# A Smart Dialogue-competent Monitoring Framework Supporting People in Rehabilitation

Thanassis Mavropoulos
Centre for Research and Technology
Thessaloniki, Greece
mavrathan@iti.gr

Georgios Meditskos
Centre for Research and Technology
Thessaloniki, Greece
gmedisk@iti.gr

Eleni Kamateri
Centre for Research and Technology
Thessaloniki, Greece
ekamater@iti.gr

Spyridon Symeonidis
Centre for Research and Technology
Thessaloniki, Greece
spyridons@iti.gr

Dimitris Tzimikas
Entranet Ltd.
Thessaloniki, Greece
tzimikas@entranet.gr

Lefteris Papageorgiou
Entranet Ltd.
Thessaloniki, Greece
papageorgiou @entranet.gr

Christos Eleftheriadis
R&D department, ELBIS P.C.
Thessaloniki, Greece
ce@elbis.gr

George Adamopoulos
Evexia S.A.
Thessaloniki, Greece
gm@evexia.gr

Stefanos Vrochidis
Centre for Research and Technology
Thessaloniki, Greece
stefanos@iti.gr

Ioannis Kompatsiaris
Centre for Research and Technology
Thessaloniki, Greece
ikom@iti.gr

## ABSTRACT

In this paper, we present work in progress on the development of a smart monitoring framework to support people with motor disabilities and their caregivers in clinical and non-clinical rehabilitation and care environments. The innovation of the platforms lies in the combination of smart monitoring solutions, such as activity recognition and lifestyle tracking, with an intelligent virtual agent that aims to empower and motivate people in need through personalized feedback and responses, as well as to assist caregivers and clinicians to easily collect information about the patients. The proposed system exploits and combines state-of-the-art technologies in speech recognition and synthesis, knowledge representation and reasoning, dialogue management and sensor data analysis, infusing clinical knowledge and patient history. Aiming for a practical, acceptable solution, the proposed system takes into account aspects of integration, security and privacy.

## CCS CONCEPTS

• Information systems~Ontologies • Computing methodologies~Natural language processing • Computing methodologies~Intelligent agents • Hardware~Sensors and actuators

## KEYWORDS

Ontology-based dialogue management, virtual assistant, sensors, NLP, semantics

**ACM Reference format:**

Thanassis Mavropoulos, Georgios Meditskos, Eleni Kamateri, et. al. 2019. A Smart Dialogue-competent Monitoring Framework Supporting People in Rehabilitation. In *Proceedings of The 12th PErvasive Technologies Related to Assistive Environments Conference (PETRA'19). ACM, Rhodes, Greece, June 5-7, 2019, 10 pages.*

## 1 INTRODUCTION

Care and rehabilitation is an important factor in maintaining and improving the quality of life of people with physical or mental injuries. At the same time, it has important short-term and long-term financial implications for both individuals and their families as well as for the health system and society. Assisted living technologies combined with intelligent interfaces and advanced dialogue systems can provide supportive technologies in homes and clinics, significantly improving the quality of life of individuals [15]. Examples are the use of multimodal sensors to monitor biometric indications and behaviors (like heart pressure and heartbeats), as well as user interactions with spoken agents acting as personal assistants, reminding exercises, drugs, etc. Although these technologies are quite popular and mature to a good extent, making full use of their potential in real environments for

monitoring, rehabilitation and caring for people with physical or mental injuries poses significant challenges. Thus, while at a global level a number of software approaches have been developed with the individual user's needs in mind that also possess advanced interaction capabilities, they are still not offering a complete and specialized solution.

With the purpose of upgrading the monitoring and support technologies currently used in clinical and non-clinical rehabilitation and patient care environments, we aim at the development of an innovative, interconnected virtual agent-centric platform with enhanced interactive, conversational and cognitive skills through natural, multimodal human-machine interaction. The platform is being developed in the context of the research project REA[1], which implements an innovative combination of advanced speech recognition, speech synthesis and user interaction technologies, with a platform for collecting, analyzing and extracting conclusions from sensor data, clinical knowledge and patient history. In particular, the system is able to: a) comprehend user needs by leveraging contextual information made available through the analysis of verbal communication of users with the system and sensor-originating data analysis, b) communicate with the user via verbal and non-verbal means, c) update the situational and conversational picture of the system, d) retrieve information from the web to satisfy the user's information requests, and e) converse with the user using ontology-based reasoning techniques. Therefore, our contribution lies mainly in the exploitation of multimodal sensor data which are handled by a sophisticated platform with conversational capabilities; the system's ability to manage a high number of fluid and natural human-agent interaction, coupled with a unique alarm-triggering, sensor-driven monitoring system offers quality of life functionalities that are currently lacking by existing competing solutions.

The main objective for the REA platform is to be integrated in: a) hospitals, rehabilitation centers and clinics, providing upgraded and innovative patient monitoring and support services in real time, through correlation of interlinked information to manage comprehensive, accurate and personalized care-related decisions, b) non-hospital environments, acting as a personal interactive assistant providing innovative voice receiving information services on issues related to the daily lives of patients (diet and medication), remote monitoring services, and initializing custom and personalized dialogues with a positive impact to their mental health and satisfaction. To this end, research revolves around four main pillars:

- Development of an efficient infrastructure for collecting multimodal sensor data, such as physiological and verbal communication data
- Implementation of innovative solutions for data analysis (verbal and non-verbal) to monitor the behavior and condition of individuals
- Application of intelligent techniques for combining multimodal data with clinical knowledge, behavioral patterns, medical history, and conversation history supporting the planning of timely and personalized interventions
- Development of a smart, virtual and interactive agent for interacting with the users and providing support to patients, caregivers and staff.

In the rest of the paper we describe the technologies that underpin each pillar, as well as the approach followed so far to address the research challenges. More specifically, after an overview of the related work relevant to the main technological and research objectives, we present in Section 3 an overview of the system architecture. In section 4, we present the data collection framework we have developed for sensor integration and data collection. Sections 5 and 6 describe the frameworks for language understanding and speech synthesis, as well as the ontology-based framework for intelligently fusing data and information to achieve context awareness. Section 7 illustrates the dialogue capabilities of the framework, while section 8 presents real-world example of the testing of the framework in realistic environment. Finally, Section 9 concludes our work and presents future direction of our research.

## 2 RELATED WORK

Speech recognition has been in the limelight since the 1990s and big companies presented relevant projects like the Sphinx-II and VAL (Voice Activated Link) legacy systems. These provided solid groundwork for modern voice-enabled and voice-first technologies [4], which ultimately led to the development of virtual agents with dialogue capabilities. Consequently, industry-leading companies like Google, Amazon, Baidu, IBM Watson, Wit, Microsoft have developed speech recognition APIs which can be integrated seamlessly with user frameworks and provide state-of-the-art results. The task of composing text or speech from non-linguistic input is called Language Generation (LG) and has been in the forefront of computational linguists and natural language processing (NLP) scientists for the past twenty years [25], while recent advances ([9], [21]) seem to focus mostly on DNNs (Deep Neural Networks). Moreover, there have been various efforts towards speech generation and the most popular ones include HMM-based speech synthesis models [29], specialized incremental dialogue systems [12], and DNN-based systems [34]. State-of-the-art approaches for text-to-speech leverage various forms of DNNs; in [20] a system is presented that builds models based on waveforms, in [28] the models are based on spectrograms, while in [23] hierarchical cascading is used to improve an LSTM network. Furthermore, day-to-day voice enabled products like Apple's Siri, Amazon's Alexa/Echo, Microsoft's Cortana and Google's Assistant are joined by more healthcare oriented agents like Aiva[2], Merit[3], Suki[4] and Robin[5].

A major advantage of the REA system is the inherent ability to leverage sensor data in order to ameliorate a patient's quality of life. Moreover, by analyzing certain data it can reach conclusions over the patient's posture or sleep quality. Falls pose a very serious

[1] https://rea-project.gr/en/home-en/
[2] https://aivahealth.com/
[3] https://merit.ai/
[4] https://www.suki.ai/
[5] https://www.robinhealthcare.com/

threat among older people and people with mobility issues. As the years pass and the proportion of the general population of the elderly is increasing, the need to find reliable ways to detect falls becomes even more relevant. Fall detection is a complicated topic because of the fact that a measuring system needs to calculate various parameters for an alarm to go off when a person falls on the ground. Many researchers work on this complicated matter and various methods are being proposed, each with its own shortcomings. Most approaches rely on the use of accelerometers and gyroscopic sensors [19], while some approaches exploit wireless signals for the detection [31]. Computer vision techniques are very popular [26], while ambient sensors are also being used in certain cases [35].

Studies of sleep in daily life have shown that there is direct relation between a person's health/well-being with the quality and quantity of everyday sleep [22]. Nowadays many approaches that monitor a patient's sleep use self-reports of the person under evaluation, while others require the presence of sensors connected to the person under evaluation [11].

An integral component of any conversational agent is the module that handles the human-machine interaction. Several approaches have been proposed to formalize the dialogue management. Common to all approaches is a) the representation of the system's knowledge of the current situation in a data structure called the *dialogue state*, and b) the use of a decision mechanism to select the action to perform in each dialogue state [14]. Several strategies have been proposed to represent, update and act upon the dialogue state including handcrafted and statistical approaches.

The incoming information (verbal and non-verbal) is semantically coupled and interlined with semantic knowledge structures by means of an ontological framework. The Web Ontology language (OWL/OWL 2) [10] is a knowledge representation language widely used within the Semantic Web community for creating ontologies. The design and semantics of OWL 2 have been strongly influenced by Description Logics (DL) [3]. Some basic notions are: a) axioms, the basic statements that an OWL ontology expresses, b) entities, elements used to refer to real-world objects, and c) expressions, combinations of entities to form complex descriptions. Ontology languages, such as OWL 2, share a common understanding of the structure and semantics of information, enabling knowledge reuse and inferencing. Capitalizing on the expressivity of the models, several approaches define one or more interpretation layers in order to elicit an understanding of the context. For example, challenges and opportunities in applying Semantic Web technologies in context-aware pervasive applications are discussed in [33]. In the domain of natural language interfaces, ontologies provide the vocabulary and semantics for content disambiguation [6]. In the domain of multimodal fusion, ontologies are used to fuse multi-level contextual information [7].

In literature, there are only few recent studies presenting personalized ontology-based dialogue agents with domain-specific and social competencies. For example, Altinok, D. [1] suggests an ontology-based dialogue manager (OntoDM) that keeps the conversation memory, provides a basis for anaphora resolution and drives the conversation via domain ontologies, while Wessel et al. [32] suggest an ontology-based dialogue manager that employs ontologies, reasoning, and ontology-based rules for a) domain model representation and reasoning, b) dialogue representation and state tracking, and c) response generation.

Ontology-based dialogue management is conceived as a rule-based approach which can bring significant advantages to the dialogue process. It enables access to a long dialogue history and domain model which is not limited to current and previous states. Ontology-based rules can also be used to compute potential system responses and implement dialogue strategies.

# 3 OVERVIEW OF SYSTEM ARCHITECTURE

## 3.1 System Overview

The proposed system follows a multidisciplinary approach to integrate and bring into effect advances in monitoring technologies and clinical expert knowledge. More specifically, the system employs a synergy of the latest advances in sensor technologies, physical device interfacing and data collection, data processing, knowledge representation and personalized end-user feedback.

**Physical device interfacing and data collection**: The Message Queuing Telemetry Transport (MQTT) broker of Amazon Web Services (AWS) is used, to achieve fast and secure asynchronous communication between the data collection servers and devices. Due to the asynchronous nature of MQTT, there is a need for a worker process to receive, store and act upon the data received from the devices. Access to the data is provided through a Web API, so that they can be consumed by the other systems (Service Oriented Architecture). The Web API also provides endpoints to configure and send commands to the devices (which transparently to the API consumer, are sent through the MQTT broker).

**Data processing**: The system currently integrates a wide selection of proprietary, ambient or wearable devices, originally intended for lifestyle monitoring, repurposed to a medical context (see next session for the complete set of sensors used). Each device is integrated by using dedicated modules that wrap their respective API, retrieve data and process them accordingly to generate atomic events from sensor observations e.g. through aggregation. In the case of image data, computer vision techniques are employed to extract events, such as recognizing falls. Sleep is also an important monitoring parameter with clinical value. To evaluate whether an individual has had a good sleep we need to measure the times the person has moved during the sleeping session, as well as the body posture he/she was in. Clinical evaluation of sleep interruptions and probable sleeping patterns will determine the threshold of what is considered a "good" sleep. Based on these heuristics a sleep assessment will be made available to the knowledge base that could trigger a system notification. Special high priority tasks running by the Real Time Operating System (RTOS) firmware implementation monitors these values and when conditions are met, alerts are generated.

**Human-Machine interfacing**: As described in section 1, the virtual assistant interacts with the users via verbal and non-verbal means. 3D character models are used to ensure natural system communication. The virtual agent supports speech synthesis techniques, transforming the results of dialogue management and responses into speech (ensuring Lip Speech synchronization / LipSync), as well as embracing sophisticated express feedback techniques, e.g. head movement, blinking eyes, facial expressions.

## 3.2 System Design and Integration

*3.2.1 System conceptualization*. The framework follows a 3-tier architecture (Figure 1), the sensors management layer, the communication understanding layer and the communication analysis layer. The communication understanding layer acts as a central hub that manages the communication between the system and the user as well as the interaction between the other two layer modules, as all information is transmitted through it.

The Sensor management layer has undertaken the task to harvest and provide user information originating from sensors installed in the area. Sensor data are either provided on demand or sent via an alert mechanism when an immediate proactive action is triggered by the system. Furthermore, it supports commands that instruct the sensors to perform some kind of movement (e.g. change the tilt angle of the bed).

The user always comes in direct contact only with the *communication understanding* layer which, as the case may be, communicates with one of the others. This layer has the responsibility of converting the user's speech to text and vice-versa as well as authenticating users that are authorized with more rights as regards to the functionality of the system (mainly clinical personnel).

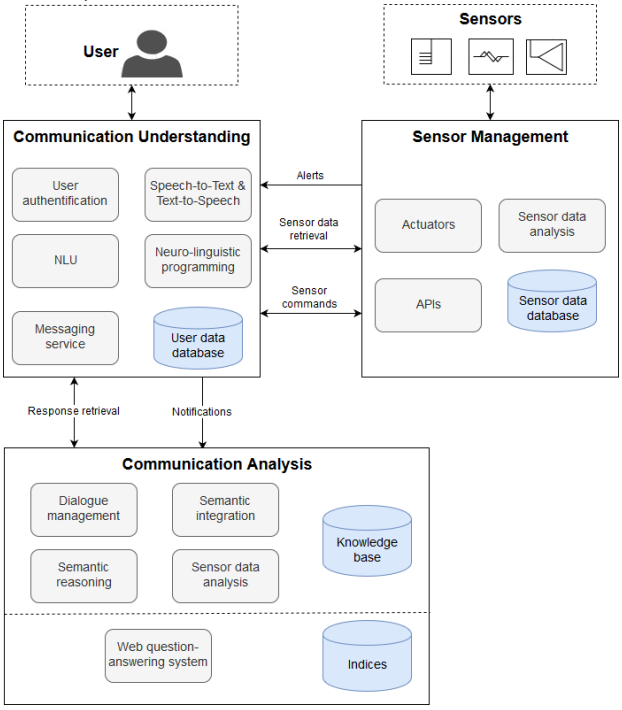


**Figure 1 System architecture**

The *communication analysis* layer has been in charge of representing the current contextual information, correlates it with domain- and dialogue-specific knowledge, managing the dialogue and selecting the appropriate response to the user according to the question he has asked. Depending on the type of the question, the Dialogue Manager (DM) accesses the Knowledge Base (KB) semantic representations or redirects the question to the web-based question-answering system to retrieve the answer through its indices. In very special cases, where the *communication understanding* layer can provide a direct answer without requiring a response from the DM, the *communication understanding* layer performs the direct system interaction with the user and sends a notification to the DM in order to keep it updated about the dialogue history.

*3.2.2 Sensor development*. The development of customized sensor systems that will be an integral part of the REA platform will be materialized by enhancing existing commercial sensor chips and devices via extra module installations and firmware upgrades. Specifically, a MicroController Unit (MCU) will be integrated in the devices that has the necessary interfaces and wireless connection capabilities to handle measurements of the desired value and send those to the cloud. Necessary firmware will be developed that will sample the values in programmable intervals of the sensor chips and devices according to the requested specifications. The skeleton of the firmware implementation will be a (RTOS). This approach will ensure the correct execution of the operational flow that the sensor device needs to follow in order to take the measurement, format it and send it to cloud. Since a RTOS is designed to process simultaneously many tasks, it is guaranteed that the abovementioned procedures will be executed swiftly.

*3.2.3 System integration*. For the system integration phase the plan is to install the implemented sensors and devices at the patients' living space, both in hospital and non-hospital environments. A familiarization phase patient-wise on the use of the voice controlled bed and other sensors is also planned.

Most sensors will be common between the tested environments, with home installations potentially being more feature-rich (e.g. door sensors for crowd detection), since these are not bound by GDPR issues. In such environments, we will be able to install sensors that check the presence of a person in a room and the duration of the stay in that room, giving an alert if the person is still standing or has fallen.

# 4 MONITORING AND DATA COLLECTION

REA implements an approach to integrate various sensors, in the sense of plug-in modules. Each device is manipulated by the system through a dedicated module that conforms to its communication protocol and data format. After raw sensor data are retrieved, processing and analysis commence on the processing layer of the framework, as described on the next section. Currently, the platform offers a rich selection of sensors that can be either ambient (e.g. a camera, mic), or wearable. The data types considered are self-contained measurements, such as physical

activity, or require further analysis to derive meaningful information, such as accelerometer movement.

In order to address the requirements of the platform (e.g. regarding data collection) and the users' needs, already existing commercial Internet of Things (IoT) systems (sensors and devices) will be integrated, their usage will be enhanced, while new sensors will be developed where necessary. More specifically, the following sensors are being developed and will be integrated in the platform:

- Blood pressure sensor: The blood pressure sensor will consist of a commercially available non-networking blood-pressure system upon which wireless network connectivity will be built.
- Blood sugar sensor: Following the same principal as the blood-pressure sensor, the blood sugar sensor will be a commercial device that will be enhanced with WiFi connectivity.
- Sleep quality sensor: The system will consist of a sensor-embedded sheet placed underneath the patient's bed. Data that have been collected through the night will be sent via a WiFi connection. To measure the patients sleep quality, the times that the individual has woken up, the body posture and other valuable parameters will be calculated.
- Wearable sensor: This device will measure the following patient biometric values while he is active: i) *temperature*, ii) *blood oxygen*, and iii) *heart rate*, while also implementing a *fall sensor* to detect whether the patient has had an accident and fell to the floor. The wearable device will transmit collected data through a Bluetooth Low Energy (BLE) gateway that will forward them to AWS through a WiFi connection.
- Camera: A 3D camera recording images and depth data online, using its open SDK and a USB interface. The data will be exploited by existing computer vision algorithms used for localization of subjects and activity recognition [2].
- Bed management system: A system that will control sections of the patient's bed by voice commands helping make patients' life better. Gyroscopic sensors will be installed on patients' beds that will assist the latter in managing the tilt angle of their back and feet via the voice controlled system; the sensors' values will provide the necessary feedback to the module that is responsible for managing the positioning of the back- and feet-section of the bed.

Each IoT sensor will have its own Universally Unique IDentifier (UUID), which will be defined when creating the devices' certificates at the AWS certificate creation environment. The certificates are used to securely connect to AWS MQTT brokers. The network connection will be encrypted using these certificates through a Transport Layer Security (TLS) connection. The actual data as a payload of the MQTT message follows the Sensor Measurement Lists (SenML) IETF standard RFC8428[6].

The data collection infrastructure is hosted on AWS. The AWS IoT service is a central part of the system, providing us with a MQTT broker for the communication with the sensors and also with a fully controlled public key infrastructure for securing the communication of the sensors. All MQTT messages from the sensors are handled by AWS Lambda functions and passed to a task engine for further processing. This processing might include storage in the database or notification to other services. All data stored in the database is available for use via a REST API.

## 4.1 Data Protection and privacy

All necessary technical and organizational measures have already been taken to protect the personal data of the natural persons involved, especially patients, escorts, medical and nursing staff. We have taken to the best of our knowledge all relevant actions in order to fully harmonize with the provisions of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27th of April 2016 on "the protection of individuals with regard to the processing of personal data and on the free movement of such data and the repeal of Directive 95/46/EC" (General Data Protection Regulation) and any other relevant national legislation. In particular, all necessary and appropriate security measures will be taken concerning the installation and use of the video surveillance system (comprising of conventional and thermal cameras), all relevant procedures towards storage times of data will be respected, while we'll ensure that consent is received promptly for all persons involved and, in general, emphasis will be placed on the respect of personal data and the harmonization of existing legislation, European and national. The latter legislation lays down, inter alia, the principles of data processing that need to be respected, the individual obligations of companies processing personal data, the rights of individuals whose data are processed, as well as the ways in which such rights are exercised.

More specifically, for data protection it is necessary to activate SSL for the security of the personal data. Furthermore, backups will be encrypted with special algorithms, while Data Centers will be controlled by all relevant stakeholders to achieve ISO 27001, PCI DSS, etc., which assure that the infrastructure, data handling and security meet the highest standards. On sensor level, data protection is handled via the use of UUID to ensure the anonymity of the patients' data and sensor devices and by using TLS data connection from the device to AWS cloud.

# 5 LANGUAGE UNDERSTANDING AND SPEECH SYNTHESIS

Speech recognition platforms, even modern ones, often get confused by regional accents and speech variations, while background noise can be difficult to manage as well. Moreover, multiple-user voice input can perplex things even further. As it is evident, in order for a system to be effective, it needs to be competent enough to distinguish between homophones, between proper names and common words and be able to surmount many more obstacles. According to the above, a list of speech recognition [30] and conversion technologies taken into consideration for use in REA are Acapela, Google and Innoetics, while in text-to-speech the candidate technologies include the Nuance, CMUSphinx and Snips suites. A deciding factor and key feature (in both speech

[6] https://tools.ietf.org/html/rfc8428

recognition and speech synthesis) even from the first phase of the system's implementation is native support for Greek and English. Another important factor is also the cost of some of these services as they require subscriptions with a cost of billing per number of words, which in the early stages of the program is a deterring factor. An additional characteristic taken into account is of course speech and voice recognition quality, as well as the level of effectiveness of the outcome result. Ultimately, the most influential criterion was support for the Greek language, which has greatly restricted the choices and has led us to choose the Google system for voice recognition and Nuance for text-to-speech since these fulfilled all the conditions that were set beforehand.

**Language understanding:** In REA, we use the Google Cloud Speech-to-Text API, enriching the results with NLP disciplines and semantic disambiguation techniques. The recognition of named entities, concepts and relations is a crucial part for verbal input understanding. The linguistic analysis involves both a syntactic and a semantic analysis, relying on tools such as part-of-speech taggers and parsers. REA implements a module for generating a dependency parse tree for the input question, based on Stanford CoreNLP [16] that identifies, among others, words, their parts of speech, whether they are names of companies, people, etc., normalize dates, times, syntactic dependencies, relations between entity mentions, etc. The results of CoreNLP are further enriched with a rule-based module [17], implementing a set of concept, domain-dependent named entity and relation extraction rules.

All entities, concepts and relations extracted need to be correctly linked to the ontologies (see next section) in order to understand the contexts of users' input and to successfully address their information needs. A semantic entity linker is implemented that is responsible for context-based disambiguation, linking natural language words and phrases to RDF/OWL ontologies, lexical databases and semantic networks (e.g. WordNet, BabelNet, ConceptNet). To this end, REA integrates and combines existing domain independent tools for word sense and context disambiguation, such as DBpedia Spotlight[7], Babelfy[8] and FRED[9], enriched with domain-dependent linking and alignment to ontologies. As a result, language understanding actually transforms user input into ontologies that capture the context and semantics of the user information needs.

**Speech synthesis** [13]: REA's text-to-speech capabilities are based solely on Nuance Vocalizer. Nuance uses large amounts of speech data to train deep learning models that learn the relations between written text and the corresponding voice characteristics, while also exploiting the utterances' context to ascertain proper intonation and expressiveness. Hence, the trained models mirror any human voice and sound more natural than the best existing Text-to-Speech frameworks.

# 6 KNOWLEDGE REPRESENTATION AND REASONING

## 6.1 Multimodal Knowledge Representation

Given the inherent requirements in multimodal environments to aggregate low-level information and integrate domain knowledge, we have used the OWL 2 ontology language [10] to capture context elements (e.g. profiles, events, activities, locations) and their pertinent relations, mapping observations and domain knowledge to class and property assertions in the DL [3] theory, fostering integration of information at various levels of abstraction and completeness. The generated models encapsulate formal and expressive semantics, harvesting several benefits brought by ontologies, e.g. modelling of complex logical relations, sharing information from heterogeneous sources, sound and complete reasoning engines.

As already described, contextual information may be collected from a variety of sources, such as ambient and wearable sensors, text analysis (verbal communication), video analysis, etc. All this information needs to be mapped on domain entities to enable the derivation of contextual descriptors that best satisfy and interpret the context. Figure 2 depicts the lightweight vocabulary we use for modelling context types. The ontology extends the leo:Event concept of LODE [27] to benefit from existing vocabularies to describe events and observations. Property assertions about the temporal extension of the observations and the agent (actor) are allowed, reusing core properties of LODE.

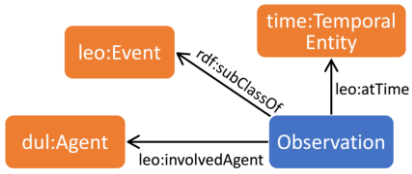


**Figure 2 Upper-level domain ontologies for observations**

Apart from observations, the platform supports domain models that capture various types of background knowledge, including user profile ontologies, ontologies for modeling routines, habits and behavioral aspects. To this end, we reuse modelling patterns described in smart homes [18], adapting the Descriptions and Situations pattern of the DolceUltralite ontology [8] to the domain of rehabilitation, e.g. for modelling physiotherapy exercises. The upper-level pattern is depicted in Figure 3. For example, behavioral aspects are captured as instances of the Aspect class, while detailed information about the aspect is captured by one or more instances of the View class.

[7] https://github.com/dbpedia-spotlight/dbpedia-spotlight
[8] http://babelfy.org/
[9] http://wit.istc.cnr.it/stlab-tools/fred/

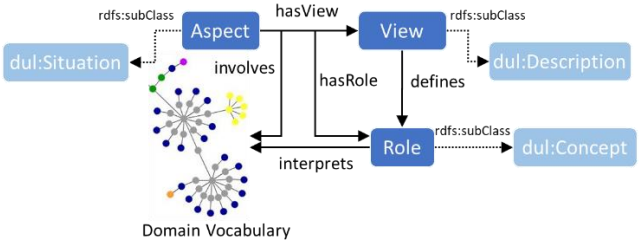


**Figure 3 Upper-level pattern for modelling complex situations**

## 6.2 Data Analytics

This layer hosts a set of modules for sensor data processing. This involves:

- Wearable sensor data analysis that provides valuable information to verify and foster high-level event detection. Such sensor data include physical activity monitoring, measured as moving intensity, stress level, various measurements regarding sleep and lifestyle monitoring. Each of the above modalities and sensors (presented in Section 4) is provided by a different processing component, dedicated to sensors and data formats in REA. For example, accelerometer movement in the three-dimensional space and skin conductance are transformed into physical activity and stress level, respectively. To do so, the library employs internal signal filtering techniques and establishes per-individual statistical baselines for those measurements, using Random forests learning [5]. Based on that, the framework extracts moving intensity and stress level in a range of zero to five.
- Human Activity Recognition that employs a set of existing computer vision techniques described in [2]. These techniques are optimized for images from ambient spaces such as those provided by the IP and depth cameras. The output is intermediate activities, such as sitting, walking, bending, etc. to higher level ones, such as preparing a meal, eating and washing the dishes.

## 6.3 Reasoning

Reasoning is performed in order to aggregate the available information and achieve context awareness. OWL 2 inherits the powerful DL reasoning services, for which efficient, sound and complete reasoning algorithms are available. For example, through subsumption, one can derive implicit taxonomic relations among concepts. Satisfiability and consistency checking are useful to determine whether a knowledge base is meaningful at all.

One of the key reasoning services supported by the platform is the interpretation of utterances in order to feed Dialogue Management with an understanding of the conversation. The supported situations of interests (e.g. from a clinical perspective) are modeled in terms of a DL theory that defines dependencies among low-level observations and high-level conceptualizations. For example, the recognition of situations when the user is complaining about pain is modelled in DL as:

$$PainContext \equiv VerbalContext \sqcap \exists contains.HurtReference$$

VerbalContext is the set of ontological concepts that have been recognized by language analysis. The DL complex class description defines that the HurtReference should be part of the verbal context in order for the current context to be classified as PainContext. DL allows for the further specialization of contextual elements through hierarchies, for example, $Headache \sqsubseteq PainContext \sqcap HeadReference$, enabling the definition of complex domain models that capture the semantics of the domain, as well as user and clinical requirements and needs. The results of classification are then forwarded to the Dialogue Management in order to be appropriately used to further drive the interaction with the user.

# 7 DIALOGUE MANAGEMENT AND WEB-BASED RETRIEVAL

As stated earlier, our proposed approach leverages existing ontologies and reasoning techniques to access a rich domain model, and inform the dialogue manager about domain insights dynamically. The followed approach also offers the benefit to include new information and answer unforeseen questions that cannot be interpreted against system's cognitive models. In this case, the system perceives that the question might have an informational perspective and searches trusted websites for relevant information in order to provide a suitable answer. The available responses will be evaluated by the dialogue manager according to the adopted dialogue strategy to decide about the final system's response.

## 7.1 Decision Making

The system aims to naturally interact with the user through complex dialogues and discussions. To this end, ontology-based dialogue management techniques are implemented, offering increased user comprehension, adaptability and guidance.

In addition, it allows the management of interactive phenomena such as hesitations and pauses. Dialogue management techniques will be based on the existing infrastructure developed in the H2020 KRISTINA [24] project and will be extended to meet the specific requirements of a personalized dialogue management system with domain-specific and social competencies.

More specifically, this component aims at developing management techniques for conducting a natural and flexible human-system dialogue. The strategies will take into account a) the available responses derived either from the semantic module or the web-based question answering module, ; b) the user profile and specific requirements aligned to the profile, c) the dialogue history; and d) the current context which is communicated through the last user's utterance. Based on these data, the dialogue management will conclude on a set of available actions to perform in the next system move. These might include motivational and interventional actions (such as notifications, alerts and reminders). In addition, the

dialogue management implements ontology-based rules to infer unconnected actions with the current state based on the dialogue history. For example, if a user at home surroundings asks many times about his workout plan, the system can provide complementary this information when the user asks about his diet. The final decision of the most appropriate response will be performed according to system's priorities (dialogue strategy).

## 7.2 Multimedia retrieval from the web

In order to enhance the background knowledge of the virtual agent and to serve the cases when the user poses a question to the agent that has a generic answer existing on Internet resources, a question answering (QA) system utilizing this type of content is going to be implemented. The web resources that we are going to focus are trusted health-related websites, websites from newspapers that continuously publish the latest news, websites that provide weather predictions and resources that inform the users of notable upcoming events.

To this end, a framework has been integrated that collects and stores data from the web in structures that will facilitate the efficient retrieval in terms of response time and quality. This mechanism handles and processes information in various modalities, such as text, images and videos. Data collection is executed using web searching, crawling and scraping techniques, while for the data storage, indexing methods are implemented. In the retrieval phase, the QA system will be able to handle natural language queries which will be parsed appropriately in order to optimize the retrieved response, for example by extending it with additional terms using machine leaning techniques. The expected response will be the exact multimedia content or passages from the web that answer the question given by the user.

An overview of the multimedia retrieval framework, including all the underlying preprocessing steps, is illustrated in Figure 4.

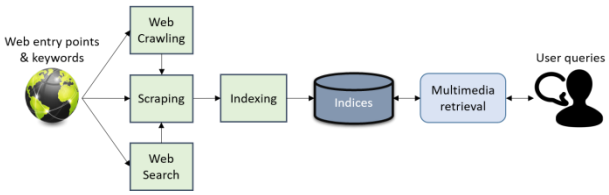


**Figure 4 Multimedia retrieval from web resources**

# 8 USE CASES AND EXAMPLES

This section introduces the environments where the system will be installed and the pilot tests that will be conducted. Then, reference is made to a usage scenario to better pinpoint the agent's advanced functions.

## 8.1 Testing environments

The testing phase will include hospital and non-hospital surroundings. The hospital environment involves the "EVEXIA"[10] rehabilitation center in Northern Greece, which operates an enhanced care unit with 165 beds. In its premises one can find single and double patient rooms, group treatment rooms, as well as head nurse offices, nurse station, examination rooms, changing rooms, day rooms, guest WC, etc. "EVEXIA" treats post-hospital patients suffering from diseases such as:

• Neurological diseases: stroke, multiple sclerosis, Parkinson's disease, motor neuron disease, cerebral palsy, CNS degenerative diseases.

• Post-operative orthopedic conditions: intertrochanteric fracture, total hip or knee replacement,

• Post-operative neurological diseases: craniocerebral injuries, paraplegia-tetraplegia, paraparesis.

• Rheumatic diseases: ankylosing spondylitis, rheumatoid arthritis, psoriatic arthritis.

The REA system will be installed in ten patient rooms, which will be monitored in terms of sensor data and camera feed during the project's lifespan. Apart from the clinical environment, REA will be also tested in two home environments of patients that follow a rehabilitation program.

## 8.2 A use case scenario

We describe a use case scenario that involves the conversation of a clinician with the REA system in order to obtain information about the health profile and clinical condition of a patient. The virtual agent acts as a mediator, coordinating and facilitating the hospital staff and therefore patients, by swiftly offering crucial information, allowing clinicians to easily obtain an overall picture of the clinical condition of the patients. Table 1 presents a relevant dialogue that is supported. In the following, we present the way the system handles this conversation and the technologies involved.

**Table 1 Hospital-oriented use case example**

| | |
|---|---|
| [i1] Doctor | REA, I would like to know the medication that is being administered to Mr. Smith. |
| [i2] REA | Mr Smith is in treatment for 3 drugs. In morning he is administered (Drug 1), at midday he is administered (Drug 2) and finally at night he is administered (Drug 3) and (Drug 1). However, he complains that the medication is not so effective for him. |
| [i3] Doctor | REA, what's his clinical condition? |
| [i4] REA | Patient's systolic pressure is 110 and diastolic is 70, i.e. at normal levels. The heartbeat of the patient is slightly elevated to 80. Finally, the temperature is 37 degrees Celsius. |
| [i5] Doctor | Thank you REA. |
| [i6] REA | You are welcome Dr. Gray. |

*8.2.1 Language understanding*. The text obtained from the speech-to-text module is being further analyzed in order to extract entities, concepts and relations. For example, from [i1] REA recognizes the key concepts "medication", "administration" and the

[10] http://www.evexia.com/en/

named entity "Smith". The concepts are mapped to the BabelNet resources:
medication -> https://babelnet.org/synset?word=bn:00054128n
administration -> https://babelnet.org/synset?word=bn:00001424n,
while the named entity is used in order to retrieve from the KB the profile of the patient.

*8.2.2 Reasoning.* Having analyzed the text of the request, the next step is to semantically understand the context of the request, in order to feed the DM with relevant knowledge. The ontology reasoning is used in this case to recognize the topics of the conversation taking into account the ontological knowledge that has been infused in the system. For example,

$$MedicationAdministration \equiv VerbalContext \sqcap \exists contains.Medication \sqcap \exists contains.Adminstration$$

In our example, the definition of MedicationAdministration is fully satisfied, and therefore the verbal context is classified as a MedicationAdministration request.

*8.2.3 Dialogue management.* Both the recognized topic of the request (MedicationAdministration) and the named entity (Smith) are propagated to the DM in order to determine the appropriate system action. Based on the provided input, DM handles the request by calling the necessary services in order to form the response that will be sent back to the user. As such, DM queries the KB to obtain information about the medication, i.e. the drug type and the schedule. In our example, the response has the following structure:

```
{
  "type": "MedicationAdministration",
  "": [
    {
      "drug": "Drug 1",
      "schedule": "Morning"
    },
    {
      "drug": "Drug 2",
      "schedule": "Midday"
    }, …
  ]
}
```

The DM can also raise system actions that are not directly connected with the current dialogue state based on history data. For example, if the user has complained more than a set amount of times during the last day about the medication (the exact number will be adapted with the aid of medical staff), REA will also inform the doctor about this. In this case, the topic of user request (MedicationAdministration) will be reacted with the pre-defined system response (DetailedMedicationAdministration) and with a dynamic system notification (UserComplain) which includes the "medication" as target entity. The intermediate representation of system actions and concepts are then sent to the speech generation module to verbalize the result ([i2]).

The conversation then progresses in a similar way. It is worth mentioning that in i3 the system needs to be able to handle coreference resolution, since the doctor does not refer to the patient with his name. This case is handled at the DM-level, taking into account the history of the conversation, mapping "his" to "Smith".

## 9 CONCLUSIONS

The work in progress described in the present paper encompasses the development currently underway for all main modules of the REA platform. Each module caters to different technological needs, which compose a system designed with the goal to facilitate natural human-machine communication in the fields of patient recuperation and rehabilitation. Medical professionals and end-users in clinical and non-clinical environments (at home) will evaluate the project's results via extensive trials that will be carried out during the prototypes' testing.

## ACKNOWLEDGMENTS

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE (project code: T1EDK-00686).

## REFERENCES

1. Duygu Altinok. 2018. An ontology-based dialogue management system for banking and finance dialogue systems. *arXiv preprint arXiv:1804.04838*.
2. Konstantinos Avgerinakis, Alexia Briassouli, and Ioannis Kompatsiaris. 2013. Recognition of activities of daily living for smart home environments. In *2013 9th International Conference on Intelligent Environments*, 173–180.
3. Franz Baader, Diego Calvanese, Deborah McGuinness, Peter Patel-Schneider, and Daniele Nardi. 2003. *The description logic handbook: Theory, implementation and applications*. Cambridge university press.
4. Jerome R Bellegarda and Christof Monz. 2016. State of the art in statistical methods for language and speech processing. *Computer Speech & Language* 35: 163–184.
5. Leo Breiman. 2001. Random forests. *Machine learning* 45, 1: 5–32.
6. Danica Damljanović, Milan Agatonović, Hamish Cunningham, and Kalina Bontcheva. 2013. Improving habitability of natural language interfaces for querying ontologies with feedback and clarification dialogues. *Web Semantics: Science, Services and Agents on the World Wide Web* 19: 1–21.
7. Sébastien Dourlens, Amar Ramdane-Cherif, and Eric Monacelli. 2013. Multi levels semantic architecture for multimodal interaction. *Applied intelligence* 38, 4: 586–599.
8. Aldo Gangemi and Peter Mika. 2003. Understanding the semantic web through descriptions and situations. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 689–706.
9. Albert Gatt and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61: 65–170.
10. Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, Bijan Parsia, Peter Patel-Schneider, and Ulrike Sattler. 2008. OWL 2: The next step for OWL. *Web Semantics: Science, Services and Agents on the World Wide Web* 6, 4: 309–322.
11. Juergen Heit, Soundararajan Srinivasan, Diego Benitez, and Burton Warren Andrews. 2013. Device and method to monitor, assess and

improve quality of sleep.

12. Julian Hough. 2011. Incremental semantics driven natural language generation with self-repairing capability. In *Proceedings of the Second Student Research Workshop associated with RANLP 2011*, 79–84.
13. Ye Jia, Yu Zhang, Ron Weiss, Quan Wang, Jonathan Shen, Fei Ren, Patrick Nguyen, Ruoming Pang, Ignacio Lopez Moreno, Yonghui Wu, and others. 2018. Transfer learning from speaker verification to multispeaker text-to-speech synthesis. In *Advances in Neural Information Processing Systems*, 4485–4495.
14. Pierre Lison. 2014. Structured probabilistic modelling for dialogue management.
15. David D Luxton, Jennifer D June, Akane Sano, and Timothy Bickmore. 2016. Intelligent mobile, wearable, and ambient technologies for behavioral health care. In *Artificial Intelligence in Behavioral and Mental Health Care*. Elsevier, 137–162.
16. Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 55–60.
17. Thanassis Mavropoulos, Dimitris Liparas, Spyridon Symeonidis, Stefanos Vrochidis, and Ioannis Kompatsiaris. 2017. A Hybrid approach for biomedical relation extraction using finite state automata and random forest-weighted fusion. In *International Conference on Computational Linguistics and Intelligent Text Processing*, 450–462.
18. Georgios Meditskos, Stamatia Dasiopoulou, and Ioannis Kompatsiaris. 2016. MetaQ: A knowledge-driven framework for context-aware activity recognition combining SPARQL and OWL 2 activity patterns. *Pervasive and Mobile Computing* 25: 104–124.
19. Behzad Mirmahboub, Shadrokh Samavi, Nader Karimi, and Shahram Shirani. 2013. Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE Transactions on Biomedical Engineering* 60, 2: 427–436.
20. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.
21. Rivindu Perera and Parma Nand. 2017. Recent advances in natural language generation: A survey and classification of the empirical literature. *Computing and Informatics* 36, 1: 1–32.
22. June J Pilcher, Douglas R Ginter, and Brigitte Sadowsky. 1997. Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. *Journal of psychosomatic research* 42, 6: 583–596.
23. Vincent Pollet, Enrico Zovato, Sufian Irhimeh, and Pier Domenico Batzu. 2017. Unit Selection with Hierarchical Cascaded Long Short Term Memory Bidirectional Recurrent Neural Nets. In *INTERSPEECH*, 3966–3970.
24. Louisa Pragst, Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 2017. Challenges for adaptive dialogue management in the KRISTINA project. In *Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents*, 11–14.
25. Ehud Reiter and Robert Dale. 2000. *Building natural language generation systems*. Cambridge university press.
26. Caroline Rougier, Jean Meunier, Alain St-Arnaud, and Jacqueline Rousseau. 2006. Monocular 3D head tracking to detect falls of elderly people. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 6384–6387.
27. Ryan Shaw, Raphaël Troncy, and Lynda Hardman. 2009. Lode: Linking open descriptions of events. In *Asian semantic web conference*, 153–167.
28. Jonathan Shen, Ruoming Pang, Ron J Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, Rj Skerrv-Ryan, and others. 2018. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4779–4783.
29. Keiichi Tokuda, Yoshihiko Nankaku, Tomoki Toda, Heiga Zen, Junichi Yamagishi, and Keiichiro Oura. 2013. Speech synthesis based on hidden Markov models. *Proceedings of the IEEE* 101, 5: 1234–1252.
30. Shubham Toshniwal, Anjuli Kannan, Chung-Cheng Chiu, Yonghui Wu, Tara N Sainath, and Karen Livescu. 2018. A comparison of techniques for language model integration in encoder-decoder speech recognition. *arXiv preprint arXiv:1807.10857*.
31. Yuxi Wang, Kaishun Wu, and Lionel M Ni. 2017. Wifall: Device-free fall detection by wireless networks. *IEEE Transactions on Mobile Computing* 16, 2: 581–594.
32. Michael Wessel, Girish Acharya, James Carpenter, and Min Yin. 2019. OntoVPAAn Ontology-Based Dialogue Management System for Virtual Personal Assistants. In *Advanced Social Interaction with Agents*. Springer, 219–233.
33. Juan Ye, Stamatia Dasiopoulou, Graeme Stevenson, Georgios Meditskos, Efstratios Kontopoulos, Ioannis Kompatsiaris, and Simon Dobson. 2015. Semantic web technologies in pervasive computing: A survey and research roadmap. *Pervasive and Mobile Computing* 23: 1–25.
34. Heiga Ze, Andrew Senior, and Mike Schuster. 2013. Statistical parametric speech synthesis using deep neural networks. In *2013 ieee international conference on acoustics, speech and signal processing*, 7962–7966.
35. Yaniv Zigel, Dima Litvak, and Israel Gannot. 2009. A method for automatic fall detection of elderly people using floor vibrations and sound. Proof of concept on human mimicking doll falls. *IEEE Transactions on Biomedical Engineering* 56, 12: 2858–2867.