Advancing Digital Humanities
This page intentionally left blank
Advancing Digital Humanities
Research, Methods, Theories

Edited by

Paul Longley Arthur
*University of Western Sydney, Australia*

Katherine Bode
*Australian National University, Australia*
# Contents

*List of Figures and Tables*  
vii

*Notes on Contributors*  
x

1 Collecting Ourselves  
*Katherine Bode and Paul Longley Arthur*

## Part I  Transforming Disciplines

2 Exercises in Battology: Digitizing Samuel Beckett’s *Watt*  
*Mark Byron*  
15

3 Stylometry of Dickens’s Language: An Experiment with Random Forests  
*Tomoji Tabata*  
28

4 Patterns and Trends in Harlequin Category Romances  
*Jack Elliott*  
54

5 The Printers’ Web: New Tools to Crack Old Chestnuts  
*Sydney J. Shep*  
68

6 Biographical Dictionaries in the Digital Era  
*Paul Longley Arthur*  
83

## Part II  Media Methods

7 Digital Methods in New Cinema History  
*Richard Maltby, Dylan Walker, and Mike Walsh*  
95

8 A ‘Big Data’ Approach to Mapping the Australian Twittersphere  
*Axel Bruns, Jean Burgess, and Tim Highfield*  
113

9 iResearch: What Do Smartphones Tell Us about the Digital Human?  
*Mark Coté*  
130

10 Screenshots as Virtual Photography: Cybernetics, Remediation, and Affect  
*Christopher Moore*  
141

## Part III  Critical Curation

11 Rethinking Collections  
*Julia Flanders*  
163
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Methods and Canons: An Interdisciplinary Excursion</td>
<td>Katherine Bode and Tara Murphy</td>
<td>175</td>
</tr>
<tr>
<td>13</td>
<td>Reading the Text, Walking the Terrain, Following the Map:</td>
<td>Øyvind Eide</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td>Do We See the Same Landscape?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Doing the Sheep Good: Facilitating Engagement in Digital Humanities</td>
<td>Deb Verhoeven</td>
<td>206</td>
</tr>
<tr>
<td></td>
<td>and Creative Arts Research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Materialities of Software: Logistics, Labour, Infrastructure</td>
<td>Ned Rossiter</td>
<td>221</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part IV</td>
<td>Research Futures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Digital Humanities: Is Bigger, Better?</td>
<td>Peter Robinson</td>
<td>243</td>
</tr>
<tr>
<td>17</td>
<td>Margins, Mainstreams and the Mission of Digital Humanities</td>
<td>Paul Turnbull</td>
<td>258</td>
</tr>
<tr>
<td>18</td>
<td>The Big Bang of Online Reading</td>
<td>Alan Liu</td>
<td>274</td>
</tr>
<tr>
<td>19</td>
<td>Getting There from Here: Remembering the Future of Digital Humanities</td>
<td>Willard McCarty</td>
<td>291</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Index</td>
<td></td>
<td>322</td>
</tr>
</tbody>
</table>
Figures and Tables

Figures

2.1 Juxta window of Watt MS 6:98–100 and Watt (1953/1959, 238), with collation histogram and comment apparatus 22

3.1 A multidimensional scaling plot based on the proximity matrix generated by RF: Dickens versus Collins 31

3.2 A multidimensional scaling plot based on the proximity matrix generated by RF: Dickens versus reference corpus 33

3.3 Result of a COHA search: with his/her eyes fixed on/upon 37

3.4 Result of a COHA search: with his/her eyes on/upon 37

3.5 Result of a COHA search: with his hands in his 39

3.6 Result of a COHA search: his hands in his pockets 40

4.1 Author rank plotted against the number of novels by that author 56

4.2 Bubble plot of writers by country, matching nationality to output 56

4.3 Publications per country by year 57

4.4 Cluster dendrogram of Harlequin Presents novels by year 59

4.5 Principal Components Analysis of Harlequin Presents, with years overlaid 60

4.6 Early period words in Harlequin Presents 61

4.7 Middle period words in Harlequin Presents 62

4.8 Late period ‘non-status’ words in Harlequin Presents 62

4.9 Late period ‘status’ words in Harlequin Presents 63

4.10 Number of authors publishing their first title and last title in the Harlequin Presents line 65

7.1 Film Weekly 1948–49 cover and page 1 100

7.2 Film Weekly data extracted to spreadsheet 101

7.3 Drive-ins and touring circuits in Victoria, 1955–56 103

7.4 Capri cinema, Goodwood 105

8.1 Thematic clusters in the Australian Twittersphere 120

8.2 Subdivisions within the politics cluster of the overall network map 121

8.3 Participation by Australian users in the hashtags #auspol (left) and #ausvotes (2010, right) 123

8.4 Participation by Australian users in the hashtags #masterchef (left) and #royalwedding (right) 124

8.5 Sharing of links to ABC News (left) and news.com.au (left) by Australian users 125

10.1 Network overview of the Flickr tag search ‘screenshot’ 142

10.2 Network overview of the tag ‘secondlife’ on Flickr 148
List of Figures and Tables

10.3 Network visualization of the tag ‘worldofwarcraft’ 150
10.4 Network visualization of the Flickr tag ‘FPS’ 153
10.5 Author’s ego network of Flickr contacts 155
10.6 Network of connections of the most prolific poster in the ‘game
screenshots’ Flickr user group 155
10.7 Visualization of the Flickr user Jokay’s ego network of contacts and
images commenter 156
10.8 Twitter network overview of user ‘Jokay’ 157
10.9 Zoomed view of the lower-center cluster of Figure 10.8 157
12.1 Australian novelists—number of novels written and number of
‘works about’ 179
12.2 Katharine Susannah Prichard: number of novels by and ‘works
about’, 1900–2006 181
12.3 Total number of ‘works about’ 182
12.4 Number of unique authors in the yearly top ten for ‘works about’,
per decade 183
12.5 Proportion of critical attention to top ten authors, per decade 185
12.6 Proportion of authors who feature in the top 50 for both academic
and newspaper criticism, 1950–2006 187
12.7 Proportion of women authors in the top ten, 20, and 50 for
academic attention, 1950s–2000s 188
12.8 Proportion of women authors in the top ten, 20, and 50 for
newspaper attention, 1950s–2000s 188
15.1 Port Botany wharf data visualization 232
15.2 Port Botany road data visualization 234

Tables

3.1 Forty most significant keywords of the Dickens corpus as
compared with the Wilkie Collins corpus (sorted according to the
log-likelihood ratios [LLR]) 29
3.2 A result of running Random Forests 31
3.3 Important variables: Dickens markers and Collins markers
(in order of importance) 32
3.4 Important V: Dickens versus the reference corpus (in order of
importance) 33
3.5 Frequent 5-grams including ‘eyes’ in the Dickens corpus 35
3.6 Frequent 5-grams including ‘eyes’ in the eighteenth-century
corpus 35
3.7 Frequent 5-grams including ‘eyes’ in the nineteenth-century
corpus 36
3.8 Frequent 5-grams including ‘hands’ in the Dickens corpus 38
3.9 Frequent 5-grams including ‘hands’ in the eighteenth-century
corpus 38
3.10 Frequent 5-grams including ‘hands’ in the nineteenth-century corpus 39
12.1 Properties of the dataset extracted from AustLit 178
12.2 Proportion of critical attention to top ten authors, per decade 186
13.1 Example of stepwise formalization from text to vector data 196
Contributors

Paul Longley Arthur is Professor of Digital Humanities, University of Western Sydney. He was previously Deputy Director, Centre for European Studies, a joint initiative of the European Commission and Australian National University. From 2010–13 he was Deputy Director, National Centre of Biography, Australian National University, and Deputy General Editor, Australian Dictionary of Biography.

Katherine Bode is Senior Lecturer in Literary and Textual Studies at the Centre for Digital Humanities Research at the Australian National University. Her most recent monograph, Reading by Numbers: Recalibrating the Literary Field (2012), explores the critical potential of quantitative book historical and bibliographical methods for literary history.

Axel Bruns is an Australian Research Council Future Fellow and Professor at Queensland University of Technology. He is the author of Blogs, Wikipedia, Second Life and Beyond: From Production to Produsage (2008) and Gatewatching: Collaborative Online News Production (2005). His current work develops new methodologies for studying public communication in social media.

Jean Burgess is Associate Professor of Digital Media Studies, Director of Research Training for the Creative Industries Faculty, and Deputy Director of the Australian Research Council Centre of Excellence for Creative Industries & Innovation at Queensland University of Technology. Her research focuses on social and mobile media platforms.

Mark Byron teaches and publishes across the genres and practices of Modernism as well as textual and editorial theory. His current work developing digital scholarly editions of complex Modernist texts and their manuscripts includes the Watt module of the Samuel Beckett Digital Manuscript Project. He is the author of Ezra Pound’s Eriugena (2014).

Mark Coté is Programme Director of the Masters in Digital Culture and Society at King’s College London. Mark writes on social media, digital culture, and the relationship between the human and technology. He is currently examining ‘big social data’ through an Arts and Humanities Research Council-funded research project with Tobias Blanke.

Øyvind Eide is a lecturer and research associate at the Chair of Digital Humanities, University of Passau, Germany. His research interests focus on the modelling of cultural heritage information, especially as a tool for critical engagement with the relationships between texts and maps as media of communication.
Jack Elliott is a doctoral candidate at the University of Newcastle. His dissertation is on machine learning and romance novels.

Julia Flanders is Professor of Practice in English and Director of the Digital Scholarship Group at Northeastern University. She directs the Women Writers Project and serves as editor in chief of Digital Humanities Quarterly. Her research focuses on digital text representation and editing, and on theories of work in digital humanities.

Tim Highfield is a Research Fellow at Queensland University of Technology and Curtin University. His research examines political and cultural uses of social media, including information flows around breaking news and popular culture. His research website is located at http://timhighfield.net/.

Alan Liu is Professor of English at the University of California, Santa Barbara. His books include The Laws of Cool: Knowledge Work and the Culture of Information (2004) and Local Transcendence: Essays on Postmodern Historicism and the Database (2008). He is founder of the 4Humanities.org advocacy initiative.

Willard McCarty is Professor in the Department of Digital Humanities, King’s College London, and in the Digital Humanities Research Group, University of Western Sydney; Editor of Interdisciplinary Science Reviews (2008–); founding Editor of Humanist (1987–); and recipient of the Roberto Busa Award (2013), Alliance of Digital Humanities Organizations. See www.mccarty.org.uk/.

Richard Maltby is Professor of Screen Studies and Executive Dean of the Faculty of Education, Humanities and Law at Flinders University, South Australia. His publications include Hollywood Cinema (2nd ed. 2003), ‘Film Europe’ and ‘Film America’: Cinema, Commerce and Cultural Exchange, 1925–1939 (1999), and Explorations in New Cinema History: Approaches and Case Studies (2011).

Christopher Moore is Lecturer in Digital Media and Communication in the School of Arts, English and Media at the University of Wollongong. He is a researcher in games studies and is currently investigating the role of online personas within indie and independent cultures of games production.

Tara Murphy is Senior Lecturer in the School of Physics at the University of Sydney. She is currently leading the VAST project to investigate astronomical objects that vary on rapid timescales. Tara is also interested in interdisciplinary applications of novel computer science techniques to data-intensive research in astronomy and beyond.

Peter Robinson is Bateman Professor of English at the University of Saskatchewan. He is developer of the texual editing program Collate, used by many textual editing projects worldwide, and of the Anastasia electronic publishing system.
He is active in the development of standards for digital resources and tools for collaborative editing.

**Ned Rossiter** is Professor of Communication at the Institute for Culture and Society at the University of Western Sydney and teaches in the School of Humanities and Communication Arts. He is currently working on a collective project entitled *Logistical Worlds: Infrastructure, Software, Labour* (logisticalworlds.org).

**Sydney J. Shep** is Reader in Book History at Victoria University of Wellington and The Printer at VUW’s Wai-te-ata Press: Te Whare Tā o Wai-te-ata. She directs the Print Culture eResearch Hub which hosts ‘The Printers’ Web’, the New Zealand Reading Experience Database, and the Digital Colenso, a prosopographical collaboratorium.

**Tomoji Tabata** is Associate Professor of Language Informatics at the Graduate School of Language and Culture, University of Osaka. His primary research interests are in forensic stylometry and corpus–stylistic investigation of the language of eighteenth- and nineteenth-century British fiction with special reference to Charles Dickens.

**Paul Turnbull** is Professor of Digital Humanities at the University of Tasmania, and an Honorary Professor of eHistory at the University of Queensland. His digital research projects include ‘South Seas’, a major online resource focused on James Cook’s Pacific voyaging; The Gugu Badhun Oral History; the Digital Guide to Sources in North Queensland History; and PaperMiner: Spatial Mapping of Australian Historical Newspapers.

**Deb Verhoeven** is Chair and Professor of Media and Communication at Deakin University. She is a founding member of the Australasian Association for the Digital Humanities (aDH), Project Director of the Humanities Networked Infrastructure initiative (HuNI), and a member of the Tasmanian government’s Digital Futures Advisory Committee.

**Dylan Walker** is a doctoral student in screen studies at Flinders University, South Australia, researching film distribution and exhibition in rural South Australia during the 1930s. He has published articles on rural cinema exhibition and his book, *Adelaide’s Silent Nights* (1995), deals with the history of cinemas in Adelaide, South Australia, during the silent era.

**Mike Walsh** is Associate Professor in the Screen and Media Department of Flinders University in Adelaide, Australia. He has published regularly on Australian cinema and is currently completing a history of the South Australian Film Corporation. He is also a programmer and writer for the Adelaide Film Festival.