UNIVERSIDADE LUSÓFONA

Disentanglement

Trabalho Final de curso

1ª Entrega Intercalar

Daniel Nascimento, a22208338, LEI

Pedro Prata, a21807403, LEI

Orientador: Zuil Pirola

Co-orientador: Bruno Saraiva

Departamento de Engenharia Informática e Sistemas de Informação

Universidade Lusófona, Centro Universitário de Lisboa

25/11/2024

www.ulusofona.pt

1

Direitos de cópia

Abstract

When confronted with multiple people interacting with each other there exists a need to separate and identify multiple conversations to participate in them. As online interactions increase, the ability to do the same within complex digital environments becomes critical for ensuring clarity and effective communication and analysis of dialogue. We address this challenge by analyzing the current best practices used, attempting potential improvements and studying the use of machine learning techniques to enhance the accuracy and efficiency of conversation disentanglement.

Disentanglement; Conversation analysis; Multi-participant social interaction; Prompt engineering; Topic modeling

Index

Figure List

List of Tables

Abbreviations

NLP	Natural Language Processing

1 introduction

1.1 Context

A **multi-participant social interaction** refers to social interactions with more than 2 participants. This kind of interaction requires a listener to differentiate which stream of dialogue each statement belongs to in order to interpret and respond appropriately (Shen et al., 2006; Elsner & Charniak, 2008), in this process we are aided not only by the content being spoken but also by the body language and attention being given by the members involved to each other, these social context clues allow people to connect easier and lead to less miscommunication to occur (Daft & Lengel, 1986).

With the development of the internet, and the proliferation of the same to the public, the ability to not only read but allow the user to also post information in 1999 with Web2.0 (Jacksi & Abass., 2019) there has been an increase in text based **multi-participant social interactions**, using various services (social media, comment sections, online chat rooms). These services remove a great majority of components present in normal communication, making it easier to misunderstand the interaction(Daft & Lengel, 1986).

1.2 Motivation and Problem Identification

These multi-party interactions, due to their particular language and casual structure, allowing any participant to interact at any time with another, causes the comprehension of different dialogues by non-participants to be difficult (Li et. al., 2022) by creating a complex discourse that needs to be organized to be interpreted, this is not helped by their synchronous nature (real-time interaction) creating quite the complicated problem.

While these interactions are self regulated, possessing an internal logic to them in turn-taking, intended recipient of messages and reactionary content like *emoticons,* helping the analysis of them (Meredith & Potter., 2014), there is still a necessity to develop tools and processes to help in the process.

1.3 Objectives

Due to the difficulty found in this process the following piece of work proposes the use and improvement of the current best practices and tools of the practice of disentanglement to produce a better understanding of the interaction between multiple users in text based multi-party interactions using various methods to be discussed and evaluated.

1.4 Document Structure

This document is organized in the following manner:

- In section 2 the viability and relevance of the work in question will be explored.
- In section 3 the basic concepts necessary to be able to read the paper will be explained in an easy to comprehend manner.
- In section 4 we will describe the current methods used in Disentanglement
- In section 5 the work that will be done to develop an annotation system for Disentanglement will be explained
- In section 6 the work plan is described with the order and type of work to be completed
- In section 7 the conclusion describes what has been accomplished and how it will help with the overall goal of this project as well as future work

2 Relevance and Viability

2.1 Relevance

The use of disentanglement in chat conversations is highly pertinent, since it is a prerequisite for further analysis, such as argument mining or complex conversational network studies. This study will also bring an understanding of how high school students in Portugal have discussed critical themes, such as expressions of love, racism, and environmental problems, which will be fundamental information for future studies and possibly a contribution to the Rede de Bibliotecas Escolares.

2.2 Viability

The study is viable since it is based on already existing data provided by the Debaqi project, supported by the "Fundação para a Ciência e Tecnologia (FCT)." This corpus contains dialogues between Portuguese secondary school students, recorded over a long period of more than one year, in a table format, and readily available, thus setting up the necessary grounds for analysis.

3 Fundamental Concepts

3.1 Theoretical Concepts

- Turn - Refers to an individual contribution of a participant in a dialogue usually in the form of individual messages. Turns are accepted as basic units of interaction because each turn is said to either respond to or introduce new ideas, within a shared context.
- Conversation - Is a structured set of exchanges that successively develop a single topic or theme. Conversations usually have cohesiveness and thematic coherence; that is, each exchange will build on or relate to previous exchanges in some shared topic.
- Thread - Refers to a continuing sequence of turns that focus on a single topic. Threads come into existence when people post contributions that add to, elaborate on, or sometimes diverge from previous statements. When there are numerous topics being discussed, several threads can be running at the same time, each representing a different line of thought or discussion topic.
- Disentanglement - Refers to the process of merging conversational turns into cohesive threads based on both thematic relevance and sequential consistency.

3.2 Technology Used

The following are the different tools used during this project:

- **PYTHON CODES (PANDAS, NUMPY, spaCy):**

  Python was utilized for data handling, pre-processing, and the application of Natural Language Processing (NLP) techniques, using the libraries Pandas, NumPy, and spaCy.

- **CUSTOM ANNOTATION TOOL:**

  This project necessitates a high volume of annotations in different areas. Therefore, alongside this TFC, another TFC is being developed, focused on creating a tool tailored for disentanglement.

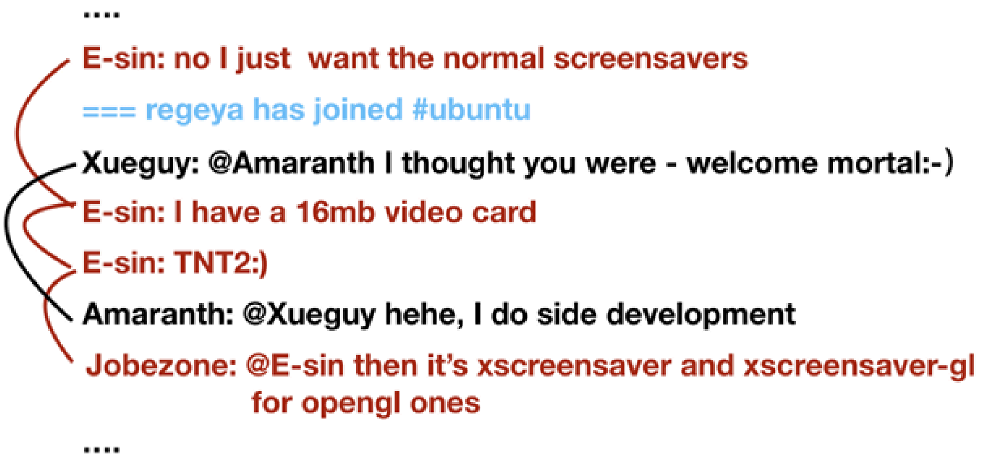
- **LARGE LANGUAGE MODELS (GPT, LHAMA):**

  Part of this project pertains to the development of prompt base models, to cluster different turns into threads. We will test both few shot as well as Chain-of-Thought to produce said prompt.

4 State of the Art

4.1 State of the Art

In many communication systems, several dialogues can be going on simultaneously without an explicit structure, which turns out to be a challenging task to distinguish between topics. To address this, it is fundamental we use a disentanglement approach, which involves partitioning the exchanged messages into different dialogues or threads. In thread detection, these interactions are structured by identifying and categorizing dialogues about specific topics, where each thread represents a distinct sequence of related turns (Shen et al., 2006; Kummerfeld et al., 2020). To help illustrate this challenge, consider the example in Figure 1 (Li, Gu, Ling, & Liu, 2022), which demonstrates a conversation where multiple threads are entangled.



....

E-sin: no I just want the normal screensavers

=== regeya has joined #ubuntu

Xueguy: @Amaranth I thought you were - welcome mortal:-)

E-sin: I have a 16mb video card

E-sin: TNT2:)

Amaranth: @Xueguy hehe, I do side development

Jobezone: @E-sin then it's xscreensaver and xscreensaver-gl for opengl ones

....

Figure 1 - Tianda Li, Jia-Chen Gu, Zhen-Hua Ling, Quan Liu, (2022), Example of dialogue disentanglement. Conversations marked with different colours are entangled together. Reprinted from Conversation- and Tree-Structure Losses for Dialogue Disentanglement by T. Li, J.-C. Gu, Z.-H. Ling, & Q. Liu, 2022, Proceedings of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering, p. 54. ©2022 Association for Computational Linguistics

Threads are turn sequences that begin with an initial turn and are followed by one or more response turns to the first turn on a given topic (Shen et al., 2006). In this context, a "turn" refers to a single exchange within a dialogue, typically corresponding to one participant's contribution in a conversation (Elsner & Charniak, 2010). Each thread thus corresponds to a specific subject of discussion, which can be complex in dynamic systems where themes frequently overlap. Identifying the focus is required in threads, but in turn streams with short turns, one statement may not contain the topic and needs more in-depth analysis (Shen et al., 2006; Kummerfeld et al., 2020).

One of the approaches to performing disentanglement is by using clustering algorithms that group the related turns into threads, each representing a single topic. The very popular Single-Pass Clustering algorithm works iteratively. In this regard, the first turn is considered a thread, and subsequent turns are matched against the existing threads. If the similarity between a turn and a thread is above a threshold, the turn is assigned to the corresponding thread; otherwise, a new thread is created for that turn (Shen et al., 2006; Traum, 2004). While simple, this method relies on parameters such as the similarity threshold, which must be carefully tuned (Shen et al., 2006, Kummerfeld et al., 2020).

Reply is a very essential concept in disentanglement, as it explicitly links turns. One turn is a reply to another when its content is a direct response to the content of the other turn; hence, the two turns are explicitly linked together. These links, along with replies, are fundamental for correctly segmenting turns into threads (Kummerfeld et al., 2020).

However, links in disentanglement go beyond just replies. They also include temporal or contextual connections that may not involve a direct response but still represent a relationship within the conversation (Kummerfeld et al., 2020). For instance, a turn may not directly reply to a previous one but still be related through the flow of the discussion, indicating a connection between topics or ideas. These broader link types help trace separate strands of conversation and show how different topics develop and interrelate over time (Kummerfeld et al., 2020). Correct identification of both replies and other kinds of links is core to the accurate segmentation of topics and the correct assignment of turns to threads (Shen et al., 2006; Elsner & Charniak, 2008).

Analysis of disentanglement in multi-party conversations requires the annotation of data, where every turn in a dataset is marked with the thread to which it belongs, taking into consideration both explicit replies and contextual or temporal links. High-quality annotation is essential as it forms the foundation for training and testing models of disentanglement. Manual annotation can be challenging due to topic overlap and the brevity or ambiguity of many turns. Thus, strong metrics are necessary to estimate the consistency and accuracy of annotations created either by human annotators or automated systems (Kummerfeld et al., 2020; Elsner & Charniak, 2008).

Several metrics are used in the evaluation of quality of annotation in disentanglement, including One-to-One Overlap and Exact Match F1. One-to-One Overlap uses a bipartite matching algorithm to calculate the similarity of conversations from different annotators, while Exact Match F1 measures precision by ensuring that extracted conversations are identical with the annotated ones, with exclusions for automated turns; see Kummerfeld et al. (2020); Elsner & Charniak (2008). These metrics are necessary to ensure the quality of the disentanglement model.

The relation of threads and topics is important because it explains how conversation is segmented. There is just one topic per thread, but different topics may be discussed in multiple threads with turns from different threads overlapping. It makes the disentanglement therefore harder since it involves finding the right turns for each of the topics while at the same time organizing these turns coherently (Shen et al., 2006; Kummerfeld et al., 2020).

5 Proposed Solution

5.1 Introduction

The objective of this work is to develop a replicable annotation system for the disentanglement task in conversations. This system will be designed to support at least four annotators working collaboratively. The annotation will be applied to chatrooms with diverse themes, and by the end, we aim to create an annotated dataset that can be utilized for future research.

Additionally, a prompt-based model will be developed for the automated generation of annotations, aiming to produce results close to the gold standard created by human annotators. After creating a sufficiently large set of annotations, different approaches, such as zero-shot, few-shot, and chain-of-thought models, will be tested.

Finally, the relationship between topic modeling and the prompt-based model will be investigated. Considering the connection between threads and topics discussed in the literature review, the proposal is to evaluate the ability of conventional topic modeling methods to group turns into threads.

5.2 Methodology

The approach to deploying the proposed solution adheres to the step-by-step workflow listed below. The operational framework is divided into three critical phases: manual annotation, automated model development, and topic analysis.

Stage 1: Manual Annotation

- Literature Review (Disentanglement)
- Development of an Annotation Manual(Appendix 1)
- Manual Changes and Training in Annotation
- Data Annotation

Stage 2: Creation of Automated Models

- Development of a Prompt-Based Model (zero-shot, few-shot, chain-of-thought)
- Performance Testing on Llama and GPT
- Evaluation metrics, for example, the F1 score,

Stage 3: Topic Models

- Application of Techniques like Hierarchical Clustering, BERT-based Cluster Models Comparison of prompt-based models with the topic modeling approaches

5.3 Data Collection

The presented work uses data obtained with the Debaqi project, which is funded by the Fundação para a Ciência e Tecnologia (FCT). The database includes dialogues gathered among high school students from Portugal. Data is presented in table and deals with a range of issues covering the fields of love, racism, environmental problems and others.

5.4 Data Description

The dataset includes the following variables:

- **Room ID**: Unique identifier for each conversation room.
- **Turns**: Individual contributions from participants, each with an associated timestamp.
- **Participants**: Unique identifiers for each interlocutor.

5.5 Data Pre-Processing

The dataset provided by the DEBAQI project is already tabulated and organized into turns. Therefore, no additional processing steps, such as removing missing values or normalizing data, will be required.

5.6 Exploratory Data Analysis

So far, no exploratory data analysis has been done. This step will be performed in the future parallel to the annotation process to discover potential patterns and meaningful features of the dataset.

5.7 Chosen Models and Algorithms

- Manual Annotation: Establishes a baseline for validation.
- Prompt-Based Model: Incorporating few-shot, and chain-of-thought approaches.
- Topic Modeling: Utilizing hierarchical clustering and BERT-based clustering methods.

5.8 Scope

Applied Curricular Units:

- Data Science: To analyze and process datasets, to understand the relationships between threads and topics, and to evaluate models.
- Artificial Intelligence: Implementation and testing of machine learning models, including prompt-based and clustering algorithms.
- Probability and Statistics: Using such metrics as inter-agreement and F1 Score to estimate the quality of annotation and the model performance.

Scientific Areas Applied:

- Computer Science
- Artificial Intelligence
- Data Science

6 Method and Planning

6.1 Initial Planning

This section describes the steps planned for the duration of this project in order:

- Bibliographic Review (disentanglement):
    - Review existing literature to provide a foundation in disentanglement for the current study
- Development of Annotation Manual(Appendix 1):
    - Annotate small sample of dialogue data to test and improve the manual
    - Test discordance between annotators to adjust the manual so it becomes clearer what differentiates different threads
    - Train the annotators to be able to use the manual and annotation system
- First submission - 01/12/2024
- Data annotation:
    - Of dialogue room AMO by trained annotators
    - Of dialogue room RAC by trained annotators
- Calculate inter agreement and development of agreement metrics to evaluate the models to be developed
- Second submission - 13/04/2014
- Occurring simultaneously:
    - Development of prompt based model:
        - Bibliographic Review(prompt engineering):
            - Review of existing literature on the topic of prompt engineering to obtain the information required for the study
        - Development of prompt few shot
        - Development of prompt chain-of-thought
        - Prompt testing using Lhama & GPT
        - Calculate F1 Score metrics to evaluate the different prompts (few shot, chain-of-thought using Llama vs GPT)
    - Development of topic modeling:
        - Bibliographic Review (topic modeling)

                - Review of existing literature on the topic of prompt engineering to obtain the foundation required for the project
            - Use of Hierarchical cluster on annotated data
            - Use of BERT based cluster model on annotated data
- Compare and evaluate prompt based versus topic based modeling
- Conclusion based of the results of the comparison between prompt and topic modeling
- Final Submission - 27/06/2024

Gantt Planning (Appendix 2)

7 Conclusion

7.1 Conclusion

This report outlines the initial stages of the project, focusing particularly on the literature review, the development of the annotation manual and the initial data annotations. A solid theoretical foundation was established with the literature review, exploring the major challenges and existing methodologies behind the disentanglement of textual interactions—an aspect that will play a crucial role in orienting methodological choices within this project.

The creation of the annotation manual was one of the first important steps, establishing the necessary guidelines for consistent data annotation. Based on this manual, the annotators received training to ensure alignment in the execution of the process. The initial data annotation was carried out according to these guidelines, allowing for the evaluation of the practical application of the established instructions.

These preliminary steps have created a very solid base for the project and consequently have allowed the orderly continuation of the current annotation phase. The next stages of the project will include checking completed annotations and applying automated models to improve efficiency and ensure scalability.

Bibliography

[ElCh10] M. Elsner & E. Charniak, Disentangling Chat, Computational Linguistics, Vol. 36, No. 3, set. 2010.

[KuGo19] J. K. Kummerfeld, S. R. Gouravajhala, J. J. Peper, V. Athreya, C. Gunasekara,J. Ganhotra, S. S. Patel, L. C. Polymenakos, & W. S. Lasecki, *A Large-Scale Corpus for Conversation Disentanglement*, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, jul. 2019

[ShYa06] D. Shen, Q. Yang, J-T Sun & Z. Chen, *Thread Detection in Dynamic Text Message Streams*, Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, aug. 2006

[DaLe86] R. L. Daft & R. H. Lengel, *Organizational information requirements, media richness and structural design*, Management Science, Vol. 32, No 5, may 1986

[LiCh22] Tianda Li, Jia-Chen Gu, Zhen-Hua Ling & Quan Liu *Conversation- and Tree-Structure Losses for Dialogue Disentanglement*. Proceedings of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering, Dublin, Ireland, may 2022

[JoMe] Joanne Meredith & Jonathan Potter, *Conversation Analysis and Electronic Interactions: Methodological, Analytic and Technical Considerations*, in Lim, H. L. (Ed.). (2013). Innovative Methods and Technologies for Electronic Discourse Analysis (Chapter 17). IGI Global.

[JaAb19] Karwan Jacksi, Shakir M. Abass, Development History Of The World Wide Web, *International Journal of Scientific & Technology Research,* Volume 8, issue 09, SEPTEMBER 2019

Appendix 1 – Annotation Manual

Introduction

As online interactions continue to expand, group communication through chats, forums, and social networks becomes increasingly complex. In chat rooms, where users may engage in multiple simultaneous conversations, distinguishing between these conversations poses a significant challenge. This issue, referred to as disentanglement - also known as "thread identification" (Khan et al. 2002), "thread detection" (Shen et al. 2006), or "thread extraction" (Adams and Martell 2008) - has been the focus of several research efforts. The purpose of this manual is to assist annotators in identifying and separating entangled text streams from online conversations to clarify which interactions belong to which thread, a task that presents challenges in various contexts.

Task Identification

The task assigned to the annotators involves reading interactions in a chat room, turn by turn, and identifying which thread each turn belongs to.

- **Turn**: consists of a set of sentences sent by the same participant to the same addressee in a chat room.
- **Chat room**: a series of turns exchanged by two or more participants.
- **Thread or conversation**: refers to a group of interconnected turns that share reply relations or(and) the same topic. In other words, a thread is a discussion where the participants are all responding and paying attention to each other. It should be clear that the turns in a conversation are connected and flow together.

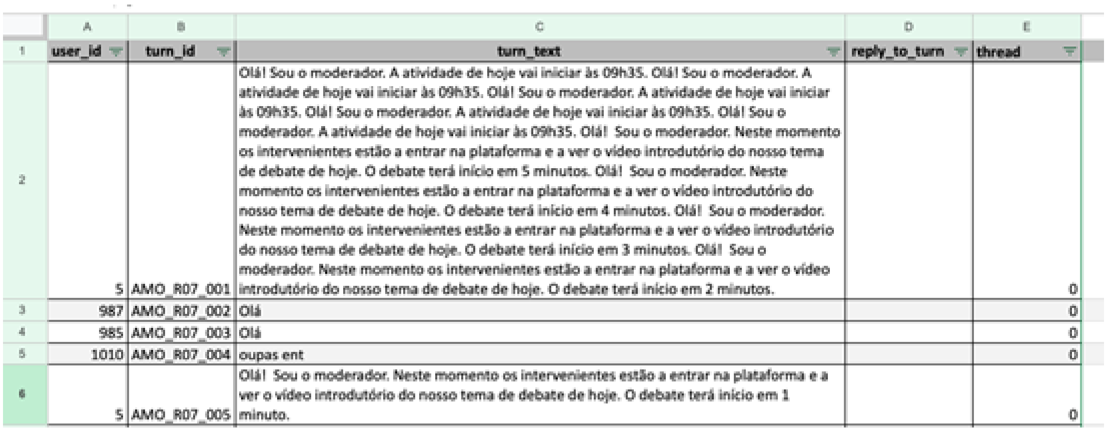
Annotation Environment and Data

As mentioned in the previous section, the data will be sent and named according to the room label. The annotators will receive an Excel file for each room. The file contains the following columns:

user_id - consists of the ID of the user who sent the turn.
turn_id - represents a unique ID for the sent turn.
turn_text - consists of a text, containing one or more sentences sent by the user.
reply_to_turn - also called reply-mark. This column indicates which turn the current turn explicitly replies to.
thread - a column that should be filled in by the annotator through the identification of the threads.

The data provided to the annotators is part of a project in which high school students were brought together in a virtual environment to discuss broad themes affecting society today, such as racism, vaccination, and others. At the beginning of each debate, students are shown a video explaining the main topic of the conversation. A moderator then interacts with the students, asking questions to guide the discussion.

Below is an example:



	A	B	C	D	E
1	user_id	turn_id	turn_text	reply_to_turn	thread
2	5	AMO_R07_001	Olá! Sou o moderador. A atividade de hoje vai iniciar às 09h35. Olá! Sou o moderador. A atividade de hoje vai iniciar às 09h35. Olá! Sou o moderador. A atividade de hoje vai iniciar às 09h35. Olá! Sou o moderador. A atividade de hoje vai iniciar às 09h35. Olá! Sou o moderador. A atividade de hoje vai iniciar às 09h35. Olá! Sou o moderador. Neste momento os intervenientes estão a entrar na plataforma e a ver o vídeo introdutório do nosso tema de debate de hoje. O debate terá início em 5 minutos. Olá! Sou o moderador. Neste momento os intervenientes estão a entrar na plataforma e a ver o vídeo introdutório do nosso tema de debate de hoje. O debate terá início em 4 minutos. Olá! Sou o moderador. Neste momento os intervenientes estão a entrar na plataforma e a ver o vídeo introdutório do nosso tema de debate de hoje. O debate terá início em 3 minutos. Olá! Sou o moderador. Neste momento os intervenientes estão a entrar na plataforma e a ver o vídeo introdutório do nosso tema de debate de hoje. O debate terá início em 2 minutos.		0
3	987	AMO_R07_002	Olá		0
4	985	AMO_R07_003	Olá		0
5	1010	AMO_R07_004	oupas ent		0
6	5	AMO_R07_005	Olá! Sou o moderador. Neste momento os intervenientes estão a entrar na plataforma e a ver o vídeo introdutório do nosso tema de debate de hoje. O debate terá início em 1 minuto.		0

Annotation Guidelines

Once the Excel file is received, the annotators should read the conversation turn by turn and assign each turn a (thread) number. Turns that are related by topic should have the same number, indicating that the conversation continued along the same thread. The result of the annotation will be an Excel spreadsheet with the 'thread' column fully completed.

Here are some annotation guidelines:

- Threads start at 0: The label assigned to threads will be numeric, starting at 0. The annotator is free to increment this number if they notice a new topic emerging in the conversation.
- Throughout the conversation, the moderator introduces questions that may either continue a previously discussed thread or open a new one. Additionally, some of the moderator's messages are only meant to encourage students to engage in conversation. When this happens, these messages should be classified as a meta-thread and grouped into a single thread from the beginning to the end of the conversation.
- Annotators should aim to be as general as possible, trying to group the maximum number of turns into a single thread. However, while students are encouraged to stick to a main topic, the conversation may branch out into subtopics. The annotator has the freedom to create a new thread if they recognize a subtopic emerging.

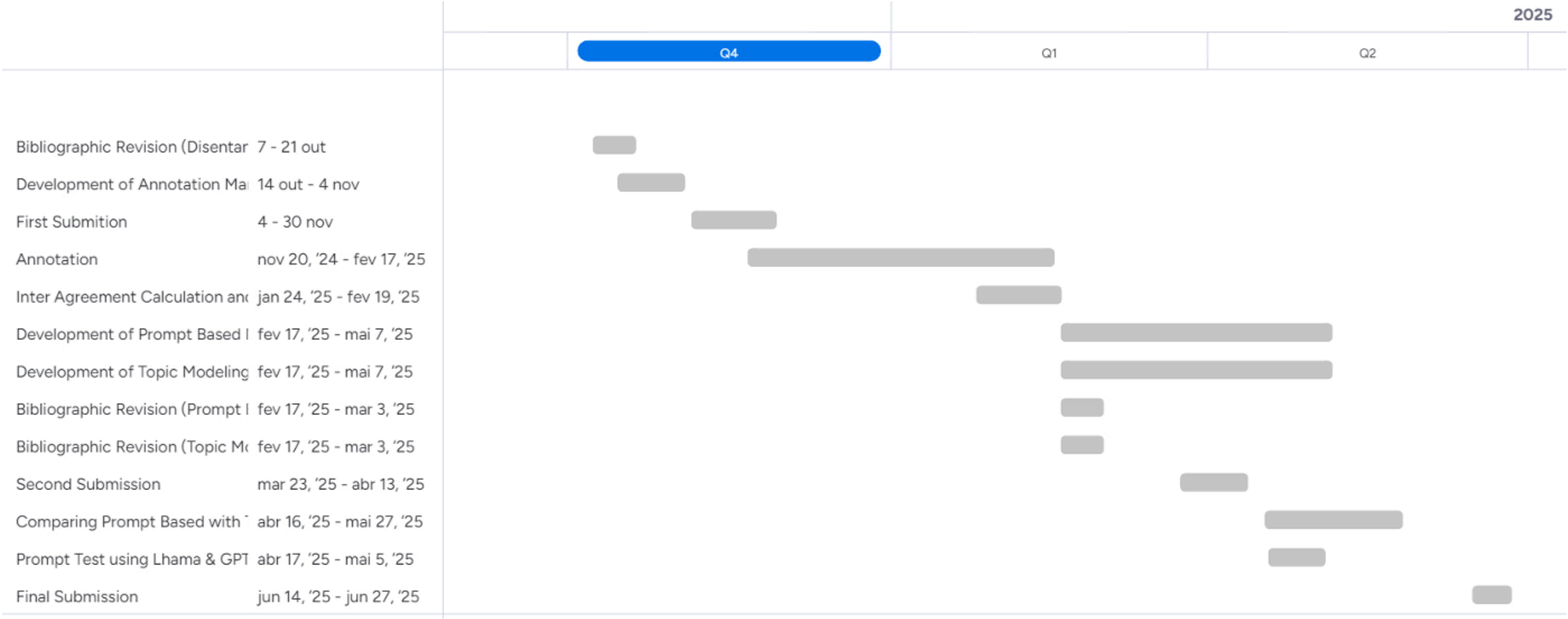
- There are some turns in which spelling and grammar errors are present. Annotators should not infer what the participants in the conversation are trying to say. In cases where the annotator cannot **understand** what is being said due to language errors or cannot identify a connection between the current turn and previous ones, they should create a new thread.
- As described in the previous section, chat rooms have a feature called *'reply_to_turn*.' Annotators should always consider this relationship between turns. Before assigning a thread number to a turn, the annotator must check whether the turn is explicitly replying to another turn and if they share the same topic or reply relation. If so, both turns should be placed in the same thread. However, there are cases where, even with an explicit reply marker, the turns do not share the same topic. In such cases, the annotator should create a new thread for the current turn being analyzed.
- The annotator should use the '*user_id'* to track messages sent in sequence by the same user. It is common for users to send a group of turns in succession, interrupted by other turns, but still referring to the same thread. Therefore, observing whether the same user has sent previous turns is an effective way to understand the thread.
- There are instances of turn flooding behavior. Some users send content like "...." or sequences of unrelated emoticons to the ongoing conversations. In such cases, whenever a turn is completely unrelated to anything previously discussed, the annotator should assign a new thread.
- Expressions like "Concordo" "sim," "Claro" or "Supostamente" should always be linked to the thread that immediately precedes the turn, unless there's a *'reply_to_turn'* event. In the case of a *'reply_to_turn',* these affirmative messages should follow the thread of the turn they are responding to.

In general, when an annotator is annotating a turn, they must follow the order of features below to assign a thread to that turn:

reply_to_turn: If there is a reply marker, check whether the current turn shares the same topic or has a reply relationship with the turn referenced by the reply marker.
user_id: Determine if the same user has sent previous turns and which threads they participated in.
turn_text: Read the content of the message and check if it relates to previous threads.

If, after analyzing the three features above, it is not possible to connect a turn to previous ones, the annotator is free to create a new thread.

Appendix 2 - Gantt Planning



Bibliographic Revision (Disentar	7 - 21 out
Development of Annotation Ma	14 out - 4 nov
First Submition	4 - 30 nov
Annotation	nov 20, '24 - fev 17, '25
Inter Agreement Calculation an	jan 24, '25 - fev 19, '25
Development of Prompt Based I	fev 17, '25 - mai 7, '25
Development of Topic Modeling	fev 17, '25 - mai 7, '25
Bibliographic Revision (Prompt I	fev 17, '25 - mar 3, '25
Bibliographic Revision (Topic M	fev 17, '25 - mar 3, '25
Second Submission	mar 23, '25 - abr 13, '25
Comparing Prompt Based with	abr 16, '25 - mai 27, '25
Prompt Test using Lhama & GPT	abr 17, '25 - mai 5, '25
Final Submission	jun 14, '25 - jun 27, '25