Abstract

India’s heritage texts have had a long history of being mined for knowledge of language and culture by Christian missionaries to India, colonial officers of the East Indian Company and the British Raj, German, European and American Indologists and later by native scholars driven by nationalist sentiments. It was during their investigative exercises that a vast body of India’s heritage texts was recovered and made the subject of rigorous study. A large number of editions in English translation as well as in modern Indian vernacular languages started appearing on the scene. The focus then was primarily on patthoddhar [retrieval of the ‘ur’-text] or making shuddhasanskarana [correct edition]. The exercise was purely manual and time-consuming and concentrated on a limited number of texts. But there still lies a vast treasure of ancient knowledge in India’s palm leaf manuscripts, waiting to be discovered, deciphered and interpreted for contemporary readers and scholars. It is impossible to ignore the ubiquitousness of Information Technology based tools and the scope that they offer for large-scale data mining. Of late, a large body of historical texts is being made available digitally.
by repositories and institutions worldwide. The time is ripe for digitally inspired editions, beginning with studies in corpus linguistics. This paper throws light on the challenges to be addressed for the preparation of a digital historical corpus edition of Sarala Mahabharata, a local version of the famous Sanskrit Mahabharata by Vyasa, from Odisha in the eastern part of India.

References


Index Terms

Computer Science

Information Sciences

Keywords
Sarala Mahabharata, Text Mining, Odia, Digitization, Palm leaf manuscript