Overview of the project

Around the world, many cultures are significantly threatened by the erosion of cultural integrity, climate change, loss of habitat, the environmental impact of globalization, and the ravages of epidemics. Cultures that do not have a strong written tradition are especially threatened – as younger generations move away in search of education and jobs, and as globalization forces linguistic shifts to more "global" languages such as English and Spanish, rich oral traditions either disappear completely or become artifacts in university or library archives, accessible only to small groups of academics. Many initiatives have been activated to preserve and revitalize indigenous cultures, traditions, and languages. While quite a few of these initiatives involve the use of computational technologies and multimedia to document and archive indigenous languages, knowledge, and cultures, much of what has been created to date is primarily designed to serve academic communities and does not provide access to either the broader public (e.g., anyone who wishes to learn about a particular culture) or to the indigenous communities that provided the data in the first place. Moreover, existing systems tend to focus on a single corpus of stories and do not provide possibilities to connect and map stories, characters, and concepts across multiple cultures and languages.

We propose to develop *StoryWorlds ITK (Indigenous Traditional Knowlege Worlds)*, an open-source web-based system for collecting, storing, connecting, and presenting Indigenous traditional knowledge (ITK) in the form of stories, myths, and testimonies from multiple cultures and languages. Moreover, the proposed system will allow users to connect ITK narratives based on topics, characters, environments, storylines, and other criteria (e.g., a particular animal appears in X stories from X cultures and acts as a positive character in X of the stories and as a negative character in the remaining X stories).

The proposed system will offer three sets of user interfaces (UIs): (1) a UI for academic users (i.e., anthropologists, linguists, botanists, ornithologists, and traditional knowledge specialists to support the uploading of stories, field notes, and associated media, concept mapping, and creation of ontologies; (2) a text-based search UI to allow any user (but especially younger indigenous users) to search through stories using multiple combinations of keywords and concept and present results in text-based or tabular format; (3) a graph-based search UI to allow any user to search through stories using multiple combinations of keywords and concept and present results as an interactive visual network of concepts.

Visualizing and exploring ITK with knowledge graphs can provide a new perspective on traditional stories and their themes, motifs, and characters. Techniques such as network analysis, tree maps, and sentiment analysis can reveal patterns and relationships within ITK that might otherwise be overlooked. For example, network analysis can uncover the relationships between the characters in a story, while sentiment analysis can help identify the emotions associated with them. Graph-based approaches to oral traditions can also be useful in comparative studies across cultures or time periods. Research into representing oral literature with knowledge graphs can uncover unique insights into traditional stories and their cultural significance. Imagine, for example, if any member of the public could quickly compare the role of a coyote in relation to various plants and animals along with key emotions evoked by a coyote across different languages of the US and Mexico. Imagine also that indigenous experts within those various traditions could upload new versions they had heard. A much more nuanced animal character might emerge than might be simply captured by the reductive category "trickster".

Tasks

Task 1: Refine and populate the graph database. A graph database is a specialized type of database where instead of storing data in the form of Excel-like tables, the data is stored as a collection of nodes where each node represents an object (e.g., story, character, animal). These nodes are linked by edges - connections that represent physical or semantic relationships between pairs or groups of nodes (Appendix 1, Figure 1).

This data representation allows a meaningful mapping of stories to concepts, physical objects, locations, characters, and to each other. For this project, we will be using Neo4J, the world's leading open-source graph database. As part of **Task 1**, we will focus on populating the database with the content from indigenous narratives, identifying connections between stories, characters, environments, locations, and languages.

Task 2: Design and implement a data entry/collection system. While the pilot of *StoryWorlds* will focus on the Amazonian Kichwa tradition, the long-term goal of the proposed system is to support the collection, annotation, and linking of traditional oral narratives across multiple cultures and languages. **Task 2** will focus on iteratively developing an intuitive web-based UI that would allow anthropologists, folklorists, linguists, and historians to upload stories and related media, annotate characters, environments, languages, and other relevant artifacts within each folkloric object, and link these concepts to related concepts in stories already stored and represented in the *StoryWorlds* system. The data collection system will be developed using the Python programming language and Flask web server for managing connections to the graph database and data read/write operations. The UIs will be developed using a combination of HTML5, Bootstrap CSS framework, and ReactJS framework (Appendix 1, Figure 2).

Task 3: Design and implement data search and display UIs. Activity 4 will be executed in parallel with **Task 2** and will focus on developing and validating UIs that will make folklore data in *StoryWorlds* accessible not just to academics, but also to the broader public. The public-facing web-based UI will allow users to search stories by keywords and filter by cultures, languages, and geographic regions. The UI will be capable of displaying search results as text (e.g., tabular format), or as an interactive graph. With the tabular data display, users will be able to click through links to "drill down" through possible connections between stories, characters, etc. With the graph data display, users will be able to see clusters of related nodes, zoom in on individual clusters or nodes, and expand and navigate sub-trees of related nodes representing folkloric artifacts. Through both search modalities, users will be able to explore relevant stories and see/experience multimedia and other artifacts associated with each story.

Skills Required

1. Python
2. Flask
3. JavaScript
4. D3js
5. Neo4j graph database