

Character Order Models In Sindhi Language and Their Exploitation In Compression Techniques

Sumbul Khawaja & Asadullah Shah

Sindhi language consists of 52 characters. The occurrence of these characters can be exploited for compression either individually or collectively. The occurrence of two characters together and many characters together is normally exploited in international languages such as English and others for compression purposes.

In this research we have studied and determined the higher order models for Sindhi language character set. The study indicates that there are many words with the characters that are always dependent upon the occurrence of each other. If one character occurs, the probability of occurrence of another is always very high. This means that many characters can be predicted on the basis of dependencies upon each other. The compression techniques can be developed on the basis of the order models.
