Setlur / Govindaraju

# Guide to OCR for Indic Scripts

Document Recognition and Retrieval

ofresearchers,scholars,andstudentsuseanddependuponInternet-basedcontent andcomputationalresources. Thechaptersinthisbookdescribeacriticallyimportantareaofinvestigation– addressingconversionofIndicscriptintomachine-readableform. Roughestimates haveitthatcurrentlymorethanabillionpeopleuseIndicscripts. Collectively,Indic historic and cultural documents contain a vast richness of human knowledge and experience. The state-of-the-art research described in this book demonstrates the multiple values associated with these activities. Technically, the problems associated with Indicscriptrecognitionareverydif?cultandwillcontributetoandinformrelated v vi Foreword scriptrecognitionefforts. Theworkalsohasenormousconsequenceforenriching andenablingthestudyofIndicculturalheritagematerialsandthehistoricrecord of its people. This in turn broadens the intellectual context for domain scholars focusingonothersocieties,ancientandmodern. Digital character recognition has brought about another milestone in coll- tivecommunicationbybringinginert,?xed-in-place,textintoaninteractivedi- talrealm. Indoingso,theinformationhasgainedadditionalfunctionalitieswhich expandourabilitiestoconnect,combine,contextualize,share,andcollaboratively pursue knowledge making. High-quality Internet content continues to grow in an explosivefashion. Inthenewglobalcyberenvironment,thefunctionalitiesandapp-cationsofdigitalinformationcontinuetotransformknowledgeintonewwndersta-ingsofhumanexperienceandtheworldinwhichwelive. Thepossibilitiesforthe futurearelimitedonlybyavailableresearchresourcesandcapabilitiesandtheim-inationandcreativityofthosewhousethem. Arlington,Virginia StephenM.

Optical Character Recognition (OCR) is a key enabling technology critical to creating indexed, digital library content, and it is especially valuable for Indic scripts, for which there has been very little digital access. Indic scripts, the ancient Brahmi scripts prevalent in the Indian subcontinent, present some challenges for OCR that are different from those faced with Latin and Oriental scripts. But properly utilized, OCR will help to make Indic digital archives practically accessible to researchers and lay users alike by creating searchable indexes and machine-readable text repositories. This unique guide/reference is the very first comprehensive book on the subject of OCR for Indic scripts, providing an overview of the state-of-the-art research in this field as well as other issues related to facilitating query and retrieval of Indic documents from digital libraries. All major research groups working in this area are represented in this book, which is divided into sections on recognition of Indic scripts and retrieval of Indic documents. Topics and features: - Contains contributions from the leading researchers in the field - Discusses data set creation for OCR development - Describes OCR systems that cover eight different scripts: Bangla, Devanagari, Gurmukhi, Gujarati, Kannada, Malayalam, Tamil, and Urdu (Perso-Arabic) - Explores the challenges of Indic script handwriting recognition in the online domain - Examines the development of handwriting-based text input systems - Describes ongoing work to increase access to Indian cultural heritage materials - Provides a section on the enhancement of text and images obtained from historical Indic palm leaf manuscripts - Investigates different techniques for word spotting in Indic scripts - Reviews mono-lingual and cross-lingual information retrieval in Indic languages This is an excellent reference for researchers and graduate students studying OCR technology and methodologies. This volume will contribute to opening up the rich Indian cultural heritage embodied in millions of ancient and contemporary documents spanning topics such as science, literature, medicine, astronomy, mathematics and philosophy. Venu Govindaraju FIEEE FIAPR, is a Distinguished Professor of Computer Science and Engineering at the University at Buffalo. He has over 20 years of research experience in pattern recognition, information retrieval and biometrics. His seminal work on handwriting recognition was at the core of the first handwritten address interpretation system used by
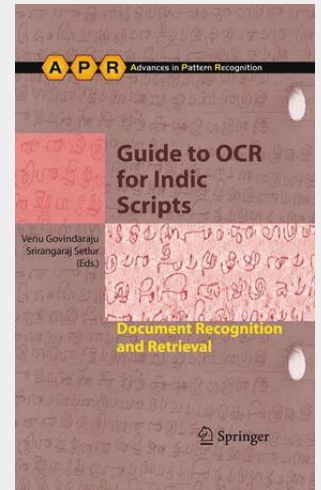
160,49 €
149,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage