Abstract

This thesis proposes a robust and innovative framework for the extraction, classification, and visualization of ontologies related to characters, topics, and their interrelationships within Indian mythological texts, particularly focusing on the *Ramayana*, the *Mahabharata*, the *Devi Bhagavata*, the *Harivamsha Purana*, the *Srimad-Bhagavatam*, the *Caitanya Caritamrita*, and *Krsna, the Supreme Personality of Godhead*. Using advanced Natural Language Processing (NLP) techniques, this work addresses the complex challenges posed by these intricate and expansive narratives. The research integrates state-of-the-art transformer models and methodologies to enhance character identification, thematic classification, and ontology development.

In the first segment, the thesis explores **Character Identification** within these ancient texts, utilizing linguistic feature extraction and machine learning models like Neural Networks and KNN classifiers. By developing a dedicated dataset for mythological characters, the study successfully disambiguates character names and their associated attributes, contributing to a more precise recognition system.

The second section presents **Pouranic Topic Classification**, introducing transformer-based models such as BERT, RoBERTa, and DistilBERT for classifying topics across mythological texts. The development of a novel annotated dataset, *PouranicTopic*, allows for accurate classification of cantos, and topics, tackling the complexity of overlapping topics and enhancing the understanding of these mythological texts.

Next, **Ontology and Character-Topic Relationship** is explored through the creation of SBC-Ontology and SBT-Ontology, offering a structured framework that maps characters to their associated topics. By employing advanced transformer models, the study uncovers deeper insights into the relationships between characters and the topics they represent in texts like the *Srimad-Bhagavatam*.

The fourth contribution is **MythoBERT 1.0**, a custom-built language model designed for Indian hindu mythology. MythoBERT surpasses general models in tasks like Named Entity Recognition (NER), text classification, and topic modeling, offering a specialized

tool for understanding the rich linguistic and cultural nuances of mythological texts.

The thesis also introduces **Mytho-Annotator**, a web-based annotation tool tailored to Hindu mythological texts. This tool provides an efficient framework for labeling entities, relationships, and events, accelerating the annotation process and enabling the creation of high-quality datasets for further research.

Finally, the study delves into the **Visualization of Character-Centric Summaries**, employing models like T5, BART, and PEGASUS to generate summaries focused on individual characters. KeyBERT is used to extract key phrases, linking them to specific characters and enhancing the interactive exploration of these texts through visual summaries.

Overall, this thesis significantly advances the application of NLP in Indian mythology, offering new models and insights for character identification, topic classification, ontology development, domain specific language model, annotation tool and finally the visualization of mythological characters of complex narratives. These contributions have broad implications for digital humanities, cultural preservation, and the creation of educational tools, enabling deeper engagement with India's hindu mythological heritage.