

# CHROME Project: Representation and Survey for AI Development

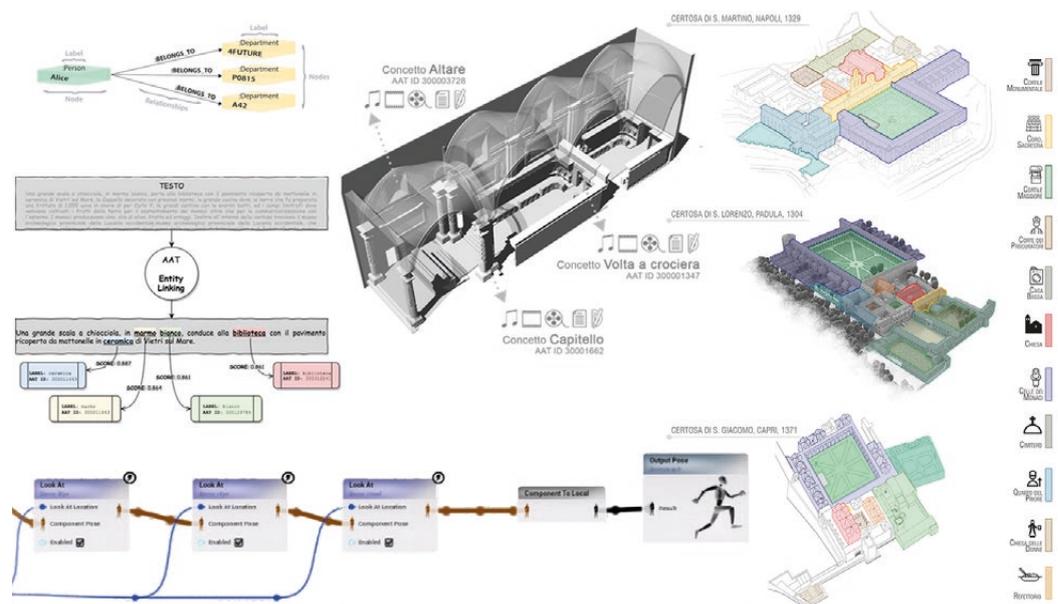
Massimiliano Campi  
Valeria Cera  
Francesco Cutugno  
Antonella di Luggo  
Domenico Iovane  
Antonio Origlia

## Abstract

The paper shows the results of the PRIN CHROME Cultural Heritage Orienting Multimodal Experiences project, about the three charterhouses of Campania, with a specific focus on research activities related to the connections between representation, survey, AI and VR. The project has formalized a methodology of collection, analysis and modeling of multimodal data, useful for designing virtual agents in 3D environments, which can be applicable in museum environments. The achievement of the goal is pursued through: (i) an integrated range-based acquisition and morphometric data modeling process coherent with VR management, (ii) the use of semantic maps linked with thesauri published as LOD to solve both the theme of ambiguity and annotation uncertainty and the inter-pretability of information by an AI; (iii) the modeling of a virtual agent with the development of a mathematical model for computational control of gestures and prosody.

## Keywords

semantic annotation, artificial intelligence, Unreal Engine 4, graph databases.



## The CHROME Project: Methodology and Procedures

The paper shows the specific results of the PRIN 2015 CHROME Cultural Heritage Resources Orienting Multimodal Experiences, developed around the case studies of the three charterhouses of Campania, with the focus on research activities related to the connections between representation, survey, artificial intelligence and virtual reality [1].

The project, that is strongly inter-disciplinary, has formalized procedures for collecting, annotating and analyzing multimodal data – such as written texts, oral presentations, 3D models – for a subsequent use by the AI.

In particular, the resources collected and annotated have served to design a virtual agent inserted in 3D virtual scenarios. This Virtual agent can be applicable in museum environments and joins the tour guides increasing the potential for intervention on the public visiting cultural sites. The virtual agent, in fact, simulates social signals through computational control of gestures and prosody according to a mathematical model based on the behavior of operators specialized in the communication of cultural contents.

The base knowledge has therefore been structured in order to build a model that allows to compose a not default and potentially adaptable to the type of interlocutor oral presentation. The achievement of the goal was pursued through the semantic association of the whole corpus of information to the geometric entities of the spatial model, that are digital clone of the real good to which the enhancement is addressed. The annotation of 3D representations made it possible to link the presentation to the automatic selection of the auxiliary material and to query it with a natural language dialogue system, in which the information is spatialized. As disclosed here, attention is focused on those investigative activities related to the inter-connections gained between the disciplinary of representation and survey and the domain of computer science, related to each other and put at the service of the development of AI applications in augmented and virtual reality environments.

Since this background, the specific research investigated the theoretical and methodological issues related to the geometric and semantic manipulation of digital representations of architectures or rather, on one hand, those of a terminological-significant nature and, on the other, the ones of geometric-formal matrix. The first ones involve the process of meaning assigning to spatial forms, the latter concern both the processes of "construction" of the digital clone and the method by which recognizing on it the geometric boundary of semantic concepts. In addition, the considered segmentation approaches have been strictly aimed at storing content in an AI-questionable system, made able to disseminate information in digital settings that can be experienced through AR/VR technologies. This last aspect involved a reflection upon the most appropriate ways of graphic simplification of the elements of the heritage in order to make their vision fluid in a system of virtual use without losing neither the realistic rendering nor the understanding of the contents.

### Representation for Semantic Structuring and Knowledge Formalization

The first phase of the study involved the realization of the digital virtual scene, to be semantically annotated, for subsequent use by the AI.

For the three case studies of the project, the charterhouse of San Martino in Naples, the San Lorenzo one in Padula and the San Giacomo one in Capri, important campaigns have been carried out. These have seen the integration of passive and active optical sensors in order to achieve accurate, precise and photorealistic three-dimensional models, returning both of the overall morphology of the different convents and the complexity of the decorative details of the interiors.

Starting from the integrated range-based acquisition of morpho-metric data, point clouds were modeled with classic triangulation algorithms and subsequent texture projection. The models obtained from the multi scalar survey were then developed for rendering in intensive 3D application development environments, initially subjecting them to a process of selective decimation of the level of detail to make their vision fluid and then subjecting them to a process of texture baking to not lose their realistic output or the understanding of the contents.

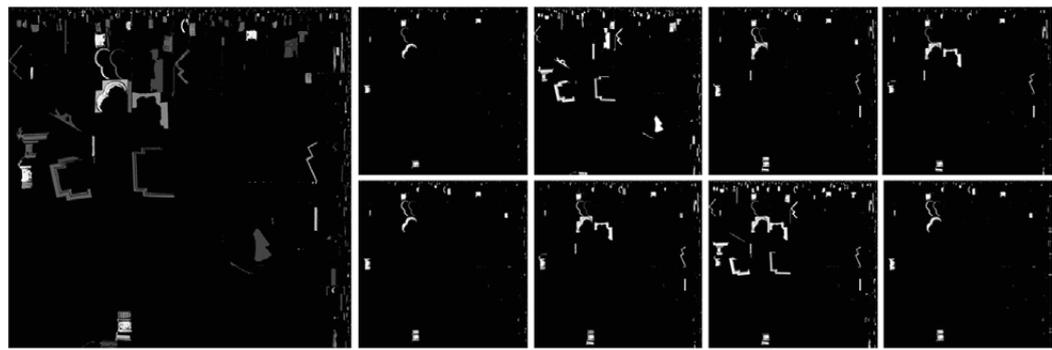
To link 3D models to the AI, annotations have been added or rather questionable metadata that encodes the knowledge of domain experts independently of applications.

In relation to domain vocabulary, the art and architecture thesaurus that is a standard of the architecture world formalized by the Getty Institute as Linked Open Data, was chosen in a format designed for compatibility with triple RDF, a flexible and extensible graph structure. The AAT has allowed to solve lexical ambiguities and to be used as a link between different data sources to allow AI to efficiently cross-check information, such as coming from Wikidata.

In order to associate semantic concepts with the corresponding spatial forms, a method that uses the correspondence between 2D/3D space, coding the annotation in the form of maps has been developed. In particular, once analysed the 3D twin and identified the semantic concepts found in it, these were searched in the Getty thesaurus to recognize its unique id code in the domain. Subsequently, for each significant and present in the model term of the is realized a monochrome map in which white indicates the polygons detected for the given concept and black does not show any relevance. The map is applied to the digital representation as well as a texture. In this way the information related to the process of attribution of meaning to spatial forms is translated into a purely visual image [Cera 2018]. The possibility of using a grey scale allows to refine the quality of information where the gradations have different percentages of relevance. So, this method makes it possible to consider semantic maps produced by multiple domain experts obtaining a final map of their degree of agreement, calculating the average values for each UV coordinate (fig.1).

**single session annotation workshop**

2 restorers [Re], 2 architecture historians [Sa], 2 art historians [St], 2 surveyors [Ri], 2 geometry experts [Ge], 2 echnologists [Te], 2 designers [Pr]



	Ge1_C04_bmp	Ge2_C04_bmp	Pv1_C04_bmp	Pv2_C04_bmp	Re1_C04_bmp	Re2_C04_bmp	Ri1_C04_bmp	Ri2_C04_bmp	Sa1_C04_bmp	Sa2_C04_bmp	St1_C04_bmp	St2_C04_bmp	Te1_C04_bmp	Te2_C04_bmp
Ge1_C04_bmp	1,0000													
Ge2_C04_bmp	0,9819	1,0000	0,6434	0,6556	0,0173	0,2873	0,7790	0,6680	0,6143	0,9477	0,7012	0,7811	0,7090	0,8246
Pv1_C04_bmp	-	-	1,0000	0,6623	0,0188	0,2907	0,7796	0,6503	0,6603	0,9229	0,6846	0,6852	0,6851	0,8349
Pv2_C04_bmp	-	-	-	1,0000	0,4837	0,5276	0,4905	0,5467	0,5190	0,4627	0,6733	0,3816	0,3812	0,3858
Re1_C04_bmp	-	-	-	-	1,0000	0,0163	0,0174	0,8834	0,7053	0,9923	0,6818	0,8279	0,8283	0,8373
Re2_C04_bmp	-	-	-	-	-	1,0000	0,8833	0,0191	0,0330	0,0163	0,0110	0,0195	0,0196	0,0193
Ri1_C04_bmp	-	-	-	-	-	-	1,0000	0,1007	0,1313	0,0168	0,0711	0,0398	0,0396	0,1732
Ri2_C04_bmp	-	-	-	-	-	-	-	1,0000	0,8297	0,8853	0,8034	0,7268	0,7259	0,7440
Sa1_C04_bmp	-	-	-	-	-	-	-	-	1,0000	0,7038	0,5637	0,8274	0,8366	0,8332
Sa2_C04_bmp	-	-	-	-	-	-	-	-	-	1,0000	0,6864	0,8247	0,8348	0,8340
St1_C04_bmp	-	-	-	-	-	-	-	-	-	-	1,0000	0,5654	0,5654	0,5719
St2_C04_bmp	-	-	-	-	-	-	-	-	-	-	-	1,0000	0,9919	0,9831
Te1_C04_bmp	-	-	-	-	-	-	-	-	-	-	-	-	1,0000	0,9893
Te2_C04_bmp	-	-	-	-	-	-	-	-	-	-	-	-	-	1,0000

Fig. 1. Semantic maps produced by different annotators for the concept of capital. Calculation of the degree of agreement with its final map.

The innovative thing of coding and using semantic maps lies in allowing to manage annotation margin of error, which is almost always ignored in the usual processes of semantic segmentation of digital representations. The margin of error of domain experts is, on the contrary, a substantial element in the knowledge process where it provides knowledge and complex cognitive mechanisms. For example, using this procedure, the margin of error due to the annotator's background, is not only recorded but also turned into a resource.

Machine learning approaches, in fact, are based on statistical models that, in this case, should model the probability of each geometric element belonging to a given category. Discretizing, in representation, is equivalent to removing information from the data on which the model is build, thus imposing a 'filtered' view to the algorithm, which has no way of modelling the existence of concepts which, for example, fade into each other or whose definition depends on multiple factors, such as the specialization of the domain expert in charge of the annotation, the aim of segmentation, the support on which the recognition process is implemented, etc. [Cera 2019].

## From Survey to Development of Artificial Intelligence

The semantic maps gathered with AAT codes, make the information contained in the map cross-referenced with that contained in the other resources annotated such as in the texts, in the AAT itself and in other LODs.

To make access to information fast and efficient for interactive applications that use real-time 3D material, knowledge has been depicted within a graph database [Webber 2012], which drastically reduces latency due to querying online resources, for example in RDF format. This allows to quickly cross-check information from different sources and compare it to adequately support the application.

Within a set of reference texts, the same concepts, described in geometry by semantic maps have been identified and annotated. In this way, you can associate, with the text that describes a resource, the geometry to which it refers independently of the application, making the material highly reusable for different purposes.

One of the possible applications achievable with the type of annotated material is the development of conversational virtual agents placed in an environment about which they have sufficient knowledge to interact with them. To study the behavior that these agents should take, a corpus [Origlia 2018] of 12 hours of audiovisual material was collected to document the behavior of art historians who illustrate the environments of the charterhouse di San Martino to small groups of visitors.

A linguistic and psychological annotation system has been created to cross-check the various levels of communication through which an experienced human transmits cultural content.

In the laboratory, motion capture data was collected to map human movements to 3D avatars. The logic of managing the gestures of the virtual agent has been defined as follows: at each frame, the system calculates the position to be assumed on the basis of a series of animations that are combined according to a series of parameters. As far as the gestures of the arms are concerned, there is a dedicated state machine which places the agent in a 'neutral' position. When an externally produced signal arrives, which corresponds to the start of an audio containing a synthetic voice, the agent switches to 'talking' mode. During speaking mode, an external system may require highlighting concepts with varying degrees of 'strength' or pointing in a certain direction. Since the location of the virtual agent is known, the only information you need to control its gestures is the location of the target.

Using the centroid of the mass of points labeled with a certain concept, for example "altare maggiore" imagining that the virtual agent is placed in the church of San Martino, it is possible to calculate the angle between the virtual agent and the concept that you want to point, thus providing the animation control system with the missing information to produce coherent deictic gestures.

The processing pipeline that allows an AI to interact is made of several modules.

First of all, a specially trained neural network transcribes audio containing a user's voice. From this transcription, an 'intent' is extracted, that is the abstract intention of the speaker and any parameters that detail the request. Based on intent and parameters, a graph database query is produced to extract the content needed to fulfill a request. The sentence to be synthesized is then passed to a second neural network that synthesizes the audio and produces the accompanying information, such as the phonetic annotation of the audio file, to allow the control of lipsync, and the indication of the temporal position of the expressed concepts, to control the gestures of the deictics. Based on this information, an interaction management engine delivers the presentation in real time.

## Results

The research developed as part of the CHROME project provided an opportunity to investigate the increasingly structured interconnections between the field of representation and survey and the themes of information technology. In particular, the paper analysed the role that the specification of architectural survey and the forms of drawing play in the development of AI applications tested in the dissemination of cultural contents related to some architectural

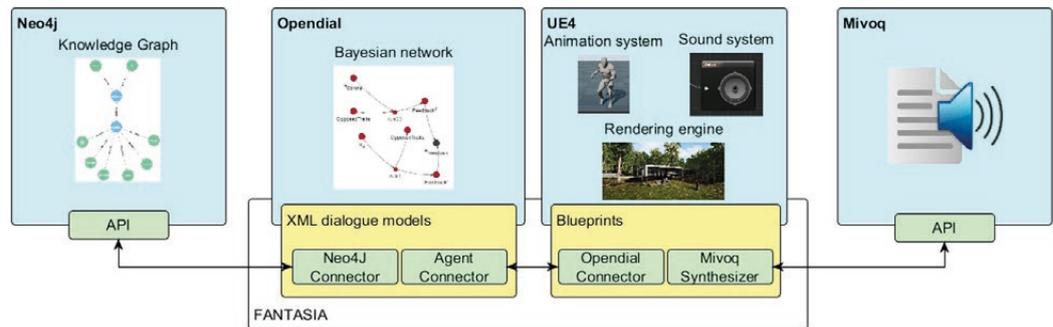


Fig. 2. The original FANTASIA architecture. An updated version was used in this work.

heritage of Campania. Validated on the case studies of the three charterhouses, the project developed a method of collection, analysis and dissemination of spatialized information resources around three-dimensional architectural models, used in digital environments whose presentation is entrusted to virtual agents modelled on human behavior:

CHROME's system architecture is designed to be generalized in a framework called FANTASIA [Origlia 2019] for developing conversational virtual agents that can be applied in any museum environment and therefore replicable (fig.2). The architecture uses graph databases to link data from different sources such as LOD, three-dimensional models, or other. It enables the use of modern peripheral devices and third-party services for capturing and analyzing input signals and integrates probabilistic decision-making systems for controlling interaction in 3D environments.

#### Notes

[1] The PI of the Italian PRIN project CHROME #B52F15000450001 is prof. F. Cutugno. The architecture unit was coordinated by profs. M. Campi and A. di Luggo. Arch. D. Iovane worked on architectural data acquisition together with arch. V. Cera who developed the research on semantics. Dr. A. Origlia worked on the A.I. development.

#### References

- Cera Valeria (2019). *La significazione digitale dell'elemento architettonico: dal rilievo alla strutturazione semantica dell'architettura*. Napoli: Editori Paparo.
- Cera Valeria, Origlia Antonio, Cutugno Francesco, Campi Massimiliano (2018). Semantically Annotated 3D Material Supporting the Design of Natural User Interfaces for Architectural Heritage. In *2018 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage*, 2091, pp. 1-4.
- Origlia Antonio, Cutugno Francesco, Rodà Antonio, Cosi Piero, Zmarich Claudio (2019). FANTASIA: a framework for advanced natural tools and applications in social, interactive approaches. In *Multimedia Tools and Applications*, 78 (10), pp. 13613-13648.
- Origlia Antonio, Savy Renata, Poggi Isabella, Cutugno Francesco, Alfano Iolanda, D'Errico Francesca, Vincze Laura, Cataldo Violetta (2018). An audiovisual corpus of guided tours in cultural sites: Data collection protocols in the chrome project. In *2018 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage*, 2091, pp. 1-4.
- Webber Jim (2012). A programmatic introduction to Neo4j. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*. New York: Association for Computing Machinery, pp. 217-218.

#### Authors

Massimiliano Campi, Dept. of Architecture, University of Naples Federico II, campi@unina.it  
 Valeria Cera, Dept. of Architecture, University of Naples Federico II, valeria.cera@unina.it  
 Francesco Cutugno, Dept. of Electrical Engineering and Information Technology, University of Naples Federico II  
 Antonella di Luggo, Dept. of Architecture, University of Naples Federico II, antonella.diluggo@unina.it  
 Domenico Iovane, Dept. of Architecture, University of Naples Federico II, domenico.iovane@unina.it  
 Antonio Origlia, Dept. of Electrical Engineering and Information Technology, University of Naples Federico II, antonio.origlia@unina.it

