Mytho-Annotator: An Annotation tool for Indian Hindu Mythology

Apurba Paul[1], Anupam Mondal[2], Sainik Kumar Mahata[1], Srijan Seal[3], Prasun Sarkar[3], Dipankar Das[1]
[1]Jadavpur University, Kolkata, India
[2]Institute of Engineering & Management, Kolkata, India
[3]JIS College of Engineering, Kalyani, India
[1]apurba.saitech@gmail.com, [2]link.anupam@gmail.com, [1]sainik.mahata@gmail.com,
[3]srijanseal03@gmail.com, [3]prasunsarkarpersonal@gmail.com, [1]dipankar.dipnil2995@gmail.com

Abstract

Mythology is a collection of myths, especially one belonging to a particular religious or cultural tradition. We observed that an annotation tool is essential to identify important and complex information from any mythological texts or corpora. Additionally, obtaining high-quality annotated corpora for complex information extraction including labeled text segments is an expensive and time-consuming process. Hence, in this paper, we have designed and deployed an annotation tool for Hindu mythology which is presented as ***Mytho-Annotator***. Its easy-to-use web-based text annotation tool is powered by Natural Language Processing (NLP). This tool primarily labels three different categories such as named entities, relationships, and event entities. This annotation tool offers a comprehensive and adaptable annotation paradigm.

1 Introduction

An annotation tool is an application or platform that helps with the process of labeling textual material with different linguistic and semantic properties in the context of Natural Language Processing (NLP) (Mondal et al., 2023). Research and development of natural language processing techniques require access to human-annotated corpora. Developing these corpora, however, is costly and time-consuming. While crowdsourcing and other remote annotation procedures on the web make it possible to quickly collect a large number of annotations, the bottleneck is frequently the annotation tool that is being used and the training that annotators must undergo. Specifically, the majority of widely used annotation tools are meant to be multipurpose. Examples of these include the web-based BRAT (Stenetorp et al., 2012) and WebAnno (Yimam et al., 2014) tools, as well as the integrated editors of GATE (Cunningham, 2002). Their extensive nature results in a more complex interface, which frequently makes the texts to be annotated less readable. Hindu mythology comprises a multitude of intricate and interwoven narratives, often spanning multiple texts and traditions. Annotation tools can be instrumental in enabling a thorough comprehension and examination of the intricate stories, symbols, and themes incorporated into the extensive collection of Hindu mythological texts. Numerous Hindu stories are rich in metaphors, allegories, and symbolism with deep spiritual and philosophical implications. An annotation tool can assist readers understand the deeper meanings buried within the myths by clarifying the cultural, theological, and philosophical importance of these symbols. In this study, we introduce Mytho-Annotator, an open-source web-based annotation tool devoted to various forms of labeling in the field of Hindu mythology. The aim of the tool is to facilitate the deeper understanding of ancient texts and enable advanced computational techniques for analysis and interpretation. It also encourages collaborative research in mythology, fostering new insights and discoveries. The motivation is to enhance the accessibility, accuracy, and collaboration in the annotation and analysis of Hindu mythological texts. All the features needed to effectively manage and operate text labeling projects are included in ***Mytho-Annotator***. Remote annotation procedures will find it especially useful because of its self-descriptive annotation interface, which simply requires a web browser. Additionally, it guarantees that the texts that need to be labeled can still be read throughout the annotation process. This procedure is server-based and subject to preemption at any time. Since all pertinent interactions between the annotators are recorded in a straightforward plain text format with a key-value basis, the annotator's progress may be continuously observed.

This research work is laid out as follows: Section 2 Related Work, Section 3 User Interface and Experience, Section 4 describes System Architecture, Section 5 describes System Description and finally Section 6 gives some concluding remarks and future works.

2 Related Work

In the realm of Natural Language Processing (NLP) annotation tools, (Bikaun et al., 2022) present *Quickgraph*, an annotation tool for knowledge graph extraction from technical text, while (Frei et al., 2022) introduce *DrNote*, an open medical annotation service. (Mondal et al., 2023) used a medical annotation system to prepare a structured medical corpus. (Dutta N., et al., 2020) proposed an annotation system to annotate healthcare information from tweets. (Bojars et al., 2018) develop a Semantic Annotation Tool for Cultural Heritage Content, whereas (Tyers et al., 2017) present *UD Annotatrix*, an annotation tool for universal dependencies. (Kiesel et al. (2017) introduce *WAT-SL*, a web annotation tool for segment labeling, and (de Castilho et al., 2014) propose *WebAnno*, a flexible, web-based annotation tool for CLARIN. (Bollmann et al., 2014) introduce *CorA*, a web-based annotation tool for historical and non-standard language data, while (Stenetorp et al., 2012) present *BRAT*, a web-based tool for NLP-assisted text annotation. (Tesconi et al., 2010) develop *KAFnotator*, a multilingual semantic text annotation tool, and (Kenter et al., 2005) utilize GATE as an annotation tool for various applications. (Maeda et al., 2004) discuss annotation tools for large-scale corpus development, and (Morton and LaCivita, 2003) introduce *WordFreak*, an open tool for linguistic annotation. (Cunningham, 2002) presents *GATE*, a general architecture for text engineering, providing a comprehensive framework for text processing and annotation tasks.

3 System Architecture

Mytho-Annotator is a multi-user tool built using the modern full-stack framework MERN-STACK[1]. Our tool consists of three components i) client side, ii) Mongo DB database, iii) Server side.

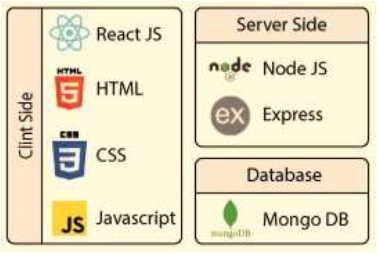

Figure 1: ***Mytho-Annotator's*** system architecture and technology stack.

Mytho-Annotator's Mongo DB database consists of the two collections: Projects and Users. Projects contain information pertinent to the projects: userId, projectName, namedEntities, eventEntities, relations, namedEntityTags, eventEntityTags, namedEntityAppearances, and eventEntityAppearances. Users collection contains information such as the users: username, hashed and salted password, email.

4 System Description

4.1 Text Document Handling

Mytho-Annotator's prowess in text document handling goes beyond mere compatibility. It excels in seamlessly processing a wide array of textual data formats, accommodating the diverse sources of text that researchers and annotators encounter. Whether it's standard text documents, PDFs, web content, or more specialized formats, ***Mytho-Annotator*** ensures flexibility and adaptability which significantly simplifies the intricate process of importing and managing documents for annotation tasks.

4.2 Annotation Sections

Mytho-Annotator's annotation framework is thoughtfully structured into three dedicated sections, each tailored to serve a specific annotation purpose:

4.3 Named Entity Annotation

Within the realm of named entity recognition, ***Mytho-Annotator*** stands out by offering annotators a comprehensive set of predefined categories. These categories encompass a wide spectrum of common named entities, including

[1] https://www.mongodb.com/mern-stack

individuals, organizations, geographic locations, and much more. Annotators benefit from a user-friendly interface that streamlines the selection and annotation of named entities throughout the text.

4.4 Relationship Annotation

Mytho-Annotator's **Relationship Annotation** section is a robust tool designed for capturing intricate connections and associations between entities present in the text. This section features two essential components that work in harmony to facilitate thorough relationship annotation. First, annotators can seamlessly select relevant phrases within the text, pinpointing specific sections that hold significance for relationship identification. This phrase selection capability serves as the foundation for building meaningful relationships. Second, annotators can assign and define relationships between the identified entities, fostering a comprehensive understanding of how different elements within the text interact.

4.5 Event Entity Recognition

Mytho-Annotator extends its annotation capabilities to encompass event entity recognition, a crucial aspect of text understanding. Within the Event Entity Recognition section, annotators have the power to assign events to phrases while considering essential attributes such as modality, frequency, and time. This feature empowers users to capture nuanced information related to events described in the text. Whether it's identifying the likelihood of an event, its occurrence frequency, or temporal context, ***Mytho-Annotator*** provides a versatile framework for annotating these critical details.

5 User Interface and Experience

In our annotation tool, we have thoughtfully structured it into three distinct sections, each aimed at providing users with a comprehensive and user-friendly annotation experience. We understand the importance of efficiency and precision in annotating mythological texts, and our tool has been designed with these considerations in mind. The **"Named Entity Annotation"**, **"Relationship Annotation",** and **"Event Entity Annotation"** sections are the cornerstones of our tool's functionality. Within the **"Named Entity Annotation"** section, we have included a sub-bar that offers two crucial functions: "**Assign Tag**" and **"Assign Gender"**. With **"Assign Tag"**, users have the flexibility to apply a diverse range of tags to specific words or phrases within the text. This feature ensures that annotations are not only accurate but also enriched with contextual information, making it a valuable resource for mythological analysis.

Figure 2: Interface portraying Assign Tag section

In the visual representation (Figure 2) of **"Named Entity Annotation"** depicted in the provided figure, we portray **"Assign Tags"** feature where set a diverse set of tags, including ***"Character", "Color", "Disease", "Event", "Facility"***, and others, is observed. Specifically, ***Lord Caitanya Mahaprabhu*** is identified and annotated as a character. The illustration further features a side bar, introducing a hierarchical structure that visually organizes the annotated entities. Notably, under the category *"Character"* a hierarchical structure is established, with a broader classification denoted as *"Human"***.** Within this hierarchy, a more specific subclass is identified as *"Brahmana"* and this annotation will be reflected in the whole text.

Furthermore, the **"Assign Gender"** function under this section is specifically designed to address the nuanced gender dynamics often present in mythological narratives.

Figure 3: Interface portraying Assign Gender section

In Figure 3, the **"Named Entity Annotation"** module includes a distinctive segment titled **"Assign Gender"**. Within this section, a range of gender categories is presented, encompassing *"Male", "Female", "Eunuch", "Bisexual", "Transgender"* and others.

In the sidebar, we can observe that both ***Lord Caitanya Mahaprabhu*** and ***Nityanada Prabhu*** selected from the text have been designated with the gender label *"Male"* shown in the right side bar. This module facilitates the assignment of specific gender attributes to identified characters within the analyzed text. Such a feature proves valuable in the context of character analysis, allowing for the nuanced classification of individuals based on their gender identity.

Now, in the second section which is **"Relationship Annotation"** section, we have incorporated two separate yet seamlessly integrated components. The first component allows users to select phrases within the text, providing a foundation for identifying key elements for relationship annotation. This phrase selection process ensures that users can pinpoint specific sections of text that are relevant to the relationships they intend to annotate.

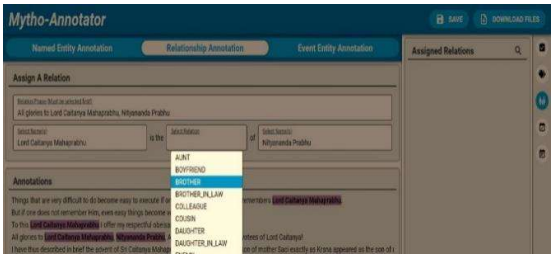

Figure 4: Interface portraying Relationship Annotation

In Figure 4, the illustration showcases **Relationship Annotation**. The accompanying sentence provides contextual information, revealing that ***Lord Caitanya Mahaprabhu*** is associated with the relationship label *"brother"* in relation to ***Nityananda Prabhu***. This feature elucidates the capability of the system to discern and annotate intricate relationships between characters, enriching the semantic understanding of their affiliations within the given textual context.

Within the **"Event Entity Annotation"** section, our tool offers a versatile platform for annotating events and associating them with specific text phrases. This feature allows users to capture information related to events, including their modality, frequency, and temporal attributes. Whether it's identifying the certainty of an event (*Modality*), specifying how often it occurs (*Frequency*), or noting the time it takes place (*Time*), our annotation tool empowers users to comprehensively capture event-related information within the text.

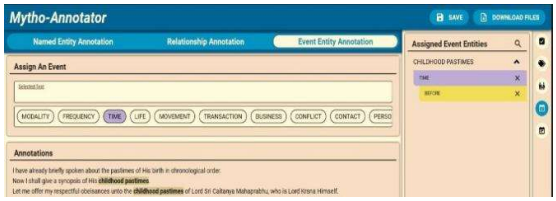

Figure 5: Interface portraying Event Entity Annotation

In Figure 5, the depiction centers on **"Event Entity Annotation"**, featuring various entities such as modality, frequency, time, and others. Notably, when the specific text segment ***"childhood pastimes"*** is selected the associated entity assigned is *"time"*. Furthermore, a contextual insight is provided through the accompanying right-hand bar, where the event is characterized as representing the past, and the temporal relationship is specified as *"before"*. This nuanced annotation of temporal attributes within the Event Entity adds a layer of sophistication, allowing for a more detailed analysis of the chronological aspects associated with events in the annotated text.

To enhance user experience and provide a visual overview of the annotation process, we have incorporated a dedicated section where we can see the results and well as the hierarchical structure.

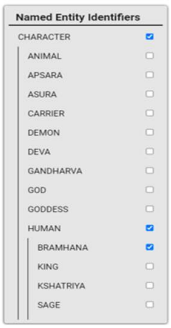

Figure 6: **Named Entities**

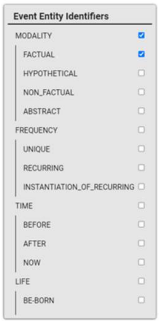

Figure 7: **Event Entities**

Thus, this visual representation (Figure 6 and Figure 7) serves as a dynamic dashboard, representing the hierarchical relationships of **Named Entities** (Figure 6) and that of **Event Entities** (Figure 7) validate their annotations, and refine their work as needed. Overall, our annotation tool is designed to cater to the unique demands of annotating mythological texts. Whether users are identifying named entities, capturing intricate relationships, or recognizing significant event entities, our tool serves as a reliable and user-

centric companion, facilitating a seamless and productive annotation journey.

6 Conclusion and Future Works

Looking ahead, ***Mytho-Annotator*** holds substantial potential for further development and expansion in several key areas. The focus lies in enhancing entity recognition, ensuring both accuracy and flexibility, and extending support for domain-specific entities. Another avenue involves enhancing interoperability with other annotation tools and platforms, supporting common data exchange formats and protocols for streamlined collaboration. Scalability and improved collaboration for large-scale annotation projects are paramount, featuring features for distributed tasks and efficient project management. The tool's future also involves exploring advanced relationship extraction techniques, including automated inference of complex semantic relationships from unstructured text data. The integration of machine learning models for semi-automated annotation is on the horizon, providing machine learning-driven suggestions and predictions.

References

Bikaun, T., Stewart, M., & Liu, W. (2022, May). Quickgraph: A rapid annotation tool for knowledge graph extraction from technical text. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (pp. 270-278).

Frei, J., Soto-Rey, I., & Kramer, F. (2022). DrNote: An open medical annotation service. PLOS Digital Health, 1(8), e0000086.

Mondal A., Cambria E., & Dey M. (2022),An annotation system of a medical corpus using sentiment-based models for applications, Computational Intelligence Applications for Text and Sentiment Data Analysis: Book Chapter(pp. 163-178)

Dutta, N., Mondal, A., Paul, P. (2020). An Annotation System to Annotate Healthcare Information from Tweets. In: Mandal, J., Bhattacharya, D. (eds) Emerging Technology in Modelling and Graphics. Advances in Intelligent Systems and Computing, vol 937. Springer, Singapore. https://doi.org/10.1007/978-981-13-7403-6_30

Bojars, Uldis & Rašmane, Anita & Žogla, Artūrs & Balina, Signe & Salna, Edgars. (2018). Semantic Annotation Tool for Cultural Heritage Content. Baltic Journal of Modern Computing. 6. 10.22364/bjmc.2018.6.4.09.Morton,

Tyers, F. M., Sheyanova, M., & Washington, J. N. (2017). Ud annotatrix: An annotation tool for universal dependencies.

Kiesel, J., Wachsmuth, H., Al Khatib, K., & Stein, B. (2017, April). WAT-SL: a customizable web annotation tool for segment labeling. In Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics (pp. 13-16).

de Castilho, R. E., Biemann, C., Gurevych, I., & Yimam, S. M. (2014). WebAnno: a flexible, web-based annotation tool for CLARIN. In Proceedings of the CLARIN Annual Conference (CAC) (pp. 4372-4380).

Bollmann, M., Petran, F., Dipper, S., & Krasselt, J. (2014, April). CorA: A web-based annotation tool for historical and other non-standard language data. In Proceedings of the 8th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH) (pp. 86-90).

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 102–107, Avignon, France. Association for Computational Linguistics.

Tesconi, M., Ronzano, F., Minutoli, S., Aliprandi, C., & Marchetti, A. (2010, January). KAFnotator: a multilingual semantic text annotation tool. In The Second International Conference on Global Interoperability for Language Resources (Vol. 1).

Kenter, T., & Maynard, D. (2005). Using gate as an annotation tool. University of Sheffield, Natural language processing group.

Maeda, K., & Strassel, S. M. (2004, May). Annotation Tools for Large-Scale Corpus Development: Using AGTK at the Linguistic Data Consortium. In LREC.

Morton, Thomas & LaCivita, Jeremy. (2003). WordFreak: An Open Tool for Linguistic Annotation.. HLT/NAACL 2003: demonstrations. 10.3115/1073427.1073436.

Hamish Cunningham. 2002. GATE, a General Architecture for Text Engineering. Computers and the Humanities, 36(2):223–254.