Studies on Computational Ontologies Towards Socio-Enactive Systems

Felipe Rodrigues Jensen *Andrey Victor Justo* *Ivo Calado*
Julio Cesar dos Reis *Rodrigo Bonacin* *Gonzaga Souza*
Ig Ibert Bittencourt *Maria Cecília Calani Baranauskas*

Technical Report - IC-18-04 - Relatório Técnico
March - 2018 - Março

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

Studies on Computational Ontologies Towards Socio-Enactive Systems

Felipe Rodrigues Jensen* Andrey Victor Justo* Ivo Calado†
Julio Cesar dos Reis* Rodrigo Bonacin‡ Gonzaga Souza‡ Ig Ibert Bittencourt§
Maria Cecília Calani Baranauskas*

Abstract

Computational ontologies refer to formal representation models that characterize domain concepts. These artifacts are relevant to shape the meanings of data in an explicit way for humans and machines. The study of socio-enactive systems demands the understanding of human actions and the adequate system response in a cyclical interaction. Ontologies can play a key role in the development of socio-enactive systems by providing meaning to data entering and leaving the systems. However, this research field still lacks thorough literature studies. This technical report presents the results achieved by a research working group to further develop methods for ontology engineering in the context of socio-enactive systems. We present a literature review, technical solutions, in addition to conceptual and practical outcomes. Our findings include a preliminary modeling of concepts involved in socio-enactive systems, and a software architecture that allows ontology-based interpretation of several types of data. This research has been developed in the socio-enactive systems project, a FAPESP's thematic project (grant #2015/16528-0).

Keywords: Enactive Systems; Socio-Enactive Systems; Ontologies; SWRL.

1 Introduction

Ontologies aim to represent semantics in computational systems (Gruber 1995) and have been designed to provide rich machine-decidable semantic representations. They refer to a formal specification of some domain, formalizing a conceptualization of a domain in terms of classes, properties and relationships between classes. Usually, a syntactic structure models the concepts of a knowledge domain, and serves as schemas that organize data expressing instances of concepts according to logical properties. The classes are the focus of ontologies because they describe the domain concepts representing groups of individuals that share properties. Data properties characterize individual attributes, whereas object properties specify the relationships between individuals of the same or different classes.

We have investigated the concept of socio-enactive systems as an extension of enactive systems (Kaipainen, Ravaja, et al. 2011). Enactive systems refer to a class of systems in which the human-computer interaction is developed by specific interfaces based on the use of sensors and the data

*Institute of Computing, University of Campinas
†Federal Institute of Alagoas
‡Center for Information Technology Renato Archer
§Federal University of Alagoas

analysis of physical interactions (Kaipainen, Ravaja, et al. 2011). To this end, the system requires capturing and interpreting psychophysiological states by including, for instance, the emotions from the involved participants in the interaction. The system should change its behaviour according to the input data sensors which must affect the involved participants of the interaction in a cyclic feedback procedure. An socio-enactive system aims to put in first place the social and cultural aspects of interaction in such class of systems.

The research challenges in the design of socio-enactive systems refer mostly to the difficulties of capturing, modeling and interpreting human social environment. To this end, we need to take into account the interaction and social context when modeling the system's knowledge and actions. The use of ontologies represents an alternative to the design of socio-enactive systems, because they are artifacts to represent semantics in computational systems, by describing concepts and interrelationships among them. These artifacts have the potential to support the representation of conceptual knowledge and behavioral repertoire, and provide interpretation capacities for the system.

This technical report presents the preliminary results, in the context of a long-term investigation, obtained from the research conducted for the investigation of ontologies in socio-enactive systems. Our research has focused on theoretical and technological issues. We assume that the study of theoretical and technological aspects together can give us clues to consider the socio perspective in enactive systems. At this stage, our aim was to obtain a comprehensive literature review and the first practical experimentations of prototypes to understand the involved concepts. In addition, our goal was to clarify the research challenges.

The objective of the literature review is to understand the way ontologies have been employed in enactive systems. We developed a proof of concept implemented to investigate the potentialities and challenges related to the use of ontologies in enactive systems. Our prototype stands for an interactive bot, which captures the user's emotion, and reacts properly based on a set of formal rules in an emotion ontology to express a series of behaviors related to affective answers. The prototype was a way to investigate how the ontologies must be explored, the involved issues, and the means to conduct experimental evaluation in order to elucidate user's requirements and needs. From a technical perspective, the practical results have informed the definition of a software architecture to organize the relevant components of the system to further support non-verbal conversations, in a way to provide context-awareness based on the interpretation of emotional states from people.

This technical report is organized as follows: Section 2 presents the findings on the literature review. Section 3 describes the conceptual and the system prototype results. Section 4 discusses our findings highlighting the detected open research challenges, and Section 5 presents the concluding remarks.

2 Literature Review

The theoretical study involved a systematic literature review to understand existing approaches exploring the use of ontologies in enactive systems. In the following, we present the method to conduct the literature review and the reached results.

2.1 Method of Analysis

In our literature review, we used the PRISMA (Moher et al. 2009) method for conducting a systematic examination. We first collected articles in several scientific bases. After the collection, we conducted a phase of analysis and synthesis to identify the most relevant and related contributions.

Our goal was to understand how ontologies are used to construct enactive systems. Our analysis allowed to define an organization of categories and to elucidate the underlying open research challenges.

Following the PRISMA framework, we first defined a title for our review. In the sequence, we defined a structured summary, a rationale statement and the objective of the review, which can be summarized as follows: "Search and review of ontology literature in enactive systems from 2005 to 2017".

Concerning the adopted method (next item of PRISMA framework), we performed the review according to following steps: (1) Search scientific databases, (2) Select papers based on their titles, abstracts, and keywords, (3) Read the entire text and application of the inclusion and exclusion criteria, (4) Perform critical evaluation, and (5) Selection of the studies and presentation on project' meetings.

In our review, we chose not to include all the PICOS (Participants, Interventions, Comparisons, Outcomes, and Study design) parameters in our search string due to the reduced number of papers on the themes in focus (as these themes are relatively new), *i.e.*, we chose to use a string that would return a larger number of results. Thus, we used reduced "AND" connectors in the search strings. Table 1 presents the search strings parameters used in this review, which includes the following themes: Enactive Interfaces (EI), Enactive Media (EM), Enactive Systems (ES), and Enactive system Ontologies (EO). Our inclusion criteria is based on the analysis of the papers' contribution for at least one of these themes. The exclusion criteria are: non-scientific works, abstracts and papers not available in digital media. The following databases were considered in this review: Academia.com, ACM DL, Google Scholar, IEEE Explorer, PubMed, Research Gate, Science Direct, Springer Link, and Web of Science.

Table 1: Search strings used in the systematic literature review.

INITIALS	Search strings
EI	Enactive interface
EM	Enactive media
ES	Enactive system
EO	Enactive system ontology; enactive ontology; emotion ontology; soft ontology; ontospace; ontological dimension

2.2 Results

Table 2 presents the results found for each one of the portals investigated and Table 3 presents the results organized by year. We focus on the description of the articles contributing regarding the ontology perspective in the five past years.

Table 2: Results found in each one of the Portals.

PORTAL	EI (10)	EM (05)	ES (11)	EO (28)	TOTAL (54)
Academia.com		2	1		3
ACM DL	2			5	7

PORTAL	EI (10)	EM (05)	ES (11)	EO (28)	TOTAL (54)
Google Scholar	2	2	2	8	14
IEEE Explorer	3		2	1	6
PubMed			3	2	5
Research Gate				3	3
Science Direct	3		1	5	7
Springer Link			2	5	7
Web of Science		1			1

Table 3: Results organized by year focused on the last 5 years.

YEAR	EI	EM	ES	EO
2013	2		4	2
2014	2	1	3	2
2015	1	1		4
2016	1			4
2017				2

Borth, T. Chen, et al. (2013) proposed a large visual sentiment ontology to enable high-level sentiment analysis of social multimedia. They presented applications for large image exploration and live media monitoring. Based on the Visual Sentiment Ontology, Borth, Ji, et al. (2013) studied sentiment analysis from visual content for understanding the visual concepts that are strongly related to sentiments. Their framework for constructing the visual sentiment ontology is used for multimodal sentiment prediction relying on machine learning classifiers. Similarly, Gil et al. (2015) proposed the *EmotionsOnto* as a generic ontology for describing emotions. They explored such ontology in experiments conducted to automatically measure users' emotional states using Brain–Computer Interfaces.

Pugliese et al. (2014) studied interactive narrative with the association of story elements with metadata for the creation of a story ontospace. They explored ontologies for the constitution of the metadata. Hautamäki (2016) studied the concept of point of view. Using a formal approach, the author defined conceptual spaces as a way to handle point of view. The work developed a new logic for points of view based on the framework of conceptual space.

Jou et al. (2015) proposed a method to discover adjective-noun pairs that represent sentiment in various languages. This method was used in the construction of the multilingual visual sentiment ontology (MVSO). The MVSO encompasses visual sentiment concepts in 12 languages. The analysis of the application of this ontology in a large image dataset points out that emotions are not necessarily culturally universal. *i.e.*, frequently visual sentiment is expressed and perceived in a culture dependent way.

Thakor and Sasi (2015) presented the Ontology-based Sentiment Analysis Process for Social Media (OSAPS). The OSAPS focused on the analysis of negative sentiments. This process was applied to the analysis of customers' tweets expressing dissatisfaction of delivery services. The results can be used by the postal company to take corrective measures, as well as they showed difficulties in evolving and maintaining the relationships between the ontology' classes, in such an ontology.

Kaipainen and Hautamäki (2015) proposed an approach for describing conceptual space in which the entities definitions depend on the perspective from which they are considered. The perspectives are related to short-term (narrative and situational factors and interpretative frames at the moment of observation) and long-term (natural conditions, evolution, or life-long learning) contexts. Kaipainen and Hautamäki (2015) presented a case study in which dimensions are scaled to range between 0 and 1 to represent properties describing coordinates of each item within the ontospace. This model allows the epistemic exploration of a conceptual space from multiple perspectives.

R.-C. Chen et al. (2016, March 1) argued that social media offers opportunities for researchers to discover knowledge about users' interest. To this end, they proposed a method to improve personalized recommendation systems based on personalized ontologies, social network and real time communication analysis. The cold start problem was minimized by the social data, such as the users' communities. Experiments with data from 100 users as a training dataset and 30 users as the testing dataset showed the effectivity of the approach in recommending movies, musics, and books.

López-Gil et al. (2016) investigated the development of an ontology for representing emotional, cognitive, and motivational state of online learners. The objective was to develop systems able to present emotional responses by interpreting the proposed ontology. The application of the proposed ontology in MOOCs revealed the capacity of representing emotions and students' motivation in a simpler way; as compared with other emotional ontologies that add (according to the authors) too much unnecessary complexity. This allowed its application even by non-experts in ontology modeling.

Meacham (2016) affirmed that the concept of embodied cognition (enaction) can be observed in a scale below the level of a single-celled organism. By analysing the behavior of proteins and utilizing Kováčs' definition of sentient proteins, as "the capacity to exhibit a variety of potential internal states responding to the immediate state of the environment", Meacham (2016) argued that the behavior of the protein can be considered as cognition and as embodied cognition because it is experimenting with the environment and evolving through this process. He also argued that by observing this phenomenon, an "enactive ontology" must consider individuation and regulation of meaning-relations across all scales, so that inferences consider this interwind of nested and overlapping environments.

Villalonga et al. (2016) proposed the combination of different types of data categorized as activities, locations and emotions. There are several sources for this data like video, audio and body sensors. This combination was proposed throughout the observation of the usual ways to process data gathered by these sources, which are generally in a sole dimension and not in a combination of them. They showed that a richer and more meaningful expression of contexts can be obtained by the combination of these categories. The authors presented an ontology-based method that combines these categories referred to low-level contexts (activities, locations and emotions); and the result of this combination is referred as high-level contexts. Their work contributed with an open ontology called "Mining Minds Context Ontology", which links low-level to high-level contexts through the integration of contextual definitions, and a framework named "Mining Minds High-Level Context Architecture" build upon this ontology. The framework relies on reasoning techniques to enable the inference of high-level contexts from low-level contexts.

Bader et al. (2017) investigated the relation between the emotions expressed on online reviews of movies and the emotion expressed by the audience while watching the corresponding movie. The authors intended to determine if the emotions extracted from online reviews could be used to construct emotional signatures of both a movie and, in a higher perspective, of an entire genre as a whole. The study relied on a large dataset of reviews originated from the IMDb movie dataset site. For each considered movie, the work analyzed if the emotions expressed on the corresponding online reviews present a parallel on the emotions elicited from the movie audience. Results confirmed the hypothesis showing that online reviews create an emotional signature from the movie. It was also possible to conclude that such emotional signatures are applicable to genres; and similar genres present closer emotional signatures.

Phan et al. (2017) studied the problem of human behaviour prediction in the context of health social networks. The authors argued that the modeling of human behaviour is a key component for systems because such capability enable trust, engagement and loyalty from users. The work proposed a ontology-based deep learning model to predict and explain human behaviours. The model considered several aspects of human condition such as biological measure (BMI, slope, wellness, etc) and social (online and offline) activit. The defined model took also into account the explicit social influences that affect a human being. The study performed an extensive experiment considering both real-world and synthetic social networks aiming to determine the accuracy of the proposed model. Results showed that it is possible to obtain a deep understanding of the human behaviour determinants.

3 Prototype of Ontology-based Enactive System

In parallel to the theoretical studies (*cf.* Section 2), we have focused on developing a prototype in small cycles to quickly obtain practical results for understanding socio-enactive systems. The prototype aimed at constructing and testing small pieces of enactive systems in simplified scenarios to experiment existing technologies. The obtained results include:

1. A preliminary modeling of underlying concepts involved in socio-enactive systems.

2. A prototype to support a preliminary proof of concept for using ontologies in enactive systems.

3. A software architecture that organizes the components of an ontology-based socio-enactive system.

3.1 Modeling of Socioenactive Systems

This task aimed to initially clarify the underlying concepts related to socio-enactive systems. The modeling was important to clarify and organize the relevant concepts involved in the conception of socio-enactive systems. For this purpose, we created a set of concept maps to understand the researchers' perception of the underlying concepts.

Figure 1 presents a conceptual map created to represent the concepts involved with socio-enactive systems. According to it, the conception of socio-enactive system relies on concepts coming from philosophy, cognitive psychology, linguistics, biology, neuroscience and sociology. For example, cognitive psychology aims to understand mental process the emotions belong to.

Socioenactive systems might make use of technologies coming from artificial intelligence, which can include computational ontologies, machine learning techniques, etc. Computational ontologies can model the mental processes. Devices and sensors belong to the technological aspect of the system.

The produced set of concept maps is a first step to the definition of socio-enactive systems. This can be used as a communication artifact among the researchers aiming at providing an harmonized view of the socio-enactive concepts. Future steps will involve the creation of further extensive maps and the potential merge of existing ones. This will require adequate methodologies to allow the contribution of other researchers involved in the project.

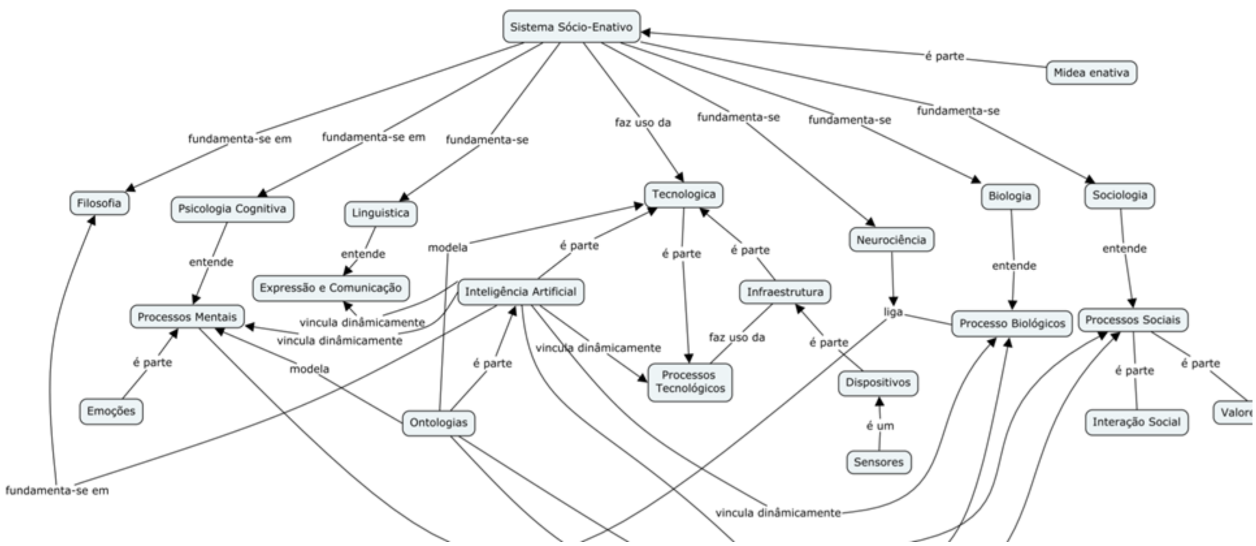

Figure 1: Example of concept map to understand socio-enactive systems.

3.2 The Constructed Prototype

We aimed at studying the role of ontologies in the development of enactive systems. Our goal was to understand how the concept of enaction is supported with the use of ontologies. The following research questions have guided our work:

- How to design, reuse and construct ontologies for the development of socio-enactive systems?
- How the system's behaviour can be guided by the ontologies?
- How to accommodate different types of ontologies in the system?
- How a software architecture based on the ontologies can be reused for considering socio-enactive systems in different domain contexts?

The prototype is an instance of an enactive system. The goal of our enactive system prototype was to recognize users' emotions and to adequately provide a feedback which is informed by some data interpretation in an ontology. We need an ontology to represent emotions and the system's action are inferred based on the user's input state. We propose the construction of an interactive bot that reacts to users' captured emotions. The bot generates and provides expressive answers relying on a set of faces expressing emotions. For the development of the software, we organized the prototype in 4 major modules. Each module holds a specific objective and provides data to others.

A module was responsible for capturing the users' emotion based on video streams (Perception module). Another module should use ontologies to represent emotional states, interpret input data and run automatic reasoning to provide the system's answers (Inference module). We considered a renderization module to create the bot's output emotional answers. An integration module was responsible to provide features for integrating all the involved software modules. In the following, we describe the modules: (1) Perception; (2) Inference; (3) Renderization; and (4) Integration.

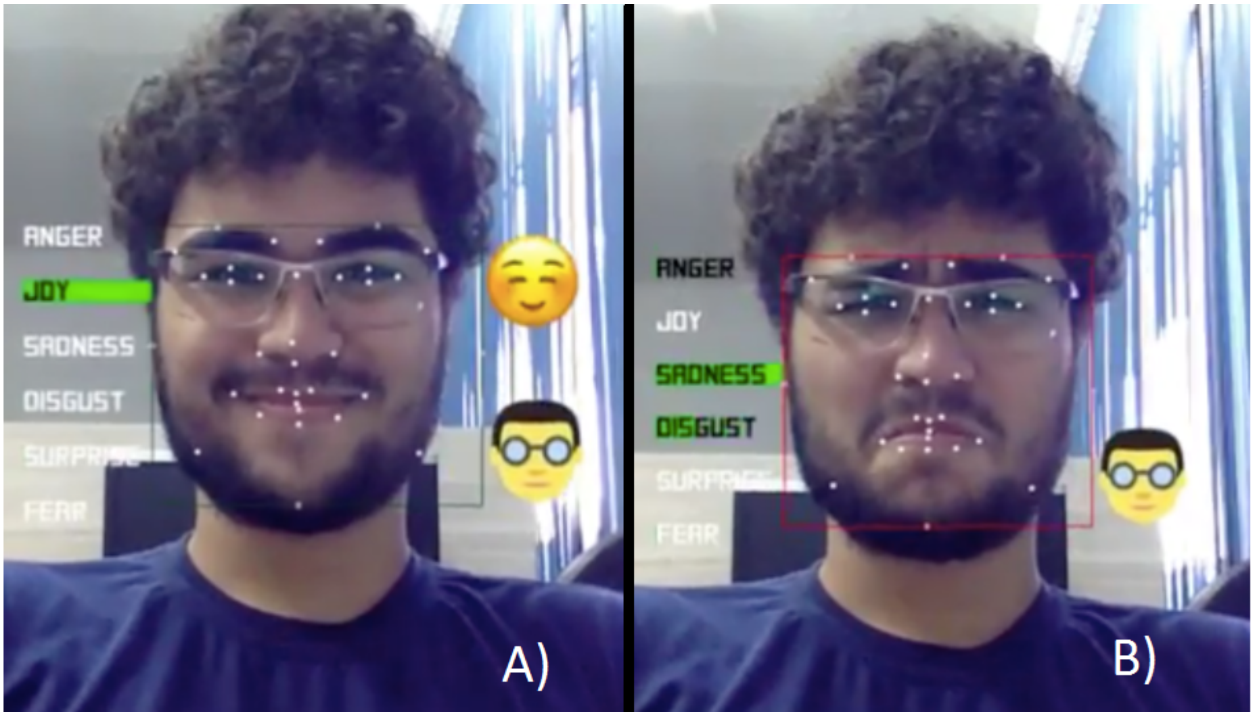

Figure 2: Example of perception module recognition.

3.2.1 Perception module

The objective of this module was capturing the user's face emotions based on the input of video streams through a webcam. In order to achieve that, we explored a customized version of the Affectivas's SDK[1] that enables the recognition of 6 emotional states (anger, joy, sadness, disgust, surprise and fear) according to Paul Ekman's emotion theory (Ekman 2007). Figure 2 presents an example of recognizing different emotions (A - joy; B - sadness in Figure 2).

The Emotion SDK for Windows is distributed as a Installer - where we made changes according to our needs. Its included assemblies enable integration with .NET and C++ Windows applications and the data folder required by API in runtime:

- affdex-native: C++ Namespace headers and library files (the Release and Debug versions).
- Affdex: .NET Namespace assembly (the Release and Debug versions).
- data: The classifier data files required by both affdex-native and Affdex libraries runtime.

In a time space of 2 seconds (customizable), the module automatically detects the most relevant user's emotion and represents it in an structured way via a file in JSOn format. The module sends this information to the integration module that handles it accordingly.

3.2.2 Inference module

Our approach is interested in understanding users' emotional states because it is an intrinsic part of human communication (Neviarouskaya et al. 2010). The users' affective expressions are interpreted

[1] https://developer.affectiva.com

to infer their emotional state. They are then used to interpret incoming information and a way to decide which expressive actions should be taken by the system. This decision is based on expected emotions to be caused. For this purpose, this work originally defined a set of rules to handle affective aspects. In the following, we present the explored technology, artifacts and theories to reach our goals.

Our defined rules to express affective states explore existing concepts in ontologies. To this end, we have used the *Emotion Ontology* (Hastings et al. 2014). This ontology aims to provide an extensive representation of affective phenomena. It defines several classes to describe and relate emotions, moods, appraisals and subjective feelings. This representation formalizes the types of emotional states, relevant to be used in virtual agents. In this sense, the *Emotion Ontology* allows us to have a formal representation for reasoning purpose in a way to provide decisions on the system affective response. The *Emotion Ontology* is available in OWL language.

Our defined rules rely on the *Semantic Web Rule Language* (SWRL). SWRL is a language useful to express logical rules using OWL ontology concepts (Horrocks et al. 2004). The logical rules is used to infer knowledge within an ontology. According to Horrocks et al. (2004), SWRL rules are defined in the form of implications between an antecedent and a consequent. The antecedents and consequents consist of a list of zero or more atoms typically presented in the form of predicates from first-order logic.

We explore SWRL rules in conjunction with ontologies to define behaviors being observed from the environment. Our proposal infers reactions and possible behaviors resulting from the reactions. We assume that this can be useful in an attempt to maintain engagement and to promote the empathy between the system and humans. In the developed architecture, the system interprets affective expressions to determine the emotional state from the person it is interacting with. The system then predicts the possible behavioral/emotional changes related to its actions. In our proposal, we intent that the system allows to choose appropriate behaviors to maintain a context-aware, comfortable and enjoyable engagement based on human interaction features.

In this work, the definition of rules for the modeling of system reactions is based on the mirror neuron theory (Heyes 2010). This theory brings forth the mirror neuron, a kind of neuron whose function is activated when the observer performs an action or when an action is observed. The neuron then mirrors the action as if it was being performed by the observer. Several researchers have acknowledged that these neurons can help in learning and in emotion understanding, like empathy (Preston and De Waal 2002).

This theory states that a person's mirror neuron observes and internally replicates what is begin observed, *i.e.*, if someone observes a motion, the mirror neurons capture that motion and assume that they know how to manipulate the body, so it can mimic the observed motion. The same goes for an emotion. If a person observes an emotion being expressed, the mirror neurons capture that emotion and assume they understand it. The mirror neurons can then "explain" how to feel that kind of emotion, and possibly what kind of reactions or emotions would pair well with that observed emotion. For example, if someone sees joy, it might feel the urge to keep the joyful situation. Otherwise, if someone sees sorrow, it might try to take actions or express emotions that might alleviate it.

The definition of rules is usually bound to the knowledge domain of interaction. In this work, we have emphasized the modeling of rules to create reaction considering emotional states. For this purpose, we explored the *Emotion Ontology*. A SWRL rule base, which we modeled, links the interpreted affective gestures with emotional states modeled in the ontology; an action set is then inferred.

The process to create our rule set was based on the *Emotion Ontology* and the mirror neuron theory. To this end, we analyzed the defined affective states in the ontology, and modeled according

to the theory the reactions according to each type of affective state. For instance, if the state of joy is detected as input, the modeled rule expresses that the system's reaction must keep the joy state.
In order to define the rules, we needed to model specific classes related to interaction. For instance, in the ontology, a class named *Person* represents people interacting with or being observed by the system. We defined a base rule as a format to interpret and act upon a person's emotional states (captured from the Affectivity Interpretation component and included as class instances in the ontology). Based on the different emotional states, distinct rules were created to represent the different actions relying on the frame of the base rule as follows:

$$Person(?p), emotionClassName(?e), personEmotionalState(?p, ?e) -> action(?p, listOfActions)$$

This base rule expresses that there is a Person p and there exists an emotion state e *emotionClassName*. The *emotionClassName* are all modeled in the *Emotion Ontology* as explicit classes which includes, but it is not limited to:

$$emotionClassName = \{Anger, Disgust, Surprise, Sadness, Joy, Fear, Boredom\}$$

The Person p is in an emotional state caused by the emotion e (of input) defined by the *personEmotionalState*. The detection of a person's emotional state derives a list of actions that the system might choose to execute to achieve a determined objective in the interaction context. The list of actions includes a set of emotions to be expressed. Formally:

$$listOfActions = \{emotionClassName_1, emotionClassName_2, ..., emotionClassName_n\}$$

For instance, following the example of a person who is feeling sadness, a possible list of actions, that might try to alleviate the sorrow, could be: feeling *surprised* with the possible tragic situation, then sharing the state of *sadness* with the person, and then trying to cheer the person up with the *joy* of positive thoughts. This sequence of actions is explicitly modeled in the rule and represents the system's output behavior. The following rule formally expresses this example:

$$Person(?p), Sadness(?e), personEmotionalState(?p, ?e) -> action(?p, \{Surprise, Sadness, Joy\})$$

Another example could be given in the instantiation of the disgust emotion. By observing someone hurling emotions of disgust, one in an attempt to revoke this emotion to bring the portrayer to a state of joy could present *surprise* feeling. The hope is that the person starts to reason upon the surprised act and realize that his/her disgust is somewhat awkward. If no result is perceived, the feeling of *sadness* could be presented to convince that his/her attitude can make people around sad. Another emotion that would revoke disgust is *anger*. This emotion could intimidate the person with the distasteful feeling in a manner that he/she would feel apprehensive to continue exposing such emotion. All in all at the final stage, *joy* is presented to try to make the person forget about his/her disgusted state and come back to enjoy other good things. The following rule express this example:

$$Person(?p), Disgust(?e), personEmotionalState(?p, ?e) -> action(?p, \{Surprise, Sadness, Anger, Joy\})$$

The reaction of the person to the system's actions is then evaluated by rules in a new interaction cycle. In a long-term perspective, we have histories of conversations and users' characteristics data, which can be explored on further action decisions.

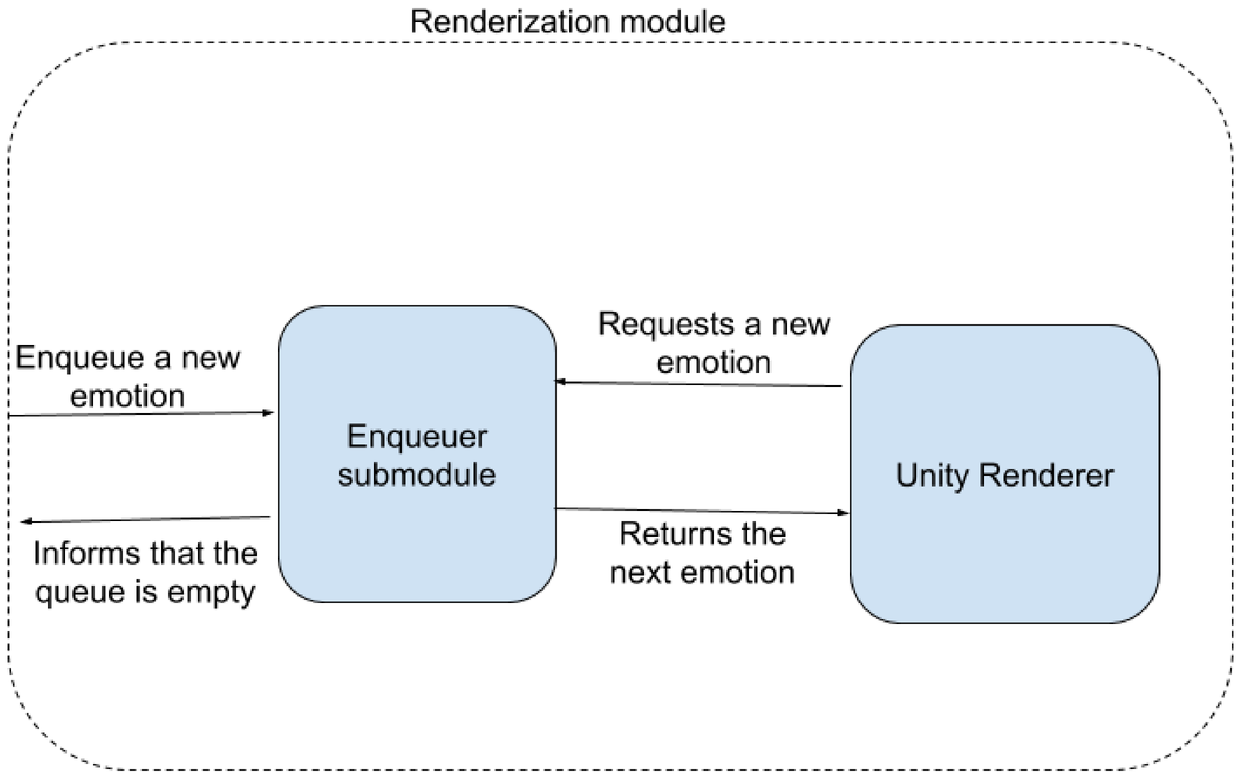

Figure 3: Organization of submodules of the renderization module.

3.2.3 Renderization module

The goal of the renderization module is the generation of a visual output from the data produced by the inference module. We rely on a 3D model of a human being as the approach to the visual representation. Two following aspects have influenced such decision:

- Present a visual representation that could apparently mimetize the actions of the user;
- Present a fluid transition between different emotions.

The renderization module is organized into two submodules, named Enqueuer and Mesh, in which the communication one has with the other is performed via HTTP REST requests (cf. Figure 3).

The first submodule implements a queue of emotions, sent by the inference module (through the integration module); these emotions are rendered. Considering that the inference module might produce emotions in a faster ratio than the renderization module can display, the enqueuer submodule implements a producer-consumer strategy where the inference module is the producer and the mesh submodule (described below) is the consumer. The Mesh submodule is responsible for consuming such emotions.

The second submodule, also named Mesh Submodule, renders the emotions obtained from the Enqueuer submodule by producing a 3D face that represents the emotion. For this purpose, we pre-built six meshes representing the six types of emotions supported in the prototype. In addition, we built an additional face mesh representing a neutral face. Figure 4 presents these pre-built meshes representing the aforementioned emotions.

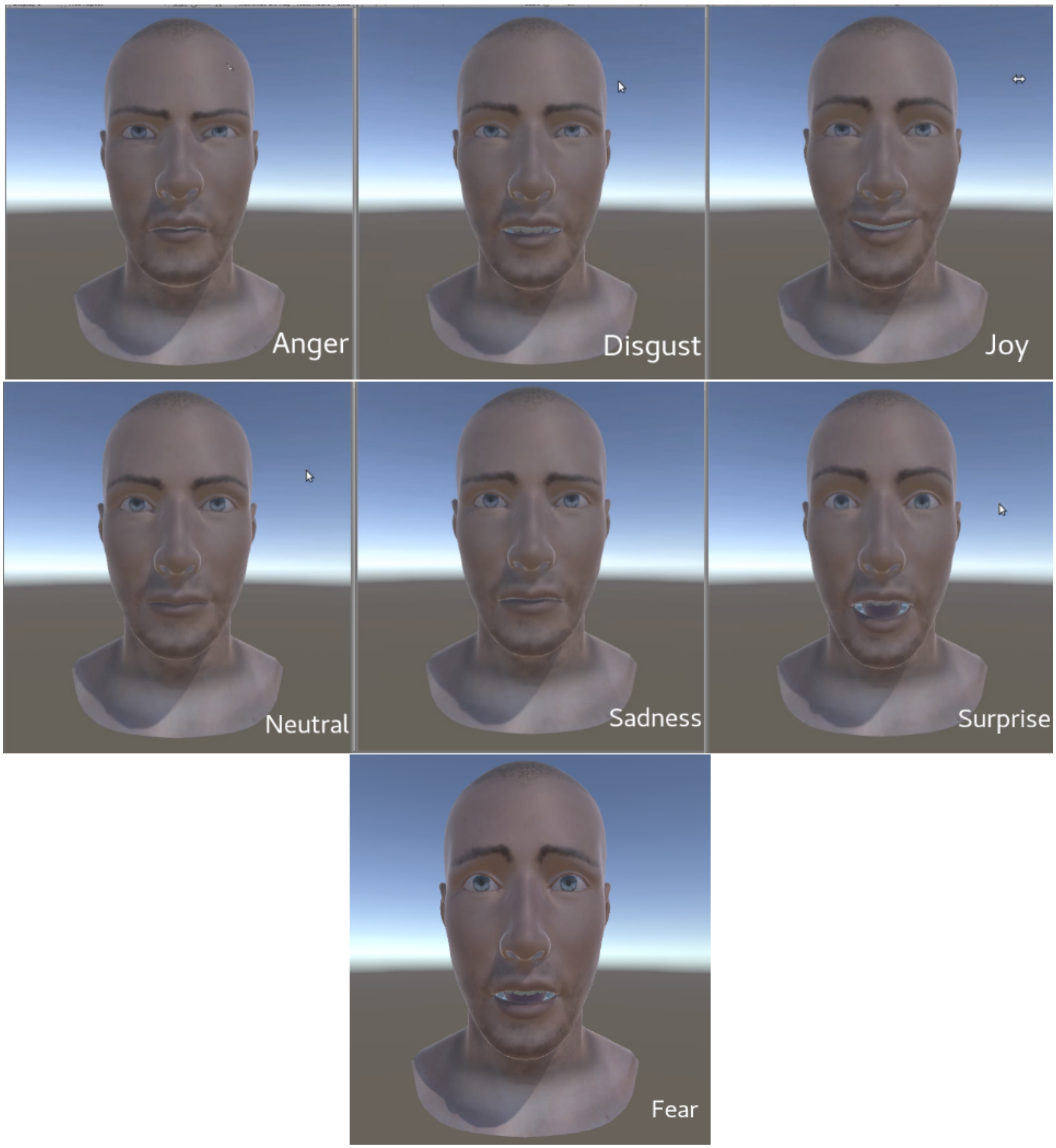

Figure 4: Faces representing the emotions of the renderization module.

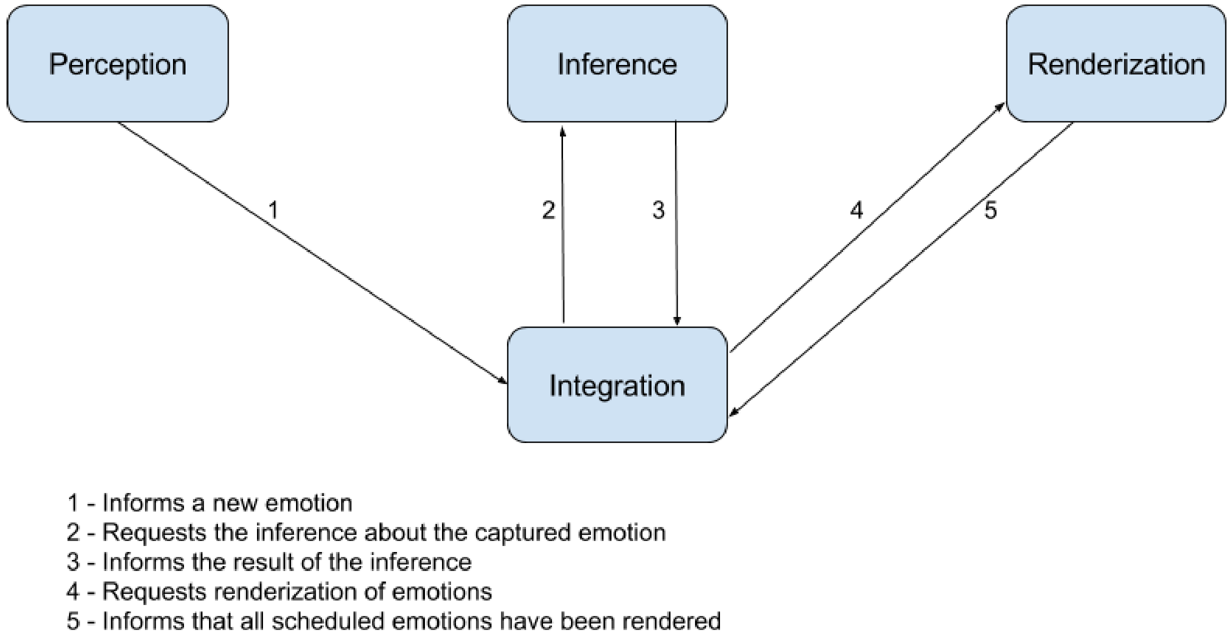

Figure 5: Orchestration of the modules.

In the technological perspective, we used the *Adobe Fuse Character Creator*[2] to render the meshes. This is a software that allows creating avatar's body and its physical characteristics, such as skin and eye color, height, *etc.* In addition, we used *Microsoft Unity framework* to control that mesh and draw the required emotion based on .NET framework.

3.2.4 Integration module

This module was responsible for orchestrating the operation of the others. Figure 5 presents the interactions among the modules. Such interactions are composed of JSON messages exchanged between different modules of the solution. The interaction starts whenever the perception module identifies a user and captures his/her emotion. It sends the recognized emotions to the integration module. Afterwards, the integration module sends the recognized emotion to the inference module, which answers back a set of emotions to be rendered. The integration module sends these answers from the inference to the renderization module, discarding all new incoming emotions from perception module. This module enqueue the received emotions to be rendered. As soon as the renderization module renders all enqueued emotions it informs the integration module which resumes the reception of new emotions from perception module.

3.3 Software Architecture

We aim to further support non-verbal conversations in a way to provide context-awareness in the system based on the interpretation of emotional states from people. For this purpose, we contribute with the definition of a software architecture to incorporate several types of input data and their interpretation with ontologies. The architecture organizes the relevant system components to provide this context (*cf.* Figure 6). This architecture proposal advances the possibilities of development defined in the prototype.

[2]https://www.adobe.com/br/products/fuse.html

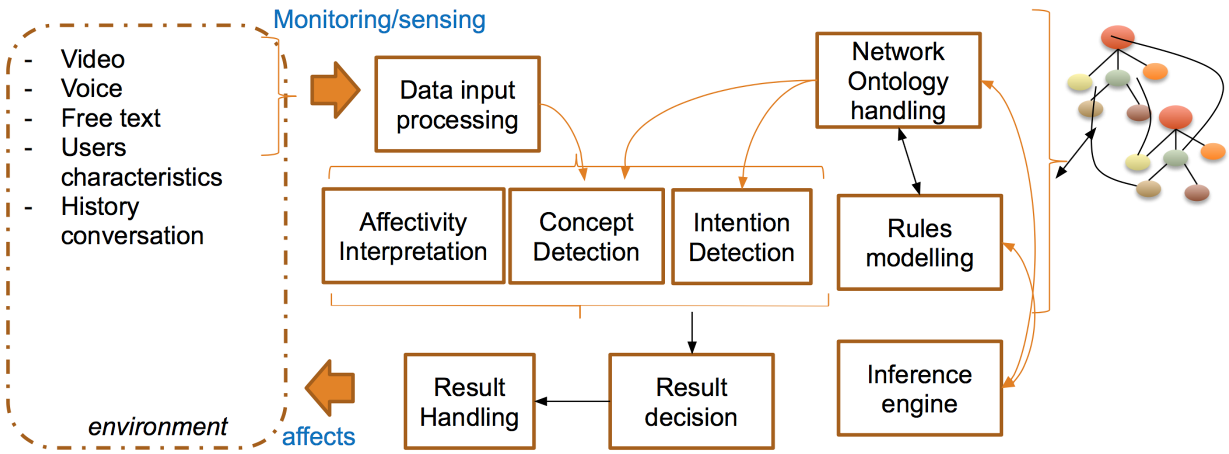

Figure 6: Proposed software architecture.

The goal is to provide means of interpreting data monitored from the environment, based on knowledge represented as ontologies. Also, our goal is to enable means to act properly based on this interpretation. Typical environment data include natural language (NL) free-texts or spoken NL. Our architecture considers various sources as sensory input, including video and images with users' gestures and face expression (as developed in our prototype), as well as users' characteristics and conversation history. Another data source to consider is stream data from sensors. For instance, the presence or the temperature. Body sensors from users can provide data regarding heart rate and skin aspects like sweat. All these data are gathered to provide environmental information.

First, data is pre-processed to enable querying ontologies' classes and instances. At this stage, speech recognition and computer vision features are used to create structured text expressions from voice and video streams. For instance, voices are transformed to structured texts and users' affective expressions are detected by analyzing videos (*cf.* the perception module of the prototype). This information enables other architectures' modules to infer and act upon the situation by means of the interpretation regarding people's aspects. We conceived specific modules to handle the detection of users' emotion states and their intentions (both implicit and explicit based on actions or gestures intended to them).

Pre-processed data generate inputs for software modules to treat specific aspects that will compose the behavior generation. We defined three components with this proposal as follows: Affective Interpretation, Concept Detection and Intent Detection. Affective Interpretation analyzes all the indicators of emotional states presented into the input data. For example, by confronting a detected change of voice tone with ontologies that represent concepts related to emotions. Also, by measuring a change in the heart rate from the user to imply some emotional state.

Concept Detection aims to identify which subjects appear in data input. For example, it implements entity recognition to make explicit the semantic interpretation of the topics under discussion in the conversation. This component is relevant to realize if there is a disinterest of the participants along the interactions and understand the situation.

Intent Detection aims to classify the users' intentions to understand some additional point of view from the interpretation of intentions. For instance, if a face of happiness is the input (in addition to other contextual data), this component detects if it refers to a valuation (*i.e.*, the user judged something) or a wish from the user.

A key aspect of the architecture proposal refers to the use of a network of ontologies to provide the data semantic interpretation. Furthermore, the inference on ontology models is used to decide the actions. This is relevant because it adds semantics on the interpretation of the situation and decision of which action to take next. The ontologies are used by the three components early described. For example, the use of a network of interconnected ontologies might inform the Concept Detection component with the aim of amplifying the system knowledge based on the recognition of similar (equivalent) concepts.

We propose the use of a network of ontologies (Suárez-Figueroa 2010) so that we can interact with several knowledge domains in a variety of ways. In addition, the use of a network of ontologies opens up the possibility of mixing up traditional hierarchical ontologies with the advantages of other types of ontologies like soft ontologies. Soft ontologies are flexible set of meta-data (Kaipainen, Normak, et al. 2008), which are useful to represent dynamically evolving information domains. These ontologies have individual elements associated with values in a non-structured *a priori* hierarchy, defining shared *m-dimensional ontospace*. In our architecture, we propose that soft ontologies might coexist with ontologies that model more rigid hierarchies to represent well-defined domains and concepts.

In addition to a network of ontologies, the process of obtaining decisions is backed up by inference rules and an inference engine (*cf.* the inference module in the prototype). The modeling of rules focuses on the understanding of the interaction effects. The rules play a fundamental role in the decision of the action, as they have their objectives implemented in the rules to provide the succession of actions, based on gathered information.

The inference engine is responsible for interpreting the ontologies' axioms and associated rules, and inferring actions. The *Result Decision* derives the result decision that verifies all types of decision and chooses the best ones. In this step, we can have more than one decision filtered. For example, send an audio response to a participant and also can change the system visual aspect, like send a smile through an avatar. At the final stage, the system must decide how to deliver all the decisions to the environment. We devised the *Result Handling* component to treat it. For instance, it could change its avatar, send a sound, change its voice tone or even send a text message (*cf.* the renderization module is equivalent to this component).

4 Discussion

The conducted literature review detected that the use of ontologies in enactive systems plays a central role, but this aspect requires further research to understand how to enable the coexistence of several types of ontologies underlying the system. In addition, the semantic annotation of media involved in the system and the use of these annotations to influence the system's behaviour (e.g., take an action rather than other) needs further investigations to achieve the required operation of socio-enactive systems.

We constructed an initial version of a proof of concept obtaining an enactive system prototype (the interactive bot). This initial version enables us to conduct experimental evaluations to refine the users' requirements. In addition, this allow us to study the drawbacks and understand the problems of considering the ontologies and their contributions to the development of socio-enactive systems. Our future work will involve an experimental evaluation to investigate the enaction properties from the system and the users' reactions with the interactive bot.

We contributed with a software architecture and infrastructure for the creation of ontology-based enactive systems. We defined an architecture that implements components for monitoring the environment and providing result decisions based on the semantic interpretation. Virtual assistants as interactive bots fail in further exploring other types of media like data sensors, images, audio and video streams to interact with people. In particular, they do not consider neither react to the users' emotional states. Usually, existing solutions only use free-text resources.

Our approach goes towards a new way of building systems that explore new kinds of interaction with virtual agents based on a combination of media. The solution relies on a set of interrelated networked ontologies. In the rule modeling component, we found that it is possible to implement and run SWRL rules that provide means for the system to interpret the input affective data. The system infers the systems' behavior considering affective states as a list of actions, and it determines the systems' actions using rules based on mirror neuron theory.

Our proposal opens directions to various research challenges. Based on the raw input data, we explored computer vision features to detect affective states from people's face. This information needs to be combined with the annotation of semantic concepts and intentions. The difficulties are mostly related to the on the fly mode of detection. The use of intentions, for instance, refers to a central aspect of human communication and it helps the interpretation and understanding of shared information. The detection of declared intentions might help to disambiguate terms according to a delimited context. This can contribute to further create adequate actions in the bot's behavior. Although there are recent investigations in this direction, research issues remain open to the automatic detection of these elements based on an ontology background knowledge.

Ontologies as formal representation of concepts in a knowledge domain that have the potential of share knowledge and increase the learning curve of algorithms. Our ontology-based enactive system relies on a network of ontologies. This network might enable the system to understand data and provide the capability of retrieving useful information for its actions. However, this entails several research challenges.

First, the use of interconnected ontologies requires an semi-automatic alignment process to create correspondences between equivalent concepts from different ontologies. In this network, domain-related ontologies, upper-level ontologies, and models that formally represent the system's actions must coexist. The creation of semantic mappings in this network can be very beneficial, but the management of such heterogeneous artifacts is a difficult problem. This is further aggravated when considering networks that contain soft ontologies as well as rigid and formal structured ontologies.

In this context, the design approach for the ontologies and the engineering process for the construction of ontologies considering the existence of this network demands additional studies. This process requires dealing with domain changes and the adequate update of the ontologies. This implies ontology evolution methods to guide the dynamic aspects of the system. These detected challenges will be further studied in our future work.

Another challenge to be undertook is the specification of SWRL rules in conjunction with soft ontologies and other less structured ontologies. This may require "soft rules" in which logical relations and consequences are not rigidly constructed, but defined by changes in ontological spaces. Fuzzy extensions of SWRL rules (*e.g.*, Pan et al. 2006) can be considered as one alternative to deal with this problem.

Our defined SWRL rule base does not include the modeling of the system's objective neither other constraints that might affect the list of actions, such as known user's personal characteristics, history of emotions, history of conversation, *etc.* However, these additional elements of the rule set are dependent upon each interaction domain. The modeling of the socio-enactive system's intentions/objectives will be further explored in future studies.

5 Conclusion

The concept of enactive systems requires the use of complex knowledge representation models to express the system's behaviour repertoire. This technical report described the initial results of the investigations conducted to understand and experiment the role of ontologies in enactive systems. Our long-term goal is to provide the adequate methodologies and software infrastructures to enable the development of socio-enactive systems based on ontologies. For this purpose, we provided a literature review to understand the existing contributions that have studied the way ontologies are used in enactive systems. This research contributed to an initial proof of concept implementing an enactive system software prototype. Our proposal was based on a network of ontologies to represent the meanings of input data semantics. In particular, we addressed the interpretation of affective states to improve the interaction based on the *Emotion Ontology*. This investigation provided a set of SWRL rules to model the system's reaction behaviors. These elements were organized in a software architecture to enable interpretation of affective expressions, semantic concepts and intentions from user's data input. Future work involves to explore additional ontologies with distinct modeling approaches and study methodologies to adequately refine the set of rules. Furthermore, we aim to develop a complete system and conduct long-term user studies to examine the interaction with a socio-enactive system.

6 Acknowledgements

This work is financially supported by the São Paulo Research Foundation (FAPESP)[3] (grant #2015/16528-0) and by National Council for Scientific and Technological Development (CNPq) (grant #308618/2014-9).

References

Bader, N., Mokryn, O., & Lanir, J. (2017). Exploring emotions in online movie reviews for online browsing. In *Proceedings of the 22nd international conference on intelligent user interfaces companion* (pp. 35–38). IUI '17 Companion. Limassol, Cyprus: ACM.

Borth, D., Chen, T., Ji, R., & Chang, S.-F. (2013). Sentibank: Large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proceedings of the 21st acm international conference on multimedia* (pp. 459–460). MM '13. Barcelona, Spain: ACM.

Borth, D., Ji, R., Chen, T., Breuel, T., & Chang, S.-F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st acm international conference on multimedia* (pp. 223–232). MM '13. Barcelona, Spain: ACM.

Chen, R.-C., Hendry, & Huang, C.-Y. (2016, March 1). A domain ontology in social networks for identifying user interest for personalized recommendations. *Journal of Universal Computer Science*, *22*(3), 319–339. |http://www.jucs.org/jucs_22_3/a_domain_ontology_in—.

Ekman, P. (2007). *Emotions revealed, second edition: Recognizing faces and feelings to improve communication and emotional life*. Holt Paperbacks.

Gil, R., Virgili-Gomá, J., García, R., & Mason, C. (2015). Emotions ontology for collaborative modelling and learning of emotional responses. *Computers in Human Behavior*, *51*, 610–617. Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Networks Era.

[3]The opinions, hypotheses and conclusions or recommendations expressed in this material are the responsibility of the authors and do not necessarily reflect the views of FAPESP.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human Computer Studies*, *43*(5-6), 907–928.

Hastings, J., Brass, A., Caine, C., Jay, C., & Stevens, R. (2014). Evaluating the emotion ontology through use in the self-reporting of emotional responses at an academic conference. *Journal of biomedical semantics*, *5*(1), 38.

Hautamäki, A. (2016). Points of view: A conceptual space approach. *Foundations of Science*, *21*(3), 493–510.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, *34*(4), 575–583.

Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). SWRL: A semantic web rule language combining OWL and RuleML. W3C Member Submission. Available at http://www.w3.org/Submission/SWRL/.

Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M., & Chang, S. (2015). Visual affect around the world: A large-scale multilingual visual sentiment ontology. *CoRR*, *abs/1508.03868*. arXiv: 1508.03868

Kaipainen, M. & Hautamäki, A. (2015). A perspectivist approach to conceptual spaces. In F. Zenker & P. Gärdenfors (Eds.), *Applications of conceptual spaces: The case for geometric knowledge representation* (pp. 245–258). Cham: Springer International Publishing.

Kaipainen, M., Normak, P., Niglas, K., Kippar, J., & Laanpere, M. (2008). Soft ontologies, spatial representations and multi-perspective explorability. *Expert Systems*, *25*(5), 474–483.

Kaipainen, M., Ravaja, N., Tikka, P., Vuori, R., Pugliese, R., Rapino, M., & Takala, T. (2011). Enactive systems and enactive media: Embodied human-machine coupling beyond interfaces. *Leonardo*, *44*(5), 433–438.

López-Gil, J.-M., Gil, R., & García, R. (2016). Web ontologies to categorialy structure reality: Representations of human emotional, cognitive, and motivational processes. *Frontiers in Psychology*, *7*, 551.

Meacham, D. (2016). How low can you go? bioenactivism, cognitive biology and umwelt ontology. *Humana.Mente Journal of Philosophical Studies*, *31*, 73–95.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLOS Medicine*, *6*(7), 1–6.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2010). Emoheart: Conveying emotions in second life based on affect sensing from text. *Advances in Human-Computer Interaction*, *2010*, 1.

Pan, J. Z., Stoilos, G., Stamou, G., Tzouvaras, V., & Horrocks, I. (2006). F-swrl: A fuzzy extension of swrl. In S. Spaccapietra, K. Aberer, & P. Cudré-Mauroux (Eds.), *Journal on data semantics vi* (pp. 28–46). Berlin, Heidelberg: Springer Berlin Heidelberg.

Phan, N., Dou, D., Wang, H., Kil, D., & Piniewski, B. (2017). Ontology-based deep learning for human behavior prediction with explanations in health social networks. *Information Sciences*, *384*, 298–313.

Preston, S. D. & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and brain sciences*, *25*(1), 1–20.

Pugliese, R., Tikka, P., & Kaipainen, M. (2014). Navigating story ontospace : Perspective-relative drive and combinatory montage of cinematic content. In *Expanding practices in audiovisual narrative :* (pp. 105–128).

Suárez-Figueroa, M. C. (2010). *Neon methodology for building ontology networks: Specification, scheduling and reuse*, Informatica. Ontology Engineering Group.

Thakor, P. & Sasi, S. (2015). Ontology-based sentiment analysis process for social media content. *Procedia Computer Science*, *53*, 199–207. INNS Conference on Big Data 2015 Program San Francisco, CA, USA 8-10 August 2015.

Villalonga, C., Razzaq, M. A., Khan, W. A., Pomares, H., Rojas, I., Lee, S., & Banos, O. (2016). Ontology-based high-level context inference for human behavior identification. *Sensors*, *16*(10), 1617.