumaa, and Mohammed, for their incredible kindness and unwavering support. Their invaluable contributions to her life and work have been instrumental in shaping her into the person that she is today.

Kholoud Aldous is forever grateful for the unwavering support and encouragement of many individuals who made the creation of this book possible. Her heartfelt appreciation goes to her beloved mom Bahria, her dad Khalil, and her husband, Mohammed, for their invaluable support and encouragement throughout the writing process. In addition, she would like to express her deepest gratitude to her five brothers, Ayman, Abdelrahman, Khalid, Ahmed, and Ali, her three sisters, Eman, Amina, and Hanady, and friends, particularly Noura, for their unwavering support and encouragement during the creation of this book. Finally, she dedicates this work to her cherished daughter Farah, who patiently waited for her return every day, eager to spend time playing together, despite her busy schedule working on this book. Her love and affection drive her efforts, and she is forever indebted to her for her constant inspiration.

Soon-gyo Jung dedicates this book to his wife Shahanieline Flores Tapang who endlessly supports and trusts him, and to the future of his daughter Joo-Ae Jung. Thanks to all the people who have helped him: especially to the Team, who always encouraged and trusted him.

Finally, the world is a better place, thanks to people who want to and actually do help and support others. It has undoubtedly been our good fortune to benefit from so much help from the people mentioned here. We thank them and everyone who strives to help and support others. We will pass it on.

Contents

Part I Introduction to Analytics

1 The Foundations of Web Analytics 3
1.1 Defining Digital Analytics 3
1.2 Realizing the Role of Web Analytics 5
1.3 Key Performance Indicators 8
References 9

2 The Foundations of Social Media Analytics 11
2.1 Definition 11
2.2 The Role of Social Media Analytics 12
2.3 Social Media Usage 13
2.4 Key Performance Indicators 14
2.5 Social Media User Engagement 14
2.6 Terminology 15
References 17

3 Web and Social Media Analytics Strategy 19
3.1 Introduction 19
3.2 The Benefits of a Strategy 20
3.3 Web Strategy 21
3.3.1 What is a Web Strategy? 21
3.3.2 Key Performance Indicators 21
3.3.3 Steps in Building a Successful Web Analytics Strategy 21
3.4 Social Media Strategy 24
3.4.1 What is a Social Media Strategy? 24
3.4.2 Key Performance Indicators 25
3.4.3 Steps in Building a Successful Social Media Strategy 25
References 29

Part II Data Sources

4 Data Collection Methods 33
4.1 Introduction 33
4.2 Online Surveys 34
4.3 Crowdsourcing 36
4.4 Eye Tracking 38
4.5 Mouse Tracking 41
4.6 Search Logs 43
4.7 Triangulation 45
4.8 Social Media APIs 46
References 50

Part III Data Quality and Preparation

5 Data Quality Assessment 55
5.1 Introduction 55
5.2 Data Quality Assessment Aspects 56
5.2.1 Completeness and Consistency 56
5.2.2 Validity, Accuracy, and Timeliness 57
5.2.3 Examples of Data Quality Assessment Aspects 57
5.3 Data Quality Assessment Methods 58
5.4 Data Quality Assessment Tools 59
5.5 Data Quality Metrics 60
5.6 The Challenges of Data Quality Assessment 60
5.6.1 Examples of Challenges in Data Quality Assessment 61
5.7 Conclusion 62
References 63

6 Data Preprocessing 65
6.1 Introduction 65
6.2 Data Preprocessing in Action 67
6.2.1 Characterizing Data 67
6.2.2 Dimensionality Reduction and Data Transformation 67
6.2.3 Data Enrichment and Validation 68
6.3 Data Preprocessing Tools and Usage 68
6.3.1 Web Usage Mining (WUM) 70
6.3.2 Data Stream Mining (Stream Learning) 70
6.3.3 Knowledge Discovery from Social Media 71
6.4 Conclusion 72
References 74

Part IV Analytical Methods

7 Social Media Analytics 79
7.1 Introduction 79
7.2 Content Analysis 80
7.3 Sentiment and Emotion Analysis 81
7.4 Social Network Analysis 84
7.5 Statistical Analysis 85
7.6 Multimedia Content Analysis 85
7.7 The Benefits of Social Media Analytics Tools 86
7.8 Recommendations for Selecting Social Media Analytics Tool 86
7.9 Tracking Social Media Performance Using Social Media Analytics Tools 89
References 91

8 Search Analytics 95
8.1 Review of Search Analytics 95
8.1.1 What is a Search Log? 96
8.2 Process of Search Log Analysis 97
8.2.1 Data Collection 97
8.2.2 Fields in a Standard Search Log 97
8.2.3 Data Preparation 98
8.2.4 Cleaning the Data 98
8.2.5 Normalizing Searching Episodes 99
8.3 Data Analysis 99
8.3.1 Analysis Levels 99
8.3.2 Conducting the Data Analysis 101
8.4 Discussion 101
8.5 Conclusions 102
References 103

9 Persona Analytics 105
9.1 What is Persona Analytics? 105
9.2 Why is Persona Analytics Needed? 107
9.3 Company Use Cases for Persona Analytics 108
9.4 Future Directions for Persona Analytics 111
References 112

10 User Study Analytics 115
10.1 Defining User Studies 115
10.2 How to Create a User Study? 116
10.2.1 Acquiring Resources for a User Study 116
10.2.2 Designing User Studies 118

10.2.3 Selecting Participants for a User Study 119
10.2.4 Who Should Run the User Study? (And How?) 119
10.3 What Are the Risks of User Studies? 121
10.4 Analyzing User Study Results 122
10.5 Implementing User Study Results in Practice 124
References 126

11 Website Analytics 127
11.1 Introduction 127
11.2 Visitor Type 129
11.3 Visit Length 129
11.4 Demographic and System Statistics 130
11.5 Internal Search 130
11.6 Visitor Path 131
11.7 Top Pages 131
11.8 Referrers and Keyword Analysis 131
11.9 Errors 132
11.10 Choosing a Web Analytics Tool 132
References 135

Part V Validation Methods

12 A Discussion of the Validity of Data Analytics 139
12.1 Introduction 139
12.2 Why Not Count Instead of Measure? 142
References 145

Glossary 147