

1. Record Nr.	UNINA9910288757203321
Autore	Moran Steven
Titolo	The Unicode cookbook for linguists : managing writing systems using orthography profiles // Steven Moran, Michael Cysouw
Pubbl/distr/stampa	Berlin, : Language Science Press, 2018 Berlin, Germany : , : Language Science Press, , [2018] ©2018
ISBN	3-96110-090-X
Descrizione fisica	1 online resource (vii, 132 pages) : PDF, digital file(s)
Collana	Translation and multilingual natural language processing ; ; 10
Disciplina	005.722
Soggetti	Unicode (Computer character set) Language and languages - Orthography and spelling
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Chapter 1: Writing systems -- Chapter 2: The Unicode approach -- Chapter 3: Unicode pitfalls -- Chapter 4: The International Phonetic Alphabet -- Chapter 5: IPA meets Unicode -- Chapter 6: Practical recommendations -- Chapter 7: Orthography profiles -- Chapter 8: Implementation
Sommario/riassunto	This text is a practical guide for linguists, and programmers, who work with data in multilingual computational environments. We introduce the basic concepts needed to understand how writing systems and character encodings function, and how they work together at the intersection between the Unicode Standard and the International Phonetic Alphabet. Although these standards are often met with frustration by users, they nevertheless provide language researchers and programmers with a consistent computational architecture needed to process, publish and analyze lexical data from the world's languages. Thus we bring to light common, but not always transparent, pitfalls which researchers face when working with Unicode and IPA. Having identified and overcome these pitfalls involved in making writing systems and character encodings syntactically and semantically interoperable (to the extent that they can be), we created a suite of open-source Python and R tools to work with languages using

orthography profiles that describe author- or document-specific orthographic conventions. In this cookbook we describe a formal specification of orthography profiles and provide recipes using open source tools to show how users can segment text, analyze it, identify errors, and to transform it into different written forms for comparative linguistics research.

---