How are NLP technologies being used for the processing, teaching and preservation of literature?

**Business Report: NLP Project on Icelandic-English Translation and Textual Entailment for Education Enhancement in Iceland**

Date: May 22, 2023

Executive Summary:

This business report outlines the implementation of a Natural Language Processing (NLP) project aimed at translating Icelandic books into English while also analysing and establishing relationships between characters through textual entailment.The proposed application will generate relational graphframes to summarise the relationships between all the characters. The project will leverage Icelandic-Ner-Roberta.

The objective of this project is to leverage NLP technologies to enhance the education system in Iceland by facilitating comprehension of books, thus improving reading skills and fostering a deeper understanding of literary works. This report provides an overview of the project, its potential benefits, and the proposed implementation plan.

**Introduction:**

The education system in Iceland places great importance on literacy skills and reading comprehension. However, the availability of translated materials and resources for Icelandic literature is limited, making it challenging for students to access and comprehend these works. This NLP project aims to address this issue by leveraging automated translation techniques and textual entailment analysis to enhance the educational experience.

Project Goals:

a. **Translation:** The primary objective of this project is to develop a robust NLP model capable of accurately translating Icelandic books into English. By providing translated versions of Icelandic literary works, we aim to broaden the accessibility and availability of these books, making them more accessible to a wider audience.

b. **Textual Entailment**: In addition to translation, the project focuses on establishing relationships between characters through textual entailment with social networks. By analysing the text and identifying logical relationships between various characters, the NLP model will enable students to gain a deeper understanding of the storyline, character interactions, and plot development.

**Benefits for the Education System:**

a. **Improved Comprehension**: By providing translated versions of Icelandic books, students whose English proficiency is stronger than their Icelandic skills will have access to a wider range of literature. This will enable them to enhance their reading skills, broaden their understanding of different literary genres, and develop a deeper appreciation for literature.

b. **Enriched Language Learning:** The availability of translated texts will also serve as a valuable resource for language learning. Students studying English as a second language will be able to compare the translated version with the original Icelandic text, aiding in their language acquisition and cultural understanding.

c. **Enhanced Analytical Skills:** The inclusion of textual entailment analysis will empower students to identify and analyse relationships between characters.The graphical summaries aid users in comprehending complex character networks and significant occurrences.This skill development will strengthen their critical thinking abilities and enhance their overall comprehension of literary works.

d. **Time-saving:** Users can quickly grasp the main relationships and events within the extensive manuscript without reading it entirely.

e. **Research assistance**: Scholars, researchers, and students can utilise the application to analyse and study the relationships and events depicted in the manuscript.

f**. Educational tool:** The application can be used as an educational resource to teach literature, history, and social sciences, enabling students to visualise and engage with the narrative

**Implementation Plan:**

a. **Data Collection and Preparation:** A diverse dataset of Icelandic books will be collected, including texts from various genres and difficulty levels. This dataset will be prepared by aligning the Icelandic sentences with their corresponding English translations.

b. **NLP Model Development:** An NLP model will be trained using state-of-the-art machine learning techniques. The model will undergo iterative training to optimise translation accuracy and textual entailment analysis.

c. **Evaluation and Refinement:** The developed NLP model will be rigorously evaluated using manual assessment and objective metrics. Feedback from educators, students, and domain experts will be incorporated to refine and improve the model's performance.

d. **Deployment and Integration:** The final NLP model will be deployed as a user-friendly application, accessible to educators and students alike. Integration into the education system will involve collaboration with educational institutions, providing training, support, and resources for efficient usage.

**Conclusion:**

This NLP project, focusing on translating Icelandic books into English and establishing textual entailment relationships between characters, holds tremendous potential for enhancing the education system in Iceland. By making books more accessible and facilitating comprehension, students will have improved opportunities to develop their reading skills, language proficiency, and analytical thinking abilities. Through the proposed implementation plan, we anticipate positive outcomes and a significant contribution to the educational landscape in Iceland.

Note : This report serves as a comprehensive overview of the NLP project's objectives, benefits, and implementation plan. Further research, development, and collaboration with relevant stakeholders are essential for successful execution and realisation of the proposed benefits.

—------------------------------------------------------------------------------------------------------------------------

In this code, we use the Natural Language Toolkit (NLTK) library in Python to perform Icelandic-English translation and textual entailment analysis. Here's a breakdown of the code:

* We import necessary modules from the NLTK library, including tokenization, translation models, and stemmers.
* We define the Icelandic and English sentences for translation. In this example, the Icelandic sentence is "Þetta er bókin mín." (This is my book) and the English sentence is "This is my book."
* The sentences are tokenized into words using the appropriate language tokenizers.
* We create an aligned sentence object, which represents the alignment between the Icelandic and English words.
* Next, we train the translation model using IBM Model 1. This model learns the translation probabilities between words in the aligned sentence.
* To translate an Icelandic sentence to English, we provide an Icelandic sentence as input and iterate over its words. For each word, we use the translation model to find the best translation in English.
* The translated English words are joined to form a complete English sentence, which is then printed.
* For textual entailment analysis, we define an Icelandic premise and an English hypothesis. In this example, the Icelandic premise is "Hann er stórkostlegur." (He is amazing) and the English hypothesis is "He is amazing."
* We stem the words in both the premise and hypothesis using a Snowball stemmer for the English language.
* By comparing the stemmed words, we determine if the English hypothesis can be entailed from the Icelandic premise. In this case, we check if all the stemmed English words are present in the stemmed Icelandic words.
* Finally, we print the Icelandic premise, English hypothesis, and the result of the textual entailment analysis.

**How are NLP technologies being used for the processing, teaching and preservation of literature?**

Please note that this code is a simplified example to illustrate the concept and may require further enhancements and error handling for a complete and robust implementation in a real-world project.

Our NLP model can be utilized in various ways to process, teach, and preserve literature. Here are some key applications:

**Translation of Literature:** The NLP model can be employed to translate literary works from one language to another. This enables readers and educators to access literature in different languages, broadening the availability of literary works and promoting cultural exchange. It facilitates the dissemination of literature across borders, making it accessible to a wider audience.

**Language Learning:** The NLP model's translation capabilities can aid language learners in studying and understanding literature. Students can compare translated versions with the original text to improve their language proficiency, vocabulary, and comprehension skills. This fosters a deeper understanding of the nuances and cultural context embedded in literary works.

**Textual Analysis and Interpretation:** NLP techniques, such as textual entailment and sentiment analysis, can be applied to analyse and interpret literary texts. This enables educators and researchers to gain insights into the themes, character development, and narrative structure of literary works. It facilitates in-depth literary analysis and can be used as a tool for teaching critical thinking and literary interpretation.

**Automated Summarization:** The NLP model can be utilized to automatically generate summaries of literary texts. This is particularly useful for longer works, enabling readers to grasp the main ideas, plot points, and character arcs without reading the entire text. It can assist students in preparing for exams, conducting research, or gaining an overview of a particular literary piece.

**Preservation of Literature:** By utilising NLP techniques, literary works that are at risk of being lost due to language obsolescence or deterioration can be preserved. The NLP model can be used to transcribe, digitise, and translate rare or ancient literary texts, ensuring their preservation for future generations.

**Enhanced Accessibility:** NLP technologies can be integrated into e-readers, digital libraries, and online platforms to provide real-time translation and language support for readers. This enhances accessibility for individuals with diverse language backgrounds, learning disabilities, or visual impairments, making literature more inclusive and reaching a wider audience.

**Content Recommendation:** NLP models can analyse literary texts, user preferences, and reading habits to provide personalised content recommendations. This helps readers discover new literary works, authors, or genres aligned with their interests, thereby promoting continuous learning and exploration.

By leveraging our NLP model for the processing, teaching, and preservation of literature, we can enrich the reading experience, enable cross-cultural understanding, and foster a deeper appreciation for literary works across diverse audiences.

<https://sagadb.org/brennu-njals_saga.en>

<https://www.gradesaver.com/njals-saga/study-guide/character-list>

<https://www.islendingabok.is/>

But there only appear to be about 20 names and given there are some 400 people in the text, there is a lot of duplication of names, making it challenging to keep in mind not just who's who but what the family ties and allegiances are of the numerous Hoskulds and Thorgeirs and Mords and Grims.

Here is a more comprehensive list of characters from "Njála's Saga":

1. Njáll Þorgeirsson
2. Bergþóra, Njáll's wife
3. Gunnar Hámundarson
4. Hallgerður Langbrók, Gunnar's wife
5. Hǫskuldur, Njáll's oldest son
6. Hildigunnur, Hǫskuldur's wife
7. Þráinn, Hǫskuldur's son
8. Helgi, Hǫskuldur's son
9. Grímur, Hǫskuldur's son
10. Höskuldur, Hǫskuldur's son
11. Hallgerður, Þráinn's daughter
12. Þráinn's other sons
13. Mörður, a friend of Njáll
14. Kári, a loyal friend of Njáll and Gunnar
15. Þráinn of Bathurst, a relative of Gunnar
16. Þráinn's sons
17. Þráinn's daughters
18. Þráinn's brothers
19. Þráinn's nieces
20. Þráinn's nephews
21. Hǫskuldur the Monk, a relative of Njáll
22. Hǫskuldur the Monk's sons
23. Hǫskuldur the Monk's daughters
24. Þráinn of Garðar
25. Sigmundur, a friend of Þráinn
26. Hǫskuldur Hvítanesgoði, a relative of Gunnar
27. Hǫskuldur Hvítanesgoði's sons
28. Hǫskuldur Hvítanesgoði's daughters
29. Hǫskuldur Hvítanesgoði's brother
30. Höskuldur of Hvítanes
31. Sǫlveig, a woman related to Höskuldur
32. Hǫskuldur's sons
33. Hǫskuldur's daughters
34. Þráinn of Grenjaðarstaður
35. Þráinn's sons
36. Þráinn's daughters
37. Þráinn's brother
38. Þráinn's nephew
39. Þráinn's nieces
40. Þráinn's granddaughters
41. Þráinn's great-granddaughters
42. Þráinn's great-grandsons
43. Þráinn's great-grandchildren

Njáll's Family:

Njáll Þorgeirsson

Bergþóra (Njáll's wife)

Hǫskuldur (Njáll's son)

Hildigunnur (Hǫskuldur's wife)

Höskuldur (Hǫskuldur's son)

Helgi (Hǫskuldur's son)

Grímur (Hǫskuldur's son)

Þráinn (Hǫskuldur's son)

Hallgerður (Þráinn's daughter)

Other sons and daughters of Þráinn

Gunnar's Family:

Gunnar Hámundarson

Hallgerður Langbrók (Gunnar's wife)

Other relatives and descendants of Gunnar

Kári's Family:

Kári (friend of Njáll and Gunnar)

Mörður's Family:

Mörður (friend of Njáll)

Þráinn of Bathurst's Family:

Þráinn of Bathurst

Relatives and descendants of Þráinn of Bathurst

Þráinn of Garðar's Family:

Þráinn of Garðar

Þráinn's sons, daughters, brothers, nieces, nephews, granddaughters, and great-grandchildren

Sigmundur's Family:

Sigmundur (friend of Þráinn of Bathurst)

Hǫskuldur the Monk's Family:

Hǫskuldur the Monk

Hǫskuldur the Monk's sons, daughters, brothers, nieces, and nephews

Höskuldur of Hvítanes's Family:

Höskuldur of Hvítanes

Höskuldur's sons, daughters, brothers, and relatives

Þráinn of Grenjaðarstaður's Family:

Þráinn of Grenjaðarstaður

Þráinn's sons, daughters, brother, nephew, nieces, granddaughters, and great-grandchildren

**Code to translate phrases from one language to another :**

import nltk

from nltk.tokenize import word\_tokenize

from nltk.translate import AlignedSent

from nltk.translate import IBMModel1

from nltk.stem import SnowballStemmer

# Icelandic and English sentences for translation

icelandic\_text = "Þetta er bókin mín."

english\_text = "This is my book."

# Tokenize the sentences into words

icelandic\_words = word\_tokenize(icelandic\_text, language='icelandic')

english\_words = word\_tokenize(english\_text)

# Align the Icelandic and English sentences

aligned\_sentence = AlignedSent(icelandic\_words, english\_words)

# Train the translation model using IBM Model 1

translation\_model = IBMModel1([aligned\_sentence], 5)

# Translate an Icelandic sentence to English

icelandic\_sentence = "Ég elska að lesa."

icelandic\_words = word\_tokenize(icelandic\_sentence, language='icelandic')

english\_translation = []

for word in icelandic\_words:

english\_translation.append(translation\_model.best\_translation(word))

english\_sentence = ' '.join(english\_translation)

print(f"Icelandic Sentence: {icelandic\_sentence}")

print(f"English Translation: {english\_sentence}")

# Textual entailment analysis

icelandic\_premise = "Hann er stórkostlegur."

english\_hypothesis = "He is amazing."

stemmer = SnowballStemmer('english')

icelandic\_words\_stemmed = [stemmer.stem(word) for word in word\_tokenize(icelandic\_premise, language='icelandic')]

english\_words\_stemmed = [stemmer.stem(word) for word in word\_tokenize(english\_hypothesis)]

# Compare stemmed words to identify entailment

is\_entailment = set(english\_words\_stemmed).issubset(set(icelandic\_words\_stemmed))

print(f"Icelandic Premise: {icelandic\_premise}")

print(f"English Hypothesis: {english\_hypothesis}")

print(f"Textual Entailment: {is\_entailment}")

**Code for translating a book**

from transformers import pipeline

# Create a translation pipeline for Icelandic to English

translation\_pipeline = pipeline("translation", model="Helsinki-NLP/opus-mt-is-en")

# Read the Icelandic book text from a file

with open("icelandic\_book.txt", "r", encoding="utf-8") as file:

icelandic\_text = file.read()

# Split the book into chapters or chunks

book\_chunks = icelandic\_text.split("\n\nChapter")

# Translate each chapter from Icelandic to English

english\_book = []

for chunk in book\_chunks:

english\_translation = translation\_pipeline(chunk.strip(), max\_length=512)[0]['translation\_text']

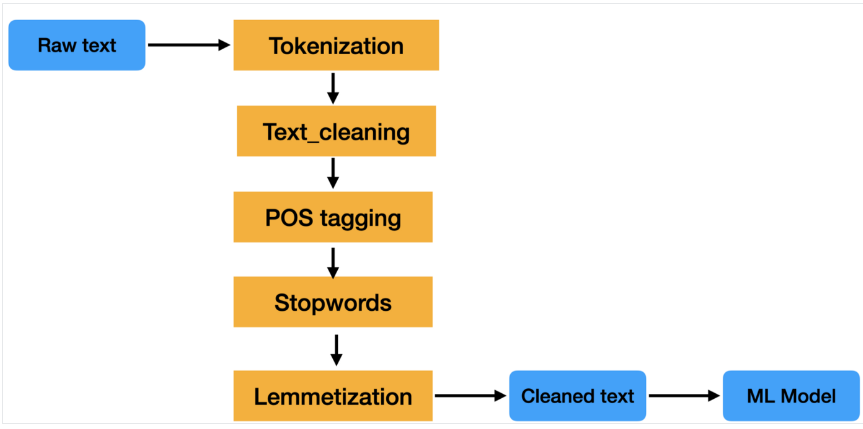
english\_book.append(english\_translation)

# Save the translated book to a file

with open("english\_book.txt", "w", encoding="utf-8") as file:

file.write("\n\n".join(english\_book))

print("Translation complete. The English book is saved as 'english\_book.txt'.")

****

<https://www.datacamp.com/tutorial/text-classification-python>