KRISTINA

A Knowledge-Based Information Agent with Social Competence and Human Interaction Capabilities

H2020-645012

D7.5

Final System

Dissemination level:	Public
Contractual date of delivery:	Month 35, 31-01-2018
Actual date of delivery:	Month 37, 01-03-2018
Workpackage:	WP7 Agent platform development and integration
Task:	T7.5 Development of the Final System
Type:	Report/Other
Approval Status:	Concept
Version:	1.0
Number of pages:	16
Filename:	D7.5-Final System_2018-03-01_v1.0.docx

Abstract

We explain the functional and technical state of the final version of the KRISTINA prototype (Deliverable D7.5), which has been implemented according to the design specified in Deliverable D7.1 using the operational system platform from Deliverable D7.2, enhancing the earlier prototypes D7.3 and D7.4. Where applicable, we explain deviations from the original plan.

co-funded by the European Union

History

Version	Date	Reason	Revised by
0.1	01-01-2018	First draft	Andries Stam
0.2	25-01-2018	Second draft	Ludo Stellingwerff
1.0	22-02-2018	Final version	All authors
1.0	01-03-2018	Revised the final version and submitted	Leo Wanner

Author list

Organization	Name	Contact Information
AL	Andries Stam	andries@almende.org
AL	Ludo Stellingwerff	ludo@almende.org
AL	Luis De Cunha	luis@almende.org
AL	Hans Abbink	hans@almende.org
AL	Alex de Mulder	alex@almende.org
UPF	Juan Sanchez Pellicer	joanpere.sanchez@upf.edu
UPF	Oriol Martinez Pujol	oriol.martinez@upf.edu
UPF	Hermann Plass	hermann.plass@upf.edu
UPF	Stamatia Dasiopoulou	stamatia.dasiopoulou@upf.edu
UPF	Roberto Carlini	roberto.carlini@upf.edu
UPF	Mireia Farrús	mireia.farrus@upf.edu
UPF	Mónica Domínguez	monica.dominguez@upf.edu
UPF	Simon Mille	simon.mille@upf.edu
UPF	Alicia Burga	alicia.burga@upf.edu
CERTH	Eleni Kamateri	ekamater@iti.gr
CERTH	Thanassis Mavropoulos	mavrathan@iti.gr
CERTH	Spyridon Symeonidis	spyridons@iti.gr
CERTH	Georgios Meditskos	gmeditsk@iti.gr
CERTH	Stefanos Vrochidis	stefanos@iti.gr
VR	Thiago Fraga	thfraga@vocapia.com
UAU	Dominik Schiller	dominik.schiller@informatik.uni-augsburg.de
UULM	Louisa Pragst	louisa.pragst@uni-ulm.de

Executive Summary

We explain the functional and technical state of the final version of the KRISTINA prototype (Deliverable D7.5), which has been implemented according to the design specified in Deliverable D7.1 using the operational system platform from Deliverable D7.2, enhancing the functionality of the earlier prototypes D7.3 And D7.4. Where applicable, we explain deviations from the original plan.

A demonstration of the final version of the KRISTINA prototype has been recorded on video. This video can be found at: **http://kristina-project.eu/en/videos/kristina-final-version/**

Abbreviations and Acronyms

AL	Almende BV (project partner)
AU	Action Unit
CERTH	Centre for Research and Technology Hellas (project partner)
DM	Dialog Manager
DoA	Description of the Action - Grant Agreement
GTI	Grup de Tecnologies Interactives – UPF (research group from project partner)
GUI	Graphical User Interface
P2T	Push-to-talk or push2talk.
REST	Representational State Transfer
SSI	Social Signal Interpretation
TALN	Tractament del Llenguate Natural – UPF (research group from project partner)
UAU	University of Augsburg (project partner)
UPF	Universitat Pompeu Fabra (project partner)
UULM	University of Ulm (project partner)
VR	Vocapia Research (project partner)
VSM	VisualSceneMaker

Table of Contents

1 INTRODUCTION

This report accompanies Deliverable D7.5, the final version of the KRISTINA agent, which has been delivered by the technical partners for Milestone 5 of the project.

The final version of the KRISTINA agent has been implemented according to the design specified in Deliverable D7.1 using the operational system platform from Deliverable D7.2. The prototype builds on top of the functionality offered by the earlier prototypes D7.3 and D7.4, with each module demonstrating an enhanced state-of-the-art. This accompanying report provides a functional and technical overview of the final version.

In Section 2, we provide a functional overview of the final version. We explain the functionality of each of its functional modules and provide an overview of the scenarios supported.

Section 3 contains a technical overview of the final version. As this version is available as a downloadable virtual machine image, the process of downloading, starting and using this image is explained. As such this document forms the documentation for handling the image and building upon the results of the project.

It is assumed that the reader is familiar with the contents of Deliverable D7.1, which acts as a reference and definition work for the terms and models used in this document, as well as with the contents of the report accompanying Deliverables D7.2, D7.3 and D7.4, which explains the set-up of the operational system platform and the earlier prototypes. In addition, the reader should have knowledge of the scenarios described in Deliverable D8.6.

A demonstration of the final version of the KRISTINA agent has been recorded on video. This video can be found at:

http://kristina-project.eu/en/videos/kristina-final-version/

2 FUNCTIONAL OVERVIEW OF THE FINAL VERSION

With the development of the final version of KRISTINA agent, we are showing the ultimate result in the various modules of the system. Most technical components are demonstrating enhanced state-of-the-art. The system as a whole is therefore more capable than the earlier version, allowing more complex interaction scenarios with the users.

The final version of the KRISTINA agent serves an important role as part of Milestone 5:

> To demonstrate the full functionality of all components and of the KRISTINA Agent as an integrated system;

In the following subsections, we list the functionalities of the individual modules of the final version (Section 2.1) and provide an overview of the implemented scenarios (Section 2.2).

2.1 Modules and their functionality

In the following sections, the different modules involved and their functionality are described. The partner responsible for the development of the module is indicated as well.

2.1.1 Speech-To-Text (VR)

As part of the final version, the STT module has been enhanced and trained to cover all five target languages of the KRISTINA project: Spanish, German, Polish, Turkish and Arabic, using the audio and text data available.

2.1.2 Gesture Analysis (ALM)

The Gesture analysis module has been extended to include detection of semantically specific gestures, like head nod, head shake and hand waving. This adds to the existing features of arousal detection and pointing gestures.

As part of the final version integration effort, the gesture analysis module is now using the output of the Face Analysis module, as described in Section 2.1.3. This solves some performance problems that were inherent to the second prototype.

2.1.3 Face Analysis (UPF)

For the final prototype, the already fully operational face analysis module -that automatically maps the facial appearance to Valence-Arousal emotional labels- has been enhanced significantly with regards to speed and quality in emotion estimation.

First, the face recognition module has been upgraded, providing a more efficient detection. Specifically, the landmark detection sub-module can now be pushed up to 50 fps. For that reason, it is shared -and therefore used-, by other components of the non-verbal analysis pipeline (gesture analysis and voice activity detection).

Second, emotional inference using the Valence-Arousal (V-A) space has also been improved by incorporating new data from the KRISTINA corpus. Besides that, all emotion estimation sub-modules (V-A and prototypical facial expressions) have been re-balanced and properly evaluated.

2.1.4 **Voice Activity Detection (UAU)**

A new feature provided by SSI is the advanced voice detection approach. Besides using audio cue's to recognize if the user is speaking, in this new approach the face analysis output is also used. Through this output, the voice detection module will only trigger on audio if it detects the user's lips moving.

This advanced feature also provides echo cancellation, by not triggering on the agent's output while the user is not speaking.

2.1.5 **Language Analysis (UPF)**

In the third prototype, language analysis is supported in German, Polish, Spanish, Turkish and Arabic. In the third year, an Arabic surface-syntactic parser has been trained; parallel surface- and deep-syntactic corpora have been annotated manually in Spanish (1,500 sentences), Turkish (1,500 sent.), German (1,000 sent.) and Polish (500 sent.); deep-syntactic parsers could be trained for Turkish, Polish and German. Graph-transduction grammars and lexicons for deep analysis have been improved (coverage, speed) so as to cover all KRISTINA Use Cases.

2.1.6 **Emotion Analysis (UAU)**

Based on the evaluation of the Second prototype, the parameters of the multi-modal fusion have been adjusted and optimized.

2.1.7 **Turn Control (UAU)**

The more advanced Voice detection, as described in section 2.1.7., has some influence on the turn control as well. It leads to more complex and flexible flows through the system, as the user can more easily barge-in or take turn.

2.1.8 **Ontology Annotation (CERTH)**

For the final system, the focus has been shifted on extending the vocabularies and semantics used to capture information at various levels of abstraction. This involves the development of additional patterns to model user profile-related information (according to the final version of the use cases), as well as models to capture verbal and non-verbal information (for fusion) and response types: More specifically:

- User models: Apart from describing preferences and habits, the final version of the user models supports the modelling of user diseases and problems. The KB of WP5 is able to provide formal information about, for example, sleep problems or diseases that people might have, such as dementia.
- Verbal information: We adopted a refined interpretation that takes into account the ontological type of the considered frames that are extracted during the linguistic analysis step. This allows for greater flexibility when it comes to mapping to the reference domain ontologies; in parallel, the preservation of the associations with the underpinning linguistic predicative resources ensures the availability of the information needed for linguistic generation. Moreover, in order to recognize

additional conversational topics, a thematic ontology has been developed that semantically associates topics with domain concepts extracted through verbal and non-verbal information.

- Non-verbal information: A number of new ontologies have been developed, while existing ones have been refined, to capture the non-verbal information that becomes available through the analysis modules of WP4, such as to model body parts, pointing gestures and facial expressions. It should be noted that facial expressions are directly mapped to emotions, based on the relevant classifications that are provided by analysis modules in WP3.
- Response models: In order to support the new functionality of the DM (e.g. extended proactiveness and verbosity), the structure and semantics of the response messages have been extended to adequately capture the interpretation results of the underlying reasoning module.

2.1.9 **Semantic Archival and Knowledge Integration (CERTH)**

The research on context interpretation and reasoning (realised through the Knowledge Integration module / KI) has mainly focused on the development of methodologies for supporting advanced conversational awareness, multimodal fusion, question answering and user-tailored responses.

More specifically, conversation awareness is now achieved by capturing dependencies among high-level topics and low-level input events in a loosely-coupled manner, rather than defining strict contextual patterns that cannot provide enough flexibility for handling the imprecise and ambiguous nature of the domain. In that way, the module is able to recognise the conversational topics more efficiently. Conversational awareness is also used to extend the KI with multimodal fusion capabilities, able to derive plausible interpretations of the ongoing interaction with the user, inferring complex situations that need to be specially treated. Question answering has been also enhanced, following a context-aware matching and extraction algorithm for retrieving content from the KB relevant to the user input. Finally, proactiveness and verbosity have been further extended, allowing the DM: a) to drive the interaction with the user by suggesting additional topics of discussion, taking into account the information needs of the user, and b) to provide richer responses to the user by combining two or more responses with similar content.

We have also extensively evaluated both the reasoning and the representation capabilities of the WP5 semantic framework using the OOPS! tool and relevant metrics (reported in D5.4).

2.1.10 **Dialog Management (ULM)**

In the final version, the DM module has been enhanced to to utilise probabilistic action selection rules in combination with general features of dialogue actions. The probabilities of the rules have been trained on the corpus recordings and with regard to the speakers culture, to enable natural and culturally appropriate behaviour. The DM's adaptability is even further improved by the capability to estimate the user's currently employed level of elaborateness and directness, and to adjust its own responses accordingly. In that regard, the identification of suitable additional information has been improved and the ability to phrase offers indirectly has been added.

2.1.11 Mode Selection (UPF)

During the third year of the project, the coverage of the mode selection rules has been improved, and classifiers for statistical assignment of the modalities have been trained.

2.1.12 Gesture Generation (UPF)

The non-verbal output pipeline of the KRISTINA agent saw some significant changes, because the planning of both non-verbal communication and idle behavior was moved into the Avatar services. A basic version of a BML Planner for the nonverbal generation and coordination of modalities was developed by UPF-GTI in order to improve the openness and interoperability of the (web-based) solution. The planner receives a message with the dialogue act, valence and arousal, overall state (waiting, listening, processing or speaking and speech information), and generates non-verbal content in a holistic manner, coordinating gaze, head or full body animations. Non-verbal content is generated in the BML Planner by rules, and when speaking it coordinates speech with lip syncing, facial expressions, head nods, gaze shifts and gestures. For states such as waiting, listening and processing, it generates autonomously non-verbal BML commands. A key novelty of the final stage is that the planner supports Mode Selection (provided by UPF-TALN) which modifies facial and gesture animations.

UPF-GTI has also developed a fully operational web-based BML Realizer as part of the Avatar Rendering module: it turns output from the planner into 3D rendered character animations. The current version supports blink, face, faceShift, facial lexemes, gaze, gazeShift, head, headDirectionShift, gestures, speech, synchronization references and composition modes. The visual avatar in the final prototype can show a larger part of the body of the character, allowing for the introduction of body gestures in addition to the face and head gestures already available in earlier prototypes. In order to cater for more natural and appealing animations in the framework of BML based holistic body animation, the main back-end improvements are (Forward Kinematics) support for skeletal animation, animation blending based on Finite States Machines, and “mood”ifiable animations through the use of Adobe Mixamo – see also later. The final prototype also renders captions.

2.1.13 Spoken Language Generation (UPF)

For language generation, the work done during the third year includes the following: improvement of coverage and flexibility of the graph-transducers and completion of lexical resources for text generation; training of statistical text generators in German, Polish and Turkish; integration of a new template-based module that generates clarifications and proactive suggestions (German, Turkish, Spanish, Polish) in direct and indirect modes; integration of the feminine Arabic voice; improvement of the prosody for the German voice (Mary TTS).

2.1.14 Face Generation (UPF)

The Face Generation module has been extended as part of the full avatar generation effort as described in 2.1.12 and 2.1.15. Technically it’s largely integrated with the Gesture Generation as described in paragraph 2.1.12. For the final prototype, the most significant enhancements come from back-end improvements, namely, in facial lexemes through improved blendshapes blending and homogeneous lip-sync configuration, partly based on more flexibility of the tools, as custom face lexemes can be generated by modifying weighting in a multi-axial valence-arousal space.

2.1.15 **Avatar Rendering (UPF)**

The Avatar Rendering module has been enhanced to support a larger body model of the avatar. Work has been done to improve the three avatars for the final prototype, based on the requirements of the user partners, which UPF-GTI gathered through questionnaires about the visual aspect, traits and background scene of the ECAs. The final virtual characters have been developed according the requested characteristics for: (i) a Caucasian female, (ii) an Arabic male, and (iii) an Arabic female. In the final stages of the project, the appearance of the characters, and the animations have been improved, in a further step to deal with a key UPF-GTI concern for the generation of those characters, namely, the homogeneity between the different models, using automatisms capable of generating better automated blend-shapes more similar to each other, to improve the quality of the lip-sync, the facial expression, and the gestures, in general. As discussed in other deliverables, this has been enabled through enhancements in the pipeline generation, where Adobe Mixamo and Adobe Fuse have been used. On the other hand, the animation system of WebGLStudio has been improved, as well as some rendering aspects.

2.1.16 **Content Extraction (CERTH)**

The crawling implementation remained unchanged, nevertheless, we utilised it to crawl more domains relevant to the needs of the final system. Scraping rules of each respective component were further updated and two different alternative approaches are employed (described in detail in D5.5), depending on the language and the domain of the resource. The updated version also manages to identify the boundaries of the paragraph of each web resource. Evaluation of both alternatives was performed using two datasets, an external and a manually created one.

Regarding the information indexing and retrieval module progress, an extra field was added to the constructed indices. This field contains the concepts extracted from each text using the extraction tool provided by UPF[1]. Thus, apart from searching and matching only the terms of a query, concepts are also taken into account (they are extracted from the query as well). To evaluate the retrieval performance, a ground truth dataset was created with the help of end-users using an annotation tool. The annotation tool has been developed on top of the open-source BioASQ tool[2] implementation which was provided for the BioASQ challenge on biomedical semantic indexing and question answering.

The original Metamap-based concept extraction approach was upgraded with the introduction of the DBpedia Spotlight service to the framework which boosted its results considerably. However, in the latest version of the concept extraction module, in lieu of the abovementioned approach, a new neural network-based methodology was adopted, which further improved results. The approach involves the exploitation of an LSTM-CRF network that leverages BILOU format annotation when handling input text which permits the extraction of medical concepts via the use of word and character embeddings. With respect

[1] http://kristina.taln.upf.edu/services_dev/language_analysis/extract_concepts

[2] https://github.com/BioASQ/AnnotationTool

to relation extraction, there have been enhancements to the three distinct modules that comprise the hybrid system. More precisely, extra features have been added to the machine learning one that gave an overall (significant) performance boost. The pattern based part was extended with new finite state automata that facilitated the handling of more use case scenarios (e.g. side effects of vaccines, factors that determine the weight and size of babies etc.). Finally, the weighting module benefited from an updated approach with regards to weight distribution; Evaluation of the results using latest features favours the use of equal weights (50%-50%) instead of the previously exploited Random Forest-based approach.

The social media module was also enhanced with more functionalities. The topic detection service was further adapted to address the final system needs. Tweet search functionalities were also integrated; criteria for search include username, keywords, recently posted tweets and tweet category. With regards to the last criterion, a separate sub-module that classifies tweets into one of a list of six predefined categories was developed. These categories were found in the IPTC news codes taxonomy as representative categories to characterise a news text[3]. The classification method involves the fusion of random forest models that were trained using different types of text vector representations, namely word2vec and n-grams. Another annotation tool was set up in order to form datasets for training and evaluating the classification models. The tweet category classification sub-module is run on a daily basis in order to continuously predict and store the category of the tweets whose category information is missing.

In newspaper retrieval, a service was added for displaying a list of newspaper article titles that exist in the newspaper indices, either random or the newest ones (when the article date of publication is available). The weather service was updated as well, in order to store and support weather predictions for more towns.

Finally, video retrieval via relevant descriptive tag exploitation is an additional feature introduced for the final system.

2.1.17 **External URL presentation (ALM)**

As described in D7.1, the system should support presenting the user with external information in the form of social network content (e.g. tweets), podcasts, videos, brochures, and other general (web)content. Compared to the second prototype, this feature has been enhanced by changing the recommended browser settings. For the final version the recommendation is to change the Chrome browsers configuration flag “Throttle expensive background timers” to “Disabled”. This will prevent the external URL from blocking the agent finishing her output while opening.

2.2 Supported scenarios

In Deliverable D8.6 various scenarios for the second pilot setup have been described. These scenarios are supported by the second prototype, as a proof of concept and as content for the pilot tests.

Depending on the scenario, the user solicits assistance in different contexts. The forms of

[3] http://www.iptc.org/site/Home/

assistance KRISTINA will provide in the pilot use cases are: (i) **social companionship**, (ii) **nursing assistance**, and (iii) **health expert**. KRISTINA operates on these different forms of assistances. Therefore KRISTINA can ask for information and provides entertainment, information and advice.

As a **social companion** the idea is that KRISTINA offers small talk and companionship. For the first prototype, KRISTINA is able to read aloud newspaper articles upon user request or agreement with such suggestion, and also to provide weather information and carry out related small talk. Through the external URL feature, the system can provide the user with more in-depth information where applicable.

As a **nursing assistant** KRISTINA inquires personal information and personal care needs. The inquiry can be performed in a dialogue or entered in a written form. Subsequently, these personal data can be offered in communication with a caregiver. For the 2nd prototype, KRISTINA supports conversations in the domain of dementia care, with some extended focus on sleep and diet habits. For example, the system can provide specific recipes that the user would appreciate.

As a **health expert** KRISTINA offers health information or information on (procedures of) the health-care system. In addition to the scenarios of the first prototype, this has been extended with topics that support the care for people suffering from various forms of dementia. The baby care scenario has also been extended, including topics like vaccination, check-up scheduling and activity suggestions.

An important enhancement of the final version is the addition of one more language.

3 TECHNICAL OVERVIEW OF THE FINAL SYSTEM

3.1 Overview of the set-up

In this section an overview is given of the deployment of the final version of the KRISTINA agents and of its various modules.

For the final version, nearly all modules are deployed into a single virtual machine image. The only exception is the Speech Recognition module, which remains hosted by Vocapia. Because all inter-module communication is RESTful, this migration from a cloud service to a VM based deployment has been relatively straight-forward.

Because of the required memory footprint of some of the modules, this image requires a quite heavy host system. The working system as shown in the accompanying video is running on an Intel i7 system with 32GB RAM. The image has been run using QEMU/KVM on an Ubuntu 17.10 host, but there is also a second image available for VirtualBox on MS Windows.

3.2 Steps to run the image, manual

Below you'll find a step by step plan to run the image.

1. Download the image from the Kristina's project website:
    a. http://kristina-project.eu/en/downloads/kristina-final-version/

2. Configure your system for running virtual machines:
   a. On Ubuntu Linux:
      i. Install qemu-kvm
      ii. Run the following script to start the image:
      echo 3 | sudo tee /proc/sys/vm/drop_caches &> /dev/null
      sudo sysctl vm.nr_hugepages=13000
      *qemu-system-x86_64 -m 23G -hda win2016.qcow2 -enable-kvm -cpu host -mem-path /dev/hugepages *
      *-smp 6,cores=6,sockets=1 -M pc -device rtl8139,netdev=n0 -netdev *
      user,id=n0,hostfwd=tcp::8000-:8000,hostfwd=tcp::8001-:8001,hostfwd=tcp::8002-:8002,
      hostfwd=tcp::8004-:8004,hostfwd=tcp::8005-:8005 &
      sleep 2; sudo sysctl vm.nr_hugepages=0
      iii. Continue step 3 below
   b. On MS Windows:
      i. Allocate at least 23GB of memory to the image in VirtualBox.
      ii. Forward ports 8000-8005 to the virtual image.
      iii. Start the image through VirtualBox.
      iv. Continue step 3 below
3. After boot, login to the image using the admin password: “Kr1st1n@”.
4. After the login, double click the “start_all.bat” icon on the desktop.
5. After about four minutes the image is fully started.
6. Prepare the Google Chrome browser on the host system:
   a. Add the certificate from: http://localhost:8004/localhost.crt to the trusted root certificate authorities. (You can follow the howto at: https://www.bonusbits.com/wiki/HowTo:Import_Certificate_Authority_Root_Certificate_in_Google_Chrome)
   b. Set chrome://flags/#expensive-background-timer-throttling to “disabled”.
7. Open https://localhost:8005/web-gui in the Google Chrome browser on the host system.
8. Accept the request for video and audio usage.
9. Select the correct scenario, through the scenario selection popup.
10. The first input will take about 1 minute longer than normal, due to the loading of the language parsers. This will be shown through the progress lights.
11. If the progress lights don’t show any reaction, including from manual text input, please reload the browser tab, and re-select the scenario.
12. If that doesn’t work, please open the red Settings popup and click the “reconnectWS” button.

3.3 GUI changes compared to Second Prototype

There are a couple of noteworthy changes to the Integrated Web-based GUI for the KRISTINA agent:

- Disabling of the reservation system. Because the final version is a single VM image there is no need for a reservation system to enforce the single-system-single-user rule that KRISTINA has employed.
- Introduction of a progress bar. The GUI now contains a progress bar, showing the server-side progress through the various modules. (Mouse-over provides module names)

- Multiple avatars. An update to the avatar has been done, including the addition of a second avatar, these can be selected through the scenario selector. The Avatar also provides a subtitle now.
- Option to disable the push2talk button. Because there is a better voice detection approach, as described in section 2.1.4, the push2talk button can be bypassed. In the Settings, there is now an option to disable the p2t button.

4 CONCLUSIONS

The final version of the KRISTINA agent has now been implemented. Thereby, Task T7.5, an important task for Milestone 5, has been concluded and the results of the task have been delivered in terms of the virtual machine image of the final version, this accompanying report and a video demonstrating the agent.