Research on Tibetan Spelling Formal Language and Automata with Application

Nyima Tashi

Research on Tibetan Spelling Formal Language and Automata with Application

Science Press
Beijing

Springer

Nyima Tashi
Tibet University
Lhasa, Xizang
China

ISBN 978-981-13-0670-9 ISBN 978-981-13-0671-6 (eBook)
https://doi.org/10.1007/978-981-13-0671-6

Jointly published with Science Press, Beijing, China

The printed edition is not for sale in the Mainland of China. Customers from the Mainland of China please order the print book from: Science Press.

Library of Congress Control Number: 2018942155

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
part of Springer Nature
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

Tibetan information technology is one of the important research fields of Chinese information technology, and the research on Tibetan information technology is of great significance for inheriting and carrying forward the excellent ethnic culture and promoting the informatization construction and economic and social development in Tibetan areas. To research the Tibetan grammar system including spelling grammar from the viewpoint of information technology is the core and basic research work for Tibetan information technology.

The writer has launched the research on Tibetan information technology since the later 1980s and has continuously analyzed and summarized the technical essence and theoretical connotation of Tibetan information technology during the long-term research. Through the practical accumulation and theoretical improvement for many years, the writer wrote the academic monograph of *Research on Tibetan Spelling Formal Language and Automata with Application.* Since he is a Tibetan professor majored in computer science and also acts as the director of library in Tibet University, the writer can realize the intensified integration of Tibetan knowledge, computer technologies, and Tibetan information technology, so as to ensure that this book possesses not only systematicness but also theoretical depth.

This book mainly features as follows: ① The Tibetan spelling grammar are analyzed and induced as fully as possible from the viewpoint of information technology; ② Formal language and automata theory are introduced into the field of Tibetan information technology, the Tibetan spelling formal language is studied, and the inherent law of Tibetan spelling grammar is revealed from the viewpoint of information technology, so as to pave the theoretical foundation for intensifying the research on Tibetan information technology; ③ On the basis of theoretical research, the application of theoretical method in optimized design and efficient realization of computer-based automatic spelling check, automatic sorting, and intelligent input method of Tibetan is detailed.

I believe that the publication of this book will extend the valuable help to researchers engaged in research on Chinese information technology, Tibetan information technology, Tibetan computational linguistics and natural language processing.

Beijing, China
March 2016

Pan Yunhe
Academician of Chinese Academy of Engineering

Preface

When I started the R&D on Tibetan information technology in the later 1980s, I felt that the previous R&D was mainly focused on the application level of Tibetan information technology and there was few basic research on the object of research (namely Tibetan and its grammar) from the viewpoint of information technology. According to the laws of development of information technology, the level of research, development, and innovation can be improved only after the basic theoretical research has been properly carried out and the technical bottleneck has been overcome. To research the Tibetan spelling grammar and word structure from the viewpoint of information technology and explore into the basic theory and application method of Tibetan spelling formal language and its automata is of great significance for improving the level of research and application of Tibetan information technology.

Tibetan is a kind of alphabetic writing, but its spelling mode is different from other languages such as English. The Tibetan has to be not only transversely spelled but also longitudinally spelled, with a nonlinear two-dimensional structure. The spelling sequence of Tibetan is as follows: prefix character, superfix character, root character, subfix character, vowel, suffix character, and postfix character, among which the superfix character, root character, subfix character, and vowel are longitudinally spelled in an overlapped manner. The root character is the core of a Tibetan word, and recognizing the root character and then every constituent of a Tibetan word is a key technology which involves many research fields of Tibetan information technology. When I was studying for a doctorate of computer application in Sichuan University, I had an idea of researching the Tibetan spelling formal language by using the formal language and automata theory and applying it into the research on recognition of constituents of Tibetan words, and I carried out the exploratory research in my doctoral dissertation and subsequent works. In 2012, I completed the draft of this book, but still felt that there were many technical issues which had to be further explained and the research result would be verified through actual application. Therefore, I adjusted the structure of the draft and added the contents on the application of Tibetan spelling formal language in computer

automatic spelling check and automatic sorting of Tibetan. More than three years later, this book was finally completed.

During the writing of this book, I received a great deal of support from many friends and colleagues. I would like to extend my sincere gratitude to all the friends and colleagues who have supported and helped me, and to my family members who have cared and encouraged me for many years!

Owing to my limited capacity, it is inevitable that there may exist something wrong in this book. All the readers are kindly requested to point them out for me.

Lhasa, China Nyima Tashi
July 2016

Contents

Introduction

This book introduces formal language and automata theory into a field of Tibetan information technology, constructs formal grammar to generate Tibetan spelling formal language and automata to recognize the language, and studies the application of Tibetan spelling formal language and its automata. There are seven chapters in this book, among which Chap. 1 briefly introduces the Tibetan language; Chap. 2 analyzes and summarizes the Tibetan spelling grammars and the basic spelling structure of Tibetan words; Chap. 3 introduces the basic knowledge about formal language; Chap. 4 formally describes the inherent Tibetan spelling grammars; Chap. 5 sets forth the formal grammars used to generate the Tibetan spelling formal languages and the automata used to recognize the languages; Chap. 6 introduces the application method of Tibetan coding standard; and Chap. 7 introduces the actual application of Tibetan spelling formal language.

This book has a referential value for researchers engaged in research on computational linguistics, information technology, and Tibetan computational linguistics and can also be used by professional teachers and graduate students majored in Tibetan information technology and Tibetan computational linguistics.