UNIVERSIDADE
LUSÓFONA

Qi-Brain: An intelligent moderator for educational group debates

Final Project

Student: Nayara Yasmym Rodrigues

Professor: Manuel Marques Pita

Final Project | LEI or LIG | Data: 23/07/2021

www.ulusofona.pt

Direitos de cópia

Resumo

Em uma era quando falsas notícias, discurso de ódio, polarização e abundância de informação estão mudando o panorama da civilidade social e a participação democrática, muitos políticos estão propondo censura centralizada, policiamento e controle. Pesquisas científica tem mostrado que tal controle é quase impossível-e indesejado- em um sistema democrático. Por outro lado, cientistas e educadores, propõem o empoderamento dos indivíduos para criar mais decisões informadas sobre os conteúdos que são consumidos online, e sobre as conversas que esses conteúdos podem gerar. Tal empoderamento é desafiado pelo rápido ritmo de crescimento da comunicação online, que implicitamente força as pessoas consumidoras do conteúdo online para decidir sua postura e resposta-gostar, reformar, comentar. Um dosresultados emergentes deste complexo sistema social é a perda de uma conversação saudável, deliberada, heterogênea e respeitosa. A conversa é substituída por reações rápidas, frequentes ideias "pro" ou "contra", criando polarização, e quebrando valores que mantém sociedades heterogênea, coesas, funcionais e vinculadas por valores similares, como a promoção da verdade. O projeto proposto é um módulo da continuação de um projeto educacional do Ministério da Educação Português, que possui o codinome "Debaqi". Nós buscamos produzir um chatbot inteligente que possa automaticamente moderar e encorajar debates saudáveis na plataforma. Uma das principais razões pelas quais esse moderador é importante é que Debaqi é planejado para ser implantado em todo o país nos próximos dois anos, significando que seria impraticável e caro contar com moderadores humanos em tal escala. O sucesso deste projeto depende do domínio de uma série de técnicas em Processamento de Linguagem Natural, por um lado, e as técnicas em Inteligência Artificial parasistemas de recomendações adaptáveis e flexíveis, por outro. Nós chamamos este novo módulo de Qi-Brain e esperamos implementar um protótipo de trabalho nos próximos meses. Este é um projeto altamente inovador, porque busca usar um conjunto de indicadores obtidos apartir de texto para tomar decisões inteligentes sobre intervenções de moderação que pretendem ser altamente contextualizadas e fazer com que os alunos sintam que o moderador esteja em sintonia com o debate específico em que está operando.

Abstract

In an era when fake news, hate speech, polarisation and infodemics are changing the landscape of social civility and democratic participation, many politicians are proposing centralised censorship, policing and controls. Scientific research has shown that such control is almost impossible—and undesirable—in a democratic system. On the other hand, scientists and educators propose the empowerment of individuals to make more informed decisions aboutthe contents they consume online, and about the conversations such content can generate. Such empowerment is challenged by the increasingly fast pace of online communications, which implicitly forces people that are consuming contents online to decide its stance and response—to like, re-share, comment. One of the emergent outcomes of this complex social system is the loss of healthy, deliberate, heterogeneous and respectful conversation. Conversation is replaced by fast reactions, often “pro” or “against” some idea, creating polarisation, and breaking the values that keep heterogeneous societies cohesive, functional and bound by similar values, such as fostering truth. The proposed project is a module inside anongoing educational project at the Portuguese Ministry of Education, the codename of which is Debaqi. We seek to produce an intelligent chatbot that can automatically moderate and encourage healthy debates in the platform. One of the main reasons such moderator is important is that Debaqi is planned to be deployed nationwide over the next two years, which means that it would be impractical, and expensive, to rely on human moderators at such scale. The success of this project depends on achieving the mastery of a number of techniques in Natural Language Processing, on the one hand, and of techniques in Artificial Intelligence for adaptive and flexible recommendation systems on the other. We call this new module **Qi-Brain** and expect to implement a working prototype in the next months. This is a highly innovative project because it seeks to use a range of indicators obtained from text to make intelligent decisions about moderation interventions that aim to be highly contextualised and make students feel the moderator is in tune with the specific debate in which it is operating.

Index

List of Figures

1 Problem identification

As a significant proportion of human communication increasingly takes place online. There are a growing number of challenges communication technologies must face to enable quality interactions in an increasingly hyperconnected information ecosystem. While technology is enabler of communication, it can also disable it. For example, chat rooms for groups are often crowded with messages that fill the screen up before they can be read. Another example are social feeds, which are mostly controlled by the so called "relevance" algorithms determining what conversations and contents people are exposed to.

This work is concerned with educational activities organized by schools and school libraries, in which groups of students engage in online debates on pre-specified topics. These debates happen, at present, using a web-based, chat-room platform (debaqi.pt). This project is supported by a government funded research project on educational social conversation online, the title of which is "Factors to promote healthy dialogue and behaviors in online school communities" DSAIPA/DS/0102/2019.

Students from up to three schools enter a single debate room. Participation is completely anonymous since people use usernames that are generated by the system. Debate rooms allow people to post their messages with a maximum length of 280 characters. Other people can click reactions on these messages and post their own messages. Debates are moderated currently by a member of the research team.

Debate dynamics are implemented by the moderator in a standardized manner for all debates. On a first part of a debate students are encouraged to share all their opinions and ideas; on a second stage students are encouraged to respond to others, argue, criticize and integrate information; on a final stage the moderator asks students to try to either reach a consensual answer to the debate topic, or that where opinions diverge, make a summary of these divergent positions before closing the debate.

Such debate activities are recent in the Portuguese educational system and are currently implemented as pilot studies that are part of a research project financed by the Foundation of Science and Technology of Portugal, whose principal investigator is this TFC's supervisor, Prof. Manuel Pita. One of the key problems identified in the early pilot projects is concerned with the type of interface used for group conversation. In this interface messages are posted in sequence. As they fill the screen up, previous messages go out of sight, requiring students to school up to read something they may have missed. As these debates generally work with about twenty students on average, the text on the chat screen can last a very short time as new messages keep coming in. It is a well-known fact that affordances in communication technology enable some aspects of meaningful interactions at the expense of hindering others[1].

The goal of this TFC is to use Artificial Intelligence and Data Science methods to compute, in real time, numerical indicators and other structures to synthesis what has been said in some debate on a few descriptive dimensions. (1) A semantic dimension that discretizes what has been said leveraging recent algorithms on key-phrase extraction and topic modelling [2] [3]; (2)indicators alignment between utterances and the debate topic; (3) indicators of how well participants are taking what others say into account; (4) intervention complexity. One of the

reasons screens fill up so fast is perhaps many people are posting very short interventions, when they could be using the reactions on previous messages instead to e.g., agree or disagree; (5) Finally, for this TFC, a last dimension is measuring the sentiment in the interventions.

In this report we have developed the scientific proposal that was described in the first report. This involved a significant amount of time spent reading scientific literature, familiarizing myself with the tools of Natural Language Processing on which my work is based, testing a few alternative designs, and settling for the mathematics and pipelines presented in section 4, tasks 1, 2 and 3. These initial tasks define the structure of the project with focus in the mathematical approach, grammatical structure and the main theories that this project is based.

The development of this project had important stages that conquered the expected results. After understanding the necessary scientific bases and define the principal methods to apply, the implementation phase began. The study of the most advanced methods was applied in base code to assemble the necessary structure for creating the desired chatbot. When establishing a relationship between defined objectives and the results acquired, the accomplishment of remarkable conclusions was distinguished. The desired research proposed based on the previously defined requirements was completed in a favorable manner. During the process some methods took place in advantage from others, for having a higher degree of precision and being a state-of-art methods.

1.1 Brief Literature Review

The use of Artificial Intelligence in educational systems has a long tradition. Indeed, the use of Natural Language Processing and AI for educational applications began in the 1960s with a famous program called Eliza [4]. Eliza was perhaps the best chatbot in history, but it did not have any algorithms to try and make sense of what humans were saying, it used regular grammatical patterns from the English language to give answers that gave the illusion that there was an ongoing conversation. These patterns were put into a SCRIPT that would direct Eliza's conversational line. The original goal of Eliza's creator Joseph Weizenbaum was to show that conversations between humans and machines would bore the human side because of being too superficial. Yet, to Weizenbaum's surprise, many humans that interacted with Eliza reported getting a sense of meaning, true interaction and empathy. From these surprising Elizaresults, evolved other systems, many of which aimed at mental health and educational applications of AI. One such system was SCHOLAR developed by Jaime Carbonell [5]. SCHOLAR was a question-answer chatbot about South American geography, which gave students immediate feedback about the quality of their answers. This is the first example of what is known today as Intelligent Tutoring System (ITS) [6]. Another contribution to the new field of ITS was the development of an expert system called MYCIN [7] that was designed to support medical doctors in diagnosing illnesses given information about symptoms and tests. The expert systems approach, by which domain knowledge is encoded in the tutoring system led toa very productive field in the last forty years.

The evolution of ITS has led to some standard models and paradigms that are still used today. The main idea is that there is an input that comes from the student --- which may or may not have been triggered by the intelligent tutor --- which is redescribed by the system into a representation that can be matched against domain knowledge existing in the ITS, following the tradition of expert systems that began with MYCIN. More recent ITS designs go beyond simple direct

matching by implementing sophisticated automated diagnostic reasoning [8].

One of the known limitations of the purely expert-system-based tutors was their inflexibility, and that it could take significant human effort to encode all relevant domain knowledge the back end of the system. Another important problem of relevance in this context is that, as learning domains become more complex, so do the types of interaction a learner can have with the system to, for example, ask questions, give answers, justify, argue, and so on. Forthis latter problem recent advances in Natural Language processing on the one hand, and gaming technology on the other, have produced methods and theory that have enriched the potential for nuanced interactions between learners and intelligent tutors vastly.

The educational problem within which this project concerns what makes the online conversation about a topic, engaging groups of people, “healthy”. This is a difficult domain for an intelligent tutor. Consider the difficulties posed by human communication: lack of precision in language, understanding what is really meant, the use of analogy and metaphor, intercultural differences and so on. However, the emergence of large-scale phenomena, such as the dissemination of fake news, hate speech cyberbullying, and others have made educatorsaware that this is a problem that needs to be tackled early, by the school [9].

2 Viability and Relevance

The proposed project is viable, since the foundational web platform that enables running debates (Debaqi) is deployed at the Ministry of Education of Portugal. This platform is the result of a master's thesis developed at COPELABS, Universidade Lusófona. Its architecture was conceived to include a "brain" in its back end that would process utterances produced by students before making decisions, at the server level, about what to send back to the clients in their chat-room front-end. The viability of brain implementations in this context could be compromised if too much processing is done on the utterances, slowing down what should be a naturally flowing sequence of student interactions in real time. For this reason, every intelligent module added to Debaqi must consider the necessity of code to be lightweight, optimized and bootstrapped by heuristics as much as possible. In future developments, the engineers and scientists working on the project will consider leveraging the powerful processors in mobile devices and scripting at the browser level. This is not done at this point to comply with very strict rules on data protection, privacy and anonymity.

The Debaqi platform currently depends on human moderators. When Debaqi is deployed on a larger scale, it will be impractical and very expensive to have an army of moderators ensuring that the debates run smoothy, that people are being motivated and that any sabotaging and/or disrespect are handled automatically, while preserving the ethical principles that must apply in this context. For this reason, the project is not only relevant, but also critical for the future success of the platform when scaled to operate across the entire country.

The intelligent moderator proposed here will be implemented using open-source technologies only. For the most part the implementation will be Python 3, with possibly some JavaScript code. In every case, we will leverage existing libraries for Natural Language Processing, and deployment of chatbots. The project will be innovative because of the specific indicators that will be computed and used to inform the interventions of the moderator.

The computational functionalities implemented in the system will be tested by the scientific team running the research project, while the educational functionalities will be tested by a team of qualified teachers collaborating in the project.

We obtained the **results that were originally planned** for Tasks 1, 2 and 3 successfully. Tasks 1 and 3 took about 40% (each) of the time I spend working on the Qi-Brain project (understanding the base theories and doing the implementation of TFiDF and Cosine Similarity in python) while Task 2 required about 20% (Reading and doing the implementation of NER and Bag of Words approach in python using the open-source library Spacy).

The final tasks of the project were around the fulfillment of the Coherence method and the application in distinguished datasets from online debates, took about 40% (Coding and fixing Part-of-speech tag imprecision). For understand this task was necessary diving in complex scientific papers and reproduce the method to interpretate the full concept. This task was considered the most important and the more complex realization, duo to the fact that the coherence approach has its own abstractions.

The proposed solution provides strong results in the development of a method to be applied in a chatbot that provide an organizational structure for a debate or conversation between customer and company support. The application required for the technical operation of a chatbot with a higher level of precision in a dialog was conceive and delivered in the correct term.

3. Requirements:

R1. The chatbot must be able to get dialogue utterances produced by clients before they are stored on the server-side database. The present requirement is a feature provided by the system Debaqi.

R2. The chatbot needs to have a pre-trained language model representing the expected space of prototypical sub-topics and named entities corresponding to a specific debate topic. It was done successfully by two types of representation were explored at prototype stage: TF-IDF and entity grids.

R3. The chatbot needs to have a representation of basic questions and answers students may ask, about the debate dynamics, and about the debate topic itself. It was partially done, was developed the first version of the language model that was not trained further.

R4. The real-time processing of utterances must be very fast and frugal because since:
Done based in the use of TF-IDF and entity grids both of which are computed in reasonable time.

- **R4.1** There is a potentially large number of utterances produced in short time spans.
- **R4.2** The chatbot moderation interventions must remain timely and relevant.

R5. The chatbot will only use group behavior and sequence indicators to inform its public interventions. This requirement is defined as a feature provided by the system Debaqi.

- **R5.1** The chatbot will generate anonymous usernames to offer an undisclosed environment to encourage the unidentified participants.

R6. Chatbot speech-acts must be simple and intelligible to an audience of young people. The structure is defined to accept data that is needed. Defined as speech act plans developed at prototype stage for question answering.

R7. The chatbot will reject utterances that have less than X words and send an alert to the producing user, asking them to consider a more substantial contribution. Requirement provided by the system Debaqi.

R8. The chatbot will compute alignment indicators for every utterance larger than X words (*cite Marques-Pita in preparation for specific info and math*). Requirement done by the computation of cosine similarity between vector representations of any part of utterances.

- **R8.1 –** The base of semantic dimension that simplifies the context, will be based on recent algorithms on key-phrase extraction and topic modelling. During the process the focus changed and the structure ended up using a system that relies on the identification of nouns and dependency parsing in utterances to use the entity coherence. This requirement changed as to TFiDF and entity grids leveraged the computation of PoS and dependencies using Spacy.
- **R8.2 –** Indicators adjustment between expressions from the participants and thedebate topic. This requirement achieved the prototype stage based in the combination.

R9. The chatbot will compute coherence indicators for every utterance using a past history window of size Y (*cite Marques-Pita in preparation for specific info and math*).

- **R9.1 –** Moderation must assurance civility between participants. Present requirement was not developed for not focus on the semantic side of the project.
- **R9.2 –** Swear Words must be regulated and controlled by the system. Requirement classified as feature provided by the system.

- **R9.3 –** Moderation must instigate topic explanation during the first stages of the discussion. Requirement defined as prototype proof that this can be done in production.
- **R9.4 –** The chatbot must stimulate the composition of a summary of the content in the final of the debate. The goal is that the group develop by themselves a main idea of what they have reached from the debate. Based in the prototype has proof that this can be done in production.
- **R.5 –** Moderation must encourage the students to keep the focus in the debate,by treat possible deviation from the topic.

R10. The chatbot will only produce the following kinds of interventions:

- **R10.1** Replies to pre-determined questions about the debate topic, defined in structured speech act triggers in tree structures.
- **R10.2** Replies to pre-determined questions about the debate activity, defined in structured speech act triggers in tree structures.
- **R10.3** Interventions triggered by significant deviations on debate topic, as determined by the computation of alignment (R8).
- **R10.4** Interventions triggered by significant deviations of debate coherence, as determined by the computation of coherence (R9).

Based in the results of the prototype it is possible to do in the production.

3.1 Architecture Model:

3.1.1 ERD - Entity Relationship Diagram:

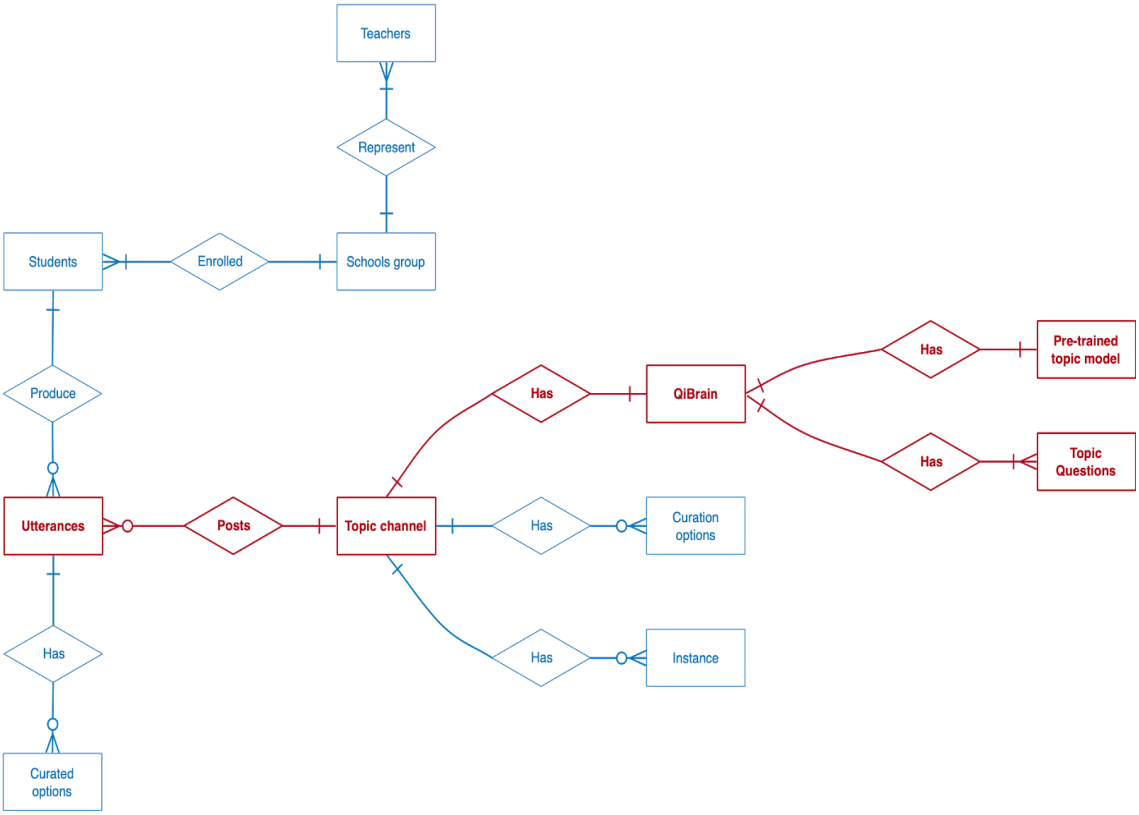
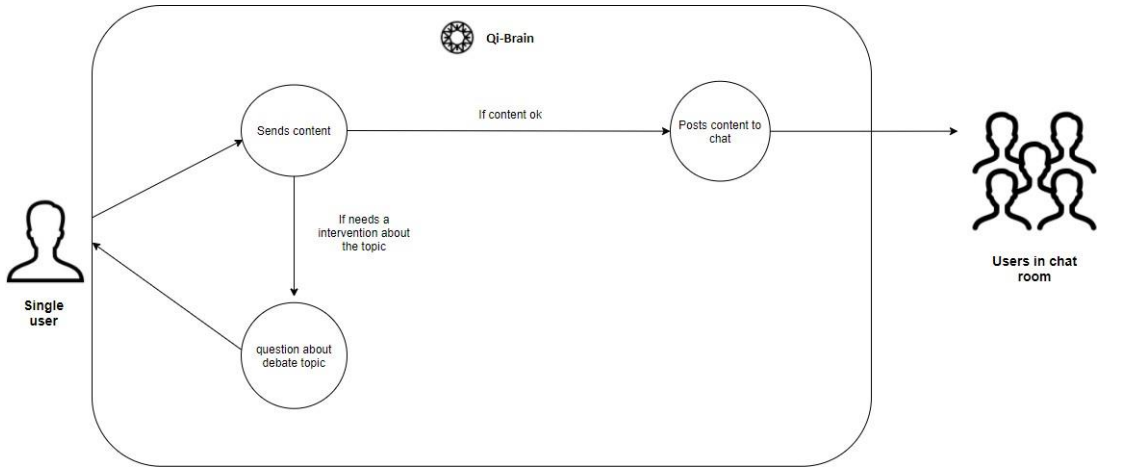
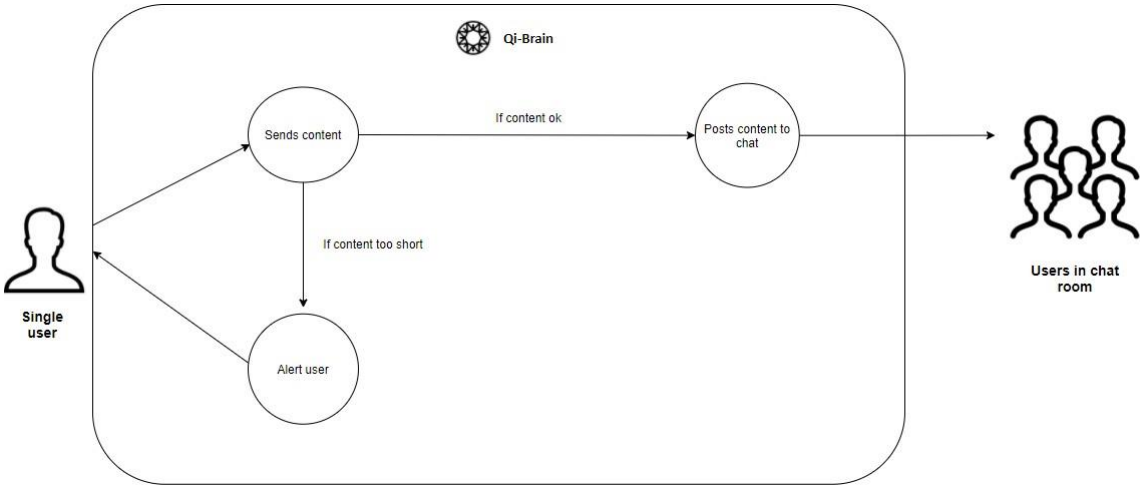


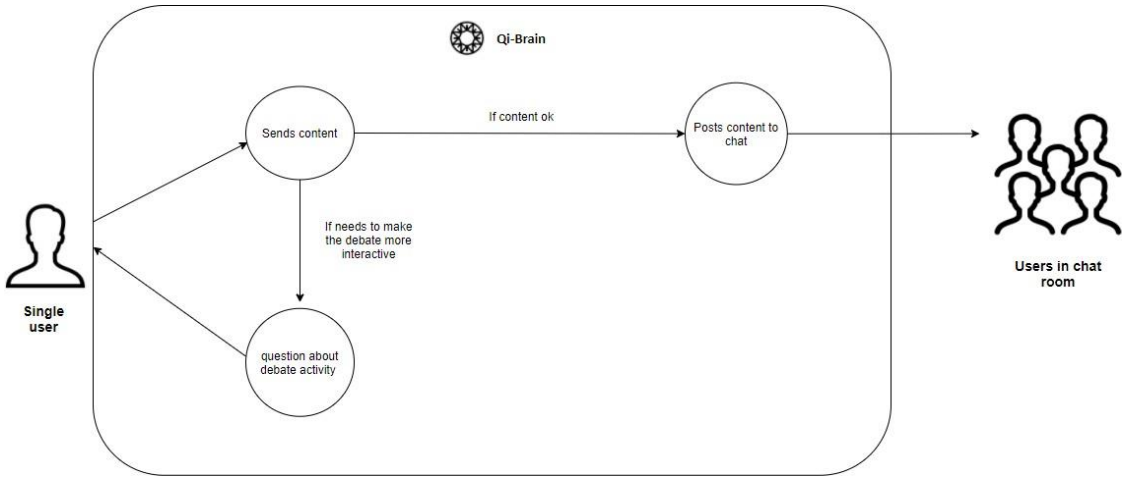
Figura 01. Red entities are the ones relevant to Qi-brain.

3.1.2 Use Case:

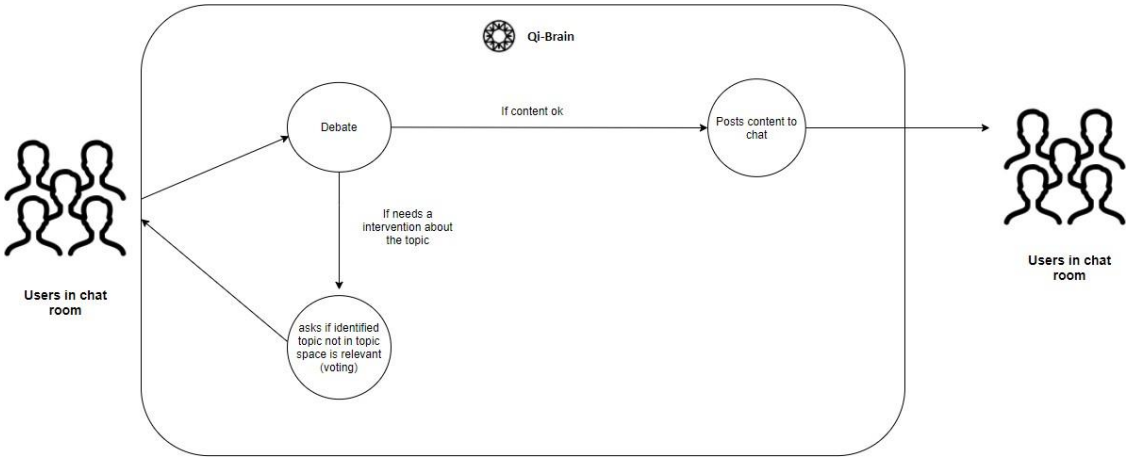
3.1.1.1 Single user:

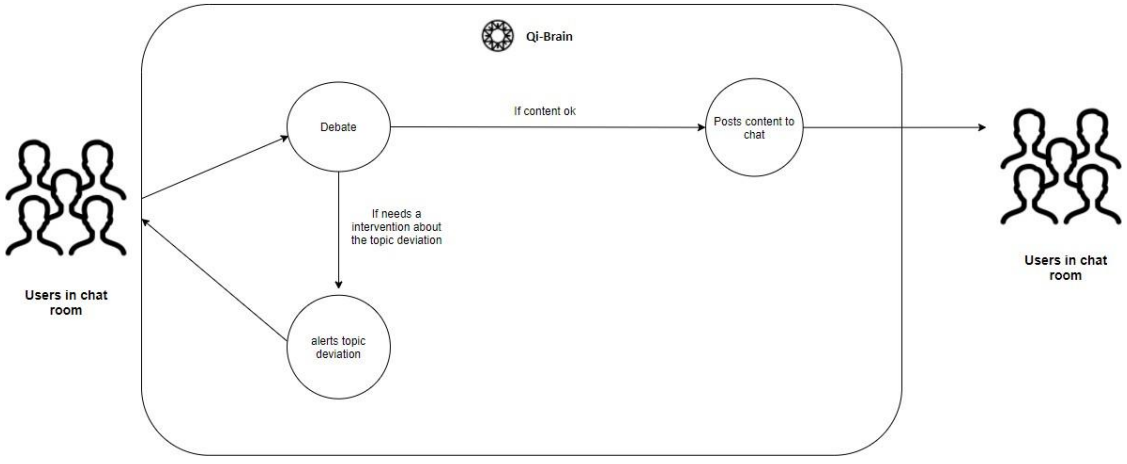


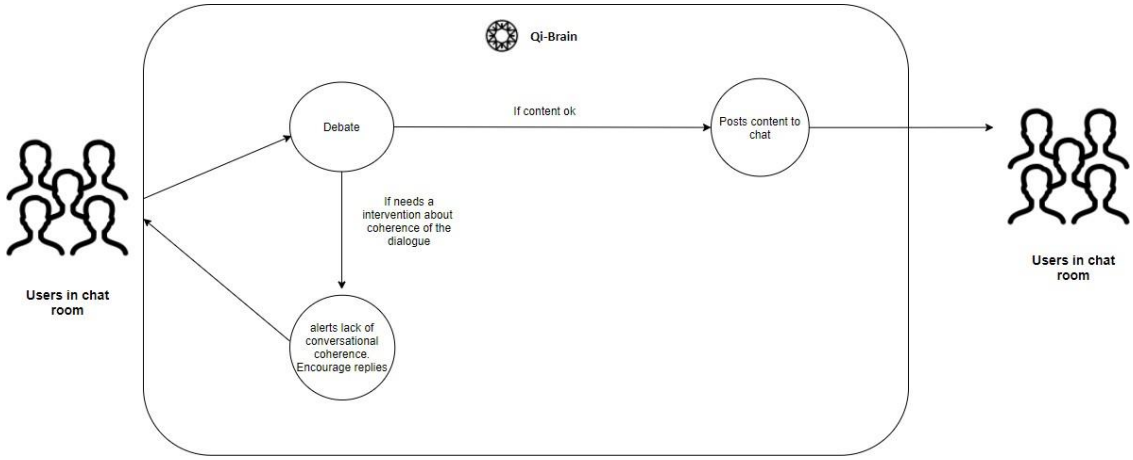


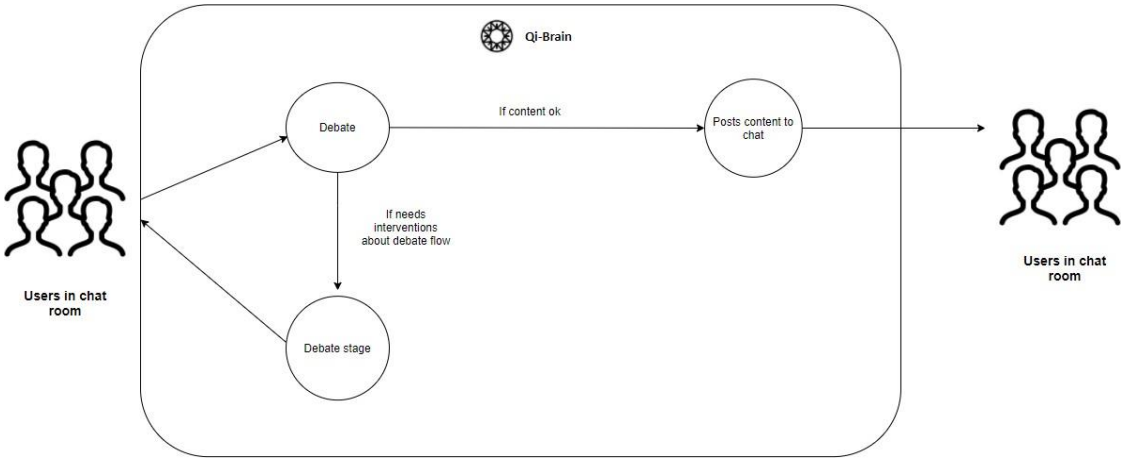


3.1.2.1 Group users:






Qi-Brain
Debate
If content ok
Posts content to chat
If needs a intervention about coherence of the dialogue
alerts lack of conversational coherence. Encourage replies
Users in chat room
Users in chat room


Qi-Brain
Debate
If content ok
Posts content to chat
If needs interventions about debate flow
Debate stage
Users in chat room
Users in chat room

4 Proposed Solution

The Debaqi client-server chat platform is implemented using an Angular front-end layer with Spring Boot for authentication and other functions. Figure 1 shows the architecture of the system developed by Prof. Pedro Perdigão for his MSc Thesis at Universidade Lusófona. The front-end layer interacts with a PostgreSQL, where information on debates is stored. That interaction is mediated by a Python Flask layer in which the Intelligent "brain" of this system is being developed. Thus, every interaction passes through this "brain" module in the back end. At this point there is no intelligence in the brain. All it does right now is to replace Portuguese swear words with **** to avoid students from using them in educational debate activities.

The platform is being used by the Ministry of Education in a sequence of pilot studies. Each study involves about sixteen different student groups, of average size 16 students. Each debateis currently moderated by a member of the scientific team. One of the most pressing features the Ministry of Education wants to see implemented in the platform is the ability to run debates autonomously. To achieve this goal there are several critical educational considerations that need to be considered:

1. Moderation must guarantee civility between participants.
2. Swear Words must be regulated and controlled by the system.
3. Moderation must encourage topic expansion during the first stages, e.g., by asking relevant and contextualized questions.
4. Moderation must encourage synthesis towards the end of the debate. The main idea is that the group self-organizes and describe the consensus they have reached on the debate topic.
5. Moderation must deal with departures from the main topic by encouraging students to stay focused.
6. Moderation must encourage quality interventions, rather than short emotionally loaded reactions.

Thus, the central goal of this project is to implement a **prototype intelligent recommender system**, **as a chatbot, that can operate as a dynamic and autonomous moderator inside the Debaqi platform**. The proposed implementation will be part of Debaqi's brain seeking to implement and test the following functionalities:

O1. Text discretization as proxy to semantics **as a vector**, using key-phrase extraction and topic modelling.

O2. Computation of **the alignment between utterances and** a reference vector representing **the topic of conversation.**

O3. Computation of the **coherence between an utterance at time *t*, and previous** utterances in a short **history** past window

O4. Computation of quantitative **indicators for utterance complexity**. This must not be based on the utterance length only but take into account language used and conversational context.

O5. Computation of indicators **of sentiment valence**, focusing mostly on "pro" and "against" attitudes that can be identified from language.

O6. Study of several situations from real data that **inform** different **automatic speech acts.** that can be produced by an automatic moderator to keep conversations going.

O7. Implementation of a basic **domain reasoner for an automatic moderator** as a recommender system

The focus of this project resides in items 01 and 04, where the study and application of the Coherence algorithm was the focus seeking to carry out the approximation for semantic vectors and computation of quantitative indicators for utterance complexity.

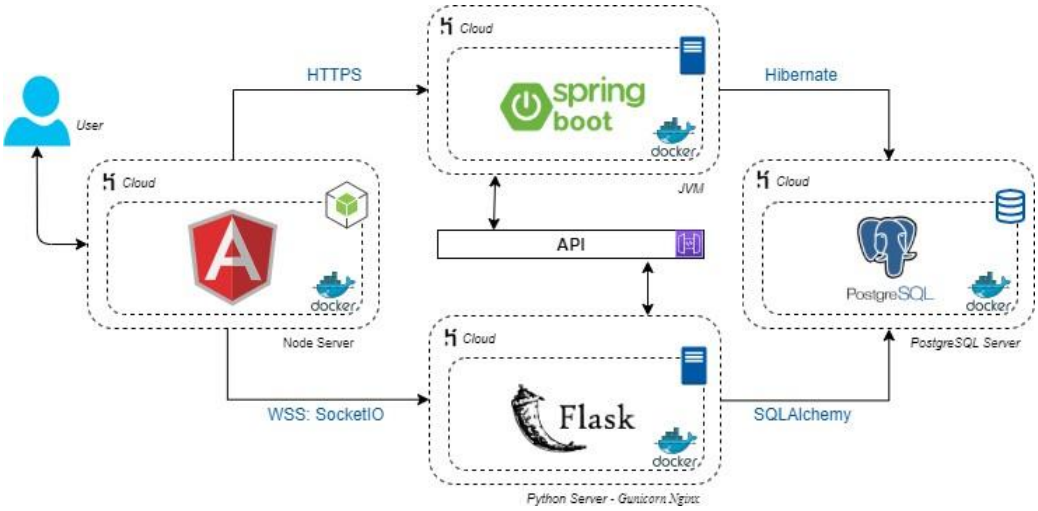


Figure 1. Architecture of the Debaqi platform.

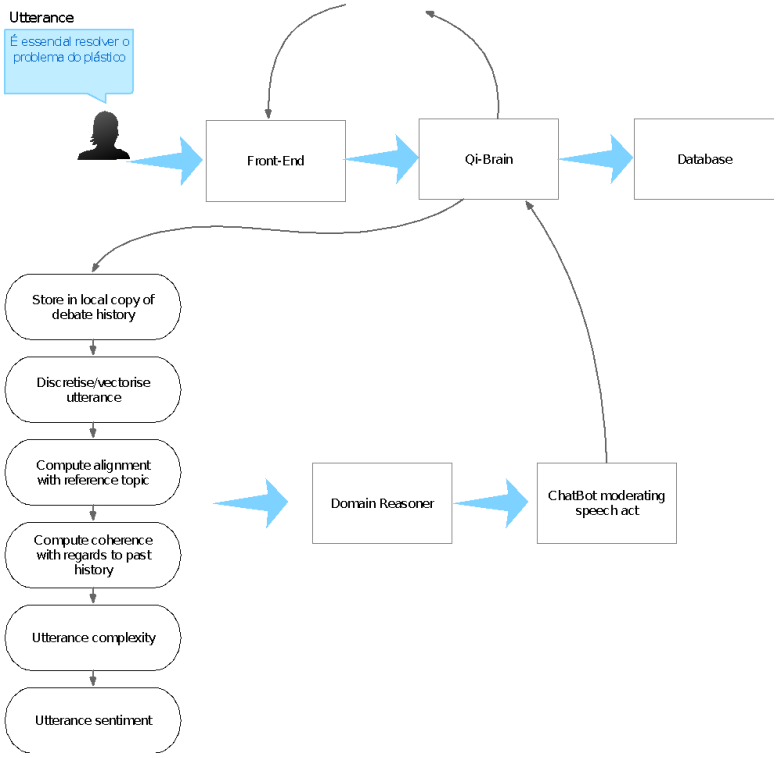


Figure 2. Proposed general architecture of Qi-Brain

Task 1. Question answering: Utterances as Vectors and Cosine Similarity.

The main goal of this task is to implement a lightweight set of algorithms that take as input questions asked by students to the automated moderator. The moderator will be able to match different linguistic variants of the questions asked and match them against a collection of predefined answers. The moderator in this TFC aims to answer questions about topics such as debate duration, debate topic, interaction rules, and so on.

In this task Qi-Brain relies on the so-called Bag of Words (BoW) approach used for information retrieval and text mining in a scope of utterances as vectors. This framework is characterised by the fact that source language data, which comes in the form of raw text, and is represented as a set of component words, disregarding grammar relationships, word order. That is, only word frequencies matter in this the BoW framework. The text in this case can be a sentence or document.

The BoW framework relies on the existence of a corpus. A corpus is simply a collection of text documents. BoW uses three essential measures computed in a specific corpus: (1) Term Frequency (TF); (2) Inverse Document Frequency (IDF) and (3) the combination of these: Term Frequency / Inverse Document Frequency (TFiDF). Term frequency refers to a word, *w*, in a document, *d*. It is the count of *w* in *d* — often normalised by the total number of words in *d*. Using TF as a measure o Although this is not usually the case in practice.

For find a better solution it is possible to apply the TFiDF weighting where to present a document as a vector and measure the degree of similarity and correlation between their corresponding vectors, which can be further quantified as the cosine of the angle between the two vectors.(Huang 2008)[10] The TFiDF formula is combined of two essential parts, first of all the TermFrequency(TF) that gives the frequency of the word in each document in the corpus, it is the ration of number of times the word appears in a document compared to the total number of words in that document. Second the Inverse Data Frequency(iDF) fig.3 used to calculate the weight of rare words across all documents in the corpus.

$$tfidf(d,t) = tf(d,t)\ x \log\left(\frac{|D|}{df(t)}\right)$$

Fig 1. TFiDF formula

$$\boldsymbol{tf_{i,j}} = \frac{\boldsymbol{n_{i,j}}}{\sum_{\boldsymbol{k}} \boldsymbol{n_{i,j}}}$$

Fig. 2 TF formula

$$\boldsymbol{idf(w) = log\left(\frac{N}{df_t}\right)}$$

Fig. 3 iDF formula

The TFiDF associated with the Cosine Similarity, define the correlation between the vectors, where each dimension represents a term with its weight in the document, which is non-negative bounded between [0,1]. The Cosine Similarity is independence of document length and documents with the same composition, but different totals will be treated identically. In fig.4 define the Cosine Similarity formula where the ta and tb are m-dimensional vectors over the term set T={t1,...tm}.

$$SIM_c\left(\overrightarrow{t_a}, \overrightarrow{t_b}\right) = \frac{\overrightarrow{t_a}.\overrightarrow{t_b}}{|t_a|\ x\ \overrightarrow{|t_b|}}$$

Fig. 4 Cosine Similarity formula

Bag of Words is a powerful tool to determine in real time the **intension** and the **topic of the conversation** that occurs in a supervised environment as a chatbot.

Task 2. Moderation I: Utterance complexity.

There are many approaches in the literature to measure the complexity of an input text. In Qi-Brain we combined two aspects of sentence complexity in a single measure. First, simply the number of words in the utterance. Utterances with very small numbers of words are not likely to contain claims and evidence to support those claims for example. On the other hand, we also consider the grammatical diversity of an utterance. The more word types are present in an utterance, the more complex we assume it to be.

We use the symbol NT to denote the number of tokens (words) in some utterance u, and VT to denote the number of word types (which we obtain from the Spacy PoS tagger). Our measurement of utterance complexity is given by the formula,

$$C(u) = \begin{cases} TTR(u), & if\ NT \geq \tau \\ \frac{TTR(u)}{\tau}, & otherwise \end{cases}$$

Where TTR(u) is given by,

$$TTR(u) = \frac{VT}{NT}$$

and **tau** (T τ) is a parameter fixed by the analyst. We set this parameter to the mean length of utterances in a given dataset of conversations.

Task 3. Moderation II: Semantic Speech Acts using Entity Coherence.

Local entity-based coherence is a method to capture text relatedness at the level of sentence-to-sentence transitions [11] In this work we adapt the single text application of local coherence to the analysis of multi-party conversations. Central to this approach is the entity-grid representation of discourse, which captures patterns of entity distribution in a text. The development of the entity-grid is based in an algorithm that automatically abstracts a text in a set of entity transitions sequences and records distributional, syntactic, and referential information about discourse entities. Thus, to understand better, the distribution of entities in locally coherent texts exhibits certain regularities, based in the Central Theory and others entity-based theories of discourse.

The entity-based approaches to local coherence have two important points that need to be highlighted: **Linguistic Modeling** and **Computational Modeling**. The first founds that discourse coherence is achieved in view of the way discourse entities are introduced and discussed; some entities are considered more **salient** than others, and consequently are expected to exhibit different properties. The second one is substantiated in automatic computation underlying discourse representation and the inference procedure.

The ranking approaches in NLP based in the **text planning** where it determines the content of a text by selecting a set of information-bearing units and arranging a structure yields an output.

The **entity-grid discourse representation** is the foundation of the study of coherence in this project where each text is represented by an entity grid, a two-dimensional array that captures the distribution of discourse entities across text sentences. For this study assuming that the corpus of a sentences is main clause with accompanying subordinate and adjunct clauses. Each grid cell thus corresponds to a string from a set of categories reflecting whether the entity in question is a subject(S), object(O), neither(X) and entities absent by gaps (-). Grammatical role information can be extracted from the output of a broad-coverage dependency parser present in the open-source library Spacy that has been used in this project.

The local entity transition is a sequence {S, O, X, -}^{n} that represents entity occurrences and their syntactic roles in *n* adjacent sentences. Each transition has a certain probability in a given grid,

where it is computed as a ration of its frequency divides by the total number of transitions of length two. Represent each text by a fixed set of transition sequences using a standard feature vector notation where each grid rendering *j* of a document *di* correspond to Φ(xij) = (p1(xij), p2(xij),...,pm(xij)), where *m* is the number of all predefined entity transition, and pt(xij)the probability of transition *t* in grid xij.

Grid construction is established in what sources of linguistic knowledge are essential for accurate prediction, and how to encode them succinctly in a discourse representation. For understand this more is necessary to divide this idea in three important concepts: **Entity Extraction** that is the accurate computation of entity classes being a key to computing meaningful entity grids, **Grammatical Function** indicates of an entity's prominence in discourse and **Salience** that is the way an entity is introduced and mentioned depends on its global role in a given discourse.

The last but important point about coherence is the **learning** where encoding texts as entity transition sequences constitutes on appropriate feature set for learning (rather than manually specifying) such a ranking function.

Practical Results:

During this phase of the project the code that was developed is based in the theorical source that was studied. In figure 1 can see the implementation of the Cosine Similarity of two Numpy vectors v1 and v2. In figure 2 is the function of TFiDF that receive as arguments the results of the functions of TF and iDF and returns the result of TFiDF as a matrix. Part of the coherence method is demonstrated in the figure 3 where preprocesses the input, apply the Spacy pipeline of Portugues model sentence-to sentence, select Nouns using pos and Subject, Object with dep_. The full code will be available in together this document.

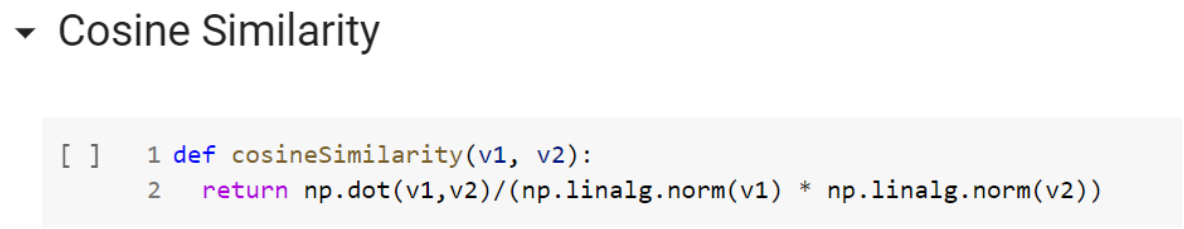


Figure 1. The implementation of Cosine Similarity

TFIDF

```
[ ]  1 def get_tf_idf(tf, idf):
     2   tf_idf = tf.copy()
     3   # iterate over each word
     4   for token in tf_idf.index:
     5     for doc_id in tf_idf.columns:
     6       tf_idf.loc[token, doc_id] = tf_idf.loc[token, doc_id] * idf[token]
     7   return tf_idf
     8
     9 TF_IDF = get_tf_idf(TF, IDF)
    10
    11 TF_IDF
```

Figure 2. Implementation of the TFiDF in Python

```
for sentence in sentences:
    sentence = re.sub("[^\w\s]", "", sentence)
    sentence = list(sentence.lower().split())
    misspelled = spell.unknown(sentence)
    sentence = [word for word in sentence if word not in misspelled]

    sentence = " ".join(sentence)
    proc_sentences.append(nlp(sentence))

for sentence in proc_sentences:
    for token in sentence:
        if (token.pos == 92 or token.pos == 96) and token.lemma_ not in nouns and translate(token.dep_)!='X':
            nouns.append(token.lemma_)
nouns.sort()
print(len(nouns))

batches = load_batches(proc_sentences, 10, 5)
```

Figure 3. Implementation of Coherence Method

The deployed implementation, after publication of the results in a scientific journal, will become part of the online debate platform Debaqi, as part of the automated moderator, that will allow the platform to work for a large population of students nationwide without the need of human moderators.

5 Benchmarking

Since this is a scientific project and not oriented to the development of a product, the Benchmarking section is replaced with the scientific validation reports. This way is important to understand that Artificial Intelligence is an attempt to replicate the human or animal intelligence using computers. Over many years, researchers explored knowledge representation and symbolic artificial intelligence, which relies in human experts creating rules to be executed by an agent. Natural Language Processing (NLP) is defined as a branch of AI, where occurs the interactions between computers and human language. This field is widely recognized for work with speech recognition, natural language understands and natural language generation.
Since Alan Turing's proposed the Turing test on the Computer Machinery and Intelligence paper, researchers are trying to create programs with the ability to act like a human on a real time conversation. With the advancements of deep learning the creation of chatbots became popular over the past years and have been used for marketing and customer service application and for this research as a moderated of debates.

This work aim creating an interaction between a human and the artificial intelligence, we built a structure of a chatbot that provides a dialogue based in Coherence. Several important methods and technologies were evaluated to develop this research from scratch, Spacy is an open-source software library for advanced natural language processing, written in the programming languages Python. Advanced methods as TFiDF combined with Cosine Similarity brings the statistical measure that evaluates how relevant a word is to a document in a collection of documents, Bag-of-word as a representation that turns arbitrary text into fixed-length vectors by counting how many times each word appears and the Coherence offers the possibility of abstract a text int a set of entity transitions sequence and records distributional, syntactic and referential information about discourse entities, were part of the constitution of this research. The implementation delivers the expected results with a good comparation.

5.1 Entity grid

The second essential stage comes from the application of the Coherence to conceive an understanding of the flow discourse during the debate using the chatbot. In fact, to achieve the desired goal of the text planning was needs to divide into subtasks to conquer the final result. As already explained, the local coherence in a text is definitely necessary for global coherence. Moreover, to accomplish the local coherence is need to select the discourse entities along the text-based (sentence-to-sentence), in the dependency parser and creating the entity-grid. An important point that needs to be emphasize is that the coherence theory considers that the distribution of entities in coherent text exhibits certain regularities reflected in grid topology. Based in this analyze is possible to create the entity grid and analyze the local entity transitions. For realizing this stage to creating the entity-grid was used two debate sources from the internet.

In the first experiment the length of the dataset was around 66 phrases in a debate and the number of transitions were 703 using the Spacy model applied in the codebase and manually the result were 772 transitions, where the dense columns were in this case subjects and objects. In this first experiment it is possible to interpret the results of the dataset, by the probability of the transitions (computed as a ratio of its frequency divided by the total number of transitions of length two) and the model achieved high percentage of acceptable results in comparation with human interpretation. In the second experiment the dataset has a slight difference in the level of coherence than the first experiment. The length of the dataset was 50 phrases and the number of transitions from the model analyze were 801 transitions against the 1024 from the manually analyze. The result

of the probability of the transitions shows that the model has a high approximation of the normal human understanding of the name entities in a phrase, however the human interpretation has better precision.

5.2 Q&A moderator based on Bag of Words.

An important support of this research is the application of the theory the bag-of-words (BoW) model is a representation that turns arbitrary text into fixed-length vectors (vectorization) by counting how many times each word appears. In this model, the data is in the form of text and is represented as the bag or multiset of its words, disregarding grammar and word order and just keeping words. Text is either a sentence or document. BoW is often used to generate features; after generating, we can derive the term-frequency of each word in a document, which can later be fed to a machine learning algorithm. The BoW theory is applied in the context of a chatbot to respond to the input added by the user, combined with the TFiDF and Cosino Similarity method, enabling greater precision in the delivery of expected results.

5.3 Utterance complexity

In this research the utterance complexity has an essential presence, considered as high, because it is data coming from online debates. The base text does not have the standard grammar structure used in the development the Coherence method. Another point to be highlighted, the entire manual analyze was carried out by a person with the standard level of knowledge of the grammatical rules of the Portuguese language.

6 Method

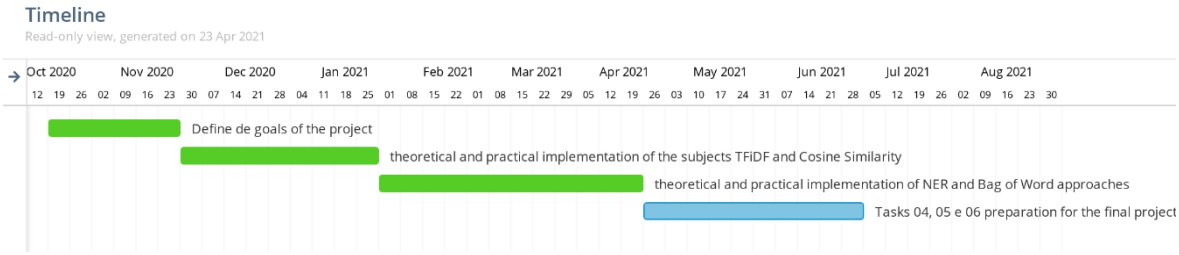


Figure 3. Work Plan (Gantt Chart)

Task O1 will deliver Python code that takes utterance strings as inputs and returns a vector representation. This can be a vector based on term-frequencies, or more complex methods to be studied during the project. This is an essential task upon which the others depend.

Tasks O2, O3, O5 and O7 will deliver Python functions to compute topic alignment and coherence between participants based on pre-existing definitions, which may need to be adapted to the type of text vectorization to be used.

Task O4 will deliver a set of qualitative norms that will inform the ground-level for universal moderation implemented in a data structure that can be read and used by the intelligent chatbot moderator given triggers created during a debate session. This structure will be designed in such a way as to make it simple to extend it to accommodate for speech acts that are relevant to specific debate topics.

Task O6 is central to this work. It will be delivered as a Python set of functions that are articulated amongst themselves in a standalone manner, and that can be readily integrated into the Debaqi "brain". It will enable Qi-Brain to "listen" to the debate and output speech acts for the intelligent chatbot moderator.

Task O8 will produce performance measures based on pilot studies in real school debates. Performance will be determined by a combination of qualitative analysis of the chatbot moderation interventions and by a questionnaire administered to the students taking part of this pilot test study.

Task O9 will document in detail all the work done for this project, include a deeper literature review section, and outline recommendations for future versions of Qi-Brain.

The plan for integration and validation:

1. **For the BoW question answering engine:**

We will produce a large synthetic dataset of questions that include many variants of a core set of questions about debate dynamics, debate topics, and other areas in which students are expected to ask questions. We will produce performance scores of how well Qi-Brain replies to the questions based on a provided gold standard. We will resort to standard practices of separating training and testing folds. The process will be iterative and may require adjustments of the vectors used to represent matching answers to eliminate redundancy and potential overlaps that may cause Qi-Brain to fail.

2. **For the utterance complexity measure**

We will select a sample of utterance from existing conversation data, perform manual annotation of complexity and match (qualitatively) against the numbers produced by Qi-Brain. This may lead to adjustments in the value of the parameter **tau** (T τ).

3. **Validation of the local coherence engine**

We will validate the local (entity based) coherence by annotating a small dataset with human perceived coherence in a small number of conversations and compare this with the results obtained from Qi-Brain. Adjustments may require further research since, currently, the method replicated from Barziley and Lapata is not parametrised.

4. **The integration**

We will test the three components working together by inviting the students of the third year of informatics engineering to take part in three short debates in the Debaqi platform where all three components are integrated. Each debate will provide information about how to best integrate the components of the domain reasoner in such a way that the debate dynamics flow naturally, and the automatic moderator becomes a good support rather than a problem. In the case that for some reason we are unable to integrate the Qi-Brain components into Debaqi on time (which depends on Debaqi engineers) we will perform one on one tests with invited students on a locally integrated version of Qi-Brain separate from the Debaqi platform.

The development of this project followed an interesting perspective during its phase. The first stage occurred in October deciding the goals of the project. The first step began with the introduction of the theoretical basis, in fact, it was the first primordial step to understand the problem and the foundation to develop a proper solution. The research was supported by state-of-art theories in the field, distinguished articles, and discussions to master the concepts of Conversation Analysis and other important pillars of Natural Language processing. The first stage was defined to take around October to February. After creating a theoretical structure around the subject, the practical stage began around February, for each method demonstrated in the articles a source code has been developed. Around Abril to June the application of Coherence method was applied.

7 Results

The algorithm created by Lapata and Barzilay was developed for the analysis of data coherence, seeking to abstract a text into a set of entity transition sequence and record distributive, syntactic and referential information about discourse entities. The research was carried out with data provided by newspapers and magazines. The situation presented in this research used the method developed by both in dealing with the automatic coherence of a text in a supervised debate in a chatbot. The elaboration of this study seeks for an approximation of the local coherence to solve the problem in question. The first stages of the research presented here were the application of the Coherence algorithm structure in Python language, the use of spaCy's Portuguese model (an open-source software library for advanced Natural Language Processing) and two datasets of online debates, in this case from the Kialo website.

7.1 Solution to the problem

During the elaboration process of this research, interesting points to be worked on were found. For the implementation of the algorithm is necessary to select the entities locally, that is, in each sentence contained in the data set. The model is able to select each entity and its respective value predicted syntactic dependencies; from that it can be classified by its type. However, the selection performed by the model using only syntactic dependences(dep_) have not the necessary precision to consider an alternative to be applied. The model only selected the nouns that were in capital letters. The best alternative used to solve the problem came from the structure created in the source code of this project, where all the nouns of each sentence are selected using the model's part-of-speech tag (pos_) and then applied the syntactic dependences tag selecting among the nouns already selected which are Subject, Object or just nouns. In this way the precision is greater and the problem of selecting only nouns with capital letters has been solved.

7.2 Answering questions.

The research developed seeks to meet the needs of a chatbot by answering questions based on the input that users provide. To achieve the necessary precision, three distinguished methods were used, bag of words combined with TFiDF and Cosino Similarity. The procedure to receive the input and process was developed and based on the similarity of the input vectors received to select the best output.

7.3 Requirements fulfilled.

Following the predefined requirements, the structure of the project was fulfilled as expected. The requirements R01, R05, R07, R08 and R9.2 were features provided by the system QiBrain. The R02 and R04 requirement was successful done by the application in two different ways in the prototype stage with the TFiDF and entity grids. The R03 were partially done for developing a first version of a language model that need to be trained. R06 requirement was fulfilled by creating as proper structure to receive data from different sources and has speech act plans developed at prototype stage for question answering. The R8.1 requirement has changed as the research progressed and explain we switched to the above namely: TFIDF, and entity grids for which leveraged the

computation of PoS and dependencies using Spacy model in Portugues. R8.2 requirement was similarly developed with the TFiDF and entity grid, however was not tested in a productive environment. Requirement R9.1 was divided in subtopics that was part feature provided by the system and other have found the evidence that this can be done based in the entity grid method. The prototype has proof that the topics can be done. The last requirement was R10 is possible to product given the results of the prototype.

8 Conclusion

The present research achieved the expected results, the functional structure of the chatbot based on automatic local coherence was achieved. The conclusions obtained from this research show the level of relationship between what the algorithm can perceive and the superiority of human understanding. The use of the model in Portuguese has been shown to be quite flawed in the matter of selection of nouns both in lowercase letters, pronominal form of the nouns and the evaluation of the word being or not a noun, especially in the beginning of sentences. The Coherence algorithm demonstrates a high level of precision and that to overcome the model's flaws it was necessary to create alternatives based on the source code of this project. Currently, the algorithm in question has a high level of precision present in the results of this study, however it still needs to be better modeled for situations of high level of complexity as the current one presented. The next steps of this research are the application of the study elaborated in the chatbot system, part of the QiBrain project, and based on this, and looking for a constant improvement in its accuracy level.

Bibliography

[1] M. Dunleavy, C. Dede and R. Mitchell, "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning.," *Journal of science Education and Technology,* vol. 18, no. 1, pp. 7-22, 2009.

[2] K. Bennani-Smires, C. Musat, A. Hossmann, M. Baeriswyl and M. Jaggi, "Simple Unsupervised Keyphrase Extraction using Sentence Embeddings," [Online]. Available: arXiv preprint arXiv:1801.04470.

[3] M. Gerlach, T. Peixoto and E. G. Altmann, "A network approach to topic models.," *Science Advances,* vol. 4, no. 7, 2018.

[4] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine.," *Communications of the ACM,* vol. 9, no. 1, 1966.

[5] J. Carbonell, "AI in CAI: An artificial-intelligence approach to computer-assisted instruction," *IEEE transactions on man-machine systems,* vol. 11, no. 4, pp. 190-202, 1970.

[6] D. Sleeman and J. S. Brown, Intelligent tutoring systems., New York: Academic Press, 1982.

[7] E. Shortliffe, Computer-based medical consultations: MYCIN, New York: Elsevier, 1976.

[8] C. Callaway, M. Dzikovska, E. Farrow, M. Marques-Pita, C. Matheson and J. D. Moore, "The Beetle and BeeDiff tutoring systems.," in *SLaTE Workshop on Speech and Language Technology in Education*, Farminton, USA, 2007.

[9] J. Alexander, "Dialogic Pedagogy in a Post-truth World.," in Dialogic Pedagogy: An International Online Journal, 7., USA, 2019.

[10] Huang, A. (2008, April). Similarity measures for text document clustering. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand (Vol. 4, pp. 9-56).

[11] Barzilay, R., & Lapata, M. (2008). Modeling local coherence: An entity-based approach. Computational Linguistics, 34(1), 1-34.

Anexo

Reinforcement on ethical guiding principles

Our system is based on ethical and civic principles to offer a better user experience:

1. Every user in the platform is completely anonymous, the account name is generated by the system.
2. Our system does not perform any analyses on individual behavior.
3. Our system only derives indicators on collective conversational patterns.
4. Users of the system are underaged, thus we have parental and student signed consents.
5. The project is overlooked by an ethics committee from University College London, led by Prof. Kaska Porayska Pomsta.

Glossary

LEI	Licenciatura em Engenharia Informática
LIG	Licenciatura em Informática de Gestão
TFC	Trabalho Final de Curso
ITS	Intelligent Tutoring System