

Standard Operating Procedure (SOP) for translating educational material in bioinformatics from Portuguese to English

Version: 1.0

Date: 1 Dec 2023

Prepared by: Liã Bárbara Arruda, Wellcome Connecting Science

SUMMARY

| | |
|---|----------|
| 1 OBJECTIVE | 1 |
| 2 SCOPE OF APPLICATION | 1 |
| 3 ACRONYMS | 1 |
| 4 REFERENCES | 2 |
| 5 GENERAL PROVISIONS | 2 |
| 5.1 Initial Preparation | 2 |
| 5.2 Glossary | 2 |
| 5.3 Example of a glossary based on the content of the course "SARS-COV-2_B4B" | 2 |
| 6 PROCEDURE | 3 |
| 6.1 Translation with Chat GPT | 3 |
| 6.1.1 Small Passages or Paragraphs | 3 |
| 6.1.2 Version Comparison | 3 |
| 6.1.3 Delving Deeper with Chat GPT | 3 |
| 6.1.4 Possible Issues When Using Chat GPT | 4 |
| 6.2 Translation of Technical Terms | 4 |
| 6.3 Review and Improvement | 5 |
| 7 EXPECTED RESULTS | 5 |
| 7.1 Distribution and Use | 5 |
| 7.2 Glossary Update | 5 |

1 OBJECTIVE

To describe procedures for translating educational materials to be used in bioinformatics training.

2 SCOPE OF APPLICATION

Bioinformatics and professional development

3 ACRONYMS

| | |
|-----------------------|--|
| Chat GTP 3.5 | Artificial Intelligence Tool |
| SOP | Standard Operating Procedure |
| PT-BR | Brazilian Portuguese |
| SARS-COV-2_B4B or B4B | SARS-CoV-2 bioinformatics course for beginners |

4 REFERENCES

Online Course: Working with Translation: Theory and Practice (<https://www.futurelearn.com/courses/working-with-translation>)

5 GENERAL PROVISIONS

5.1 Initial Preparation

Familiarize yourself with the material's context. Understand the topic and key bioinformatics concepts covered in the document. Identify specific technical terms that need translation and make a list of them.

5.2 Glossary

Research technical terminology in both languages, using resources such as specialized dictionaries, bioinformatics glossaries, and related scientific articles.

Create a personalized glossary with preferred translations of technical terms. This will help maintain consistency throughout the document.

5.3 Example of a glossary based on the content of the course "SARS-COV-2_B4B"

| English | Portuguese |
|---------------------------|--------------------------------------|
| Amplification | Amplificação |
| Assembling | Montagem |
| Assignment | Atribuição |
| Cleaned | Limpo |
| Command line | Linhas de comando |
| Consensus | Consenso |
| Cover | Cobertura |
| Data set | Conjunto de dados |
| Depth | Profundidade |
| Dropout | Declínio; desistência |
| Kmer | K-mero |
| Merging | Fusão |
| NGS | Sequenciamento de Nova Geração (NGS) |
| Overlapping | Sobreposição |
| Overrepresented | Hiper-representado |
| Per base sequence quality | Qualidade da sequência por base |
| Pipelines | Conjunto de regras e instruções |

| | |
|--------------------------|--|
| Poor-quality data | Dados de baixa qualidade |
| Raw reads | Leituras brutas |
| Read length distribution | Distribuição do comprimento da leitura |
| Reads | Leituras |
| Software | Programa |
| String | Cadeia/linha de caracteres |
| Trimming | Cortar / Remover |

6 PROCEDURE

6.1 Translation with Chat GPT

Translate the document overall, focusing on understanding the original content. For this purpose, you may use tools like Chat GPT or Google Translate to assist.

6.1.1 Small Passages or Paragraphs

To start the translation process, it's advisable to divide the document into small passages or paragraphs to ease translation and comparison between the original and translated versions. Copy and paste the excerpts into Chat GPT 3.5 (<https://chat.openai.com/>) and request to translate into Portuguese.

Example command:

"Translate the following paragraph into EN-GB while keeping the technical terms in the original language."

The translation of technical terms will be explained in subsequent steps; do not concern yourself with them at this stage.

6.1.2 Version Comparison

After translating each segment, it's crucial to compare the original version with the translated one. This helps to identify potential errors that may arise due to automatic translation. During comparison, pay special attention to potential translation errors, such as ambiguities, incoherent phrases, or loss of meaning.

6.1.3 Delving Deeper with Chat GPT

To enhance comprehension and translation quality, it's possible to use artificial intelligence to delve deeper into the meaning of certain phrases or terms. Chat GPT can provide more detailed and contextual explanations to ensure the translation accurately reflects the original content.

Example command:

'Provide a detailed explanation about the following section from the Bioinformática Decodificada course, and suggest a translation:

As investigações iniciais sugeriram que o agente causador poderia ser um patógeno viral zoonótico. Foi realizada coleta de lavado broncoalveolar para a realização de rastreo molecular.'

Chat GPT's response:

The given text describes initial investigations that indicated the potential involvement of a zoonotic viral pathogen as the causative agent. A bronchoalveolar lavage was conducted to perform molecular screening.

In this context, "investigações iniciais" (initial investigations) refer to the early stages of research or inquiry into a specific issue or problem. "Agente causador" (causative agent) signifies the entity responsible for causing a particular disease or condition. "Patógeno viral zoonótico" (zoonotic viral pathogen) refers to a virus capable of being transmitted from animals to humans.

The "coleta de lavado broncoalveolar" (bronchoalveolar lavage collection) involves obtaining a sample of fluid from the lungs, specifically the bronchioles and alveoli, for further analysis or testing. "Rastreo molecular" (molecular screening) denotes the use of molecular techniques to detect, identify, or study genetic material related to the suspected pathogen.

Suggested Translation:

"The initial investigations suggested that the causative agent could be a zoonotic viral pathogen. Collection of bronchoalveolar lavage was performed for molecular screening."

This translation aims to maintain the essence and key information from the original text while ensuring clarity and coherence in English.

6.1.4 Possible Issues When Using Chat GPT

Unsolicited translation of file names, codes in R/Python/Bash, or other programming languages.

It is good programming practice to add comments next to command codes. These comments are indicated with # in Python, for example. While translating these comments is essential for understanding the command to be executed, it is not appropriate to copy the entire text of the code cell and request translation from Chat GPT. In some cases, parts of the code may be translated, causing problems in executing the command. Therefore, great care is needed in cases where parts that should be translated are mixed with codes or any elements that should not undergo translation.

Inadvertent translation of file names can also cause issues in command execution.

Attention must be paid to translating technical terms, which are often routinely used in English or another original language. For example: *internet*, *in house*. In these cases, including a glossary describing their meaning may be more beneficial than the translation itself.

6.2 Translation of Technical Terms

After the initial translation in step 6.1.1, return to the glossary and begin replacing the technical terms in the document with the appropriate translation.

If manually replacing all terms is too laborious, you can use the word search tool by opening the document in Google Docs, clicking on 'Edit', and selecting 'Find and Replace'. Choose the word to be replaced and click on 'Replace All'.

Remember that this tool must be used carefully, as some terms may have different translations depending on the context.

6.3 Review and Improvement

Conduct a comprehensive review of the translated document to ensure text fluency and coherence.

Check if technical terms are consistent throughout the document. If possible, consult external sources on the subject or experts in the field of study to review the translation and ensure technical terms are correct.

Perform a final review of the document to correct any grammar or spelling errors.

Whenever possible, ask a second person to review and confirm the translation.

7 EXPECTED RESULTS

7.1 Distribution and Use

Ensure that the translated document is widely and effectively distributed to the target audience, facilitating access to bioinformatics knowledge in the Portuguese language.

7.2 Glossary Update

Keep the technical glossary updated with new terms and translations as new concepts and discoveries emerge in the field of bioinformatics.