

Eliciting Image Schemas for Urban Digital Twins

Maria Rosaria Stufano Melone¹, Stefano Borgo¹, Domenico Camarda² and Stefano De Giorgis³

¹Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy

²Politecnico di Bari, Italy

³Vrije Universiteit Amsterdam, Netherlands

Abstract

The construction of urban digital twins (UDT) is a broad objective that, to be successful, must go beyond structural and functional information to include the human factor with its cognitive and experiential knowledge. In this paper, we investigate this issue by focusing on a typical, yet less complex, element of the urban scenario, namely the square. More specifically, our general aim is to study how human knowledge of the square may be elicited and structured for inclusion in UDTs. To do this in a fairly broad perspective, we study which knowledge characterises squares by considering two quite different real cases: Navona square in Rome and Djemaa el-Fna square in Marrakech. Starting from three images of each square, we use an LLM to obtain a descriptive text which is then used to generate a knowledge graph. Both the text and the knowledge graph are then provided again to the LLM to elicit relevant image schemas. The results on each square are then compared to select the image schemas that are more characteristic of squares. The hope is that this kind of effort can lead to enrich our knowledge of urban spaces and, subsequently, to develop UDT that are more comprehensive representations of our cities by including along with structural and functional information also human perspectives and values.

Keywords

Urban planning, architecture, creativity, applied ontology, image schema

1. Introduction

The study and construction of Digital Twins (DT) is a live and central research goal of the last 20 years or so [13] which has attracted attention also in urban planning [1]. While characterizing what cities are is still an ongoing effort [2] and there are doubts about the scope and use of DTs for urban areas, the so called Urban Digital Twins (UDT) [9], it remains important to study these representation systems and to explore how to enrich them.

This paper presents a step within a longer research project aiming to analyse how to understand and exploit Urban Digital Twins. Our focus goes beyond the structural and functional aspects that are typically addressed in DT studies, to concentrate on how to build DTs that comprise the cognitive and experiential perspective typical of humans. We address this problem by focusing on a common, if not typical, element of cities, namely the square. This focus is not a restriction in interest but a way to deal with the complexity of the goal without constraining the generality of our project.

In this contribution, we posit the question of how to elicit cognitive-based information in a way suitable for enriching DTs. Generally speaking, we believe that the elicitation of this kind of information should be based on cognitive representation approaches like the image schemas. However, extracting and analysing image schemas on a varied and complex structure, which includes both static and dynamic information, is a particularly complicated process which is too difficult and error prone to be done manually. For this reason, we investigate the use of AI software over a set of sources (text, graphs, images). The results of this analysis highlight the variety and utility of cognitive-centred knowledge and the kind of knowledge that today's structural and functional UDTs clearly lack.

Proceedings of the Joint Ontology Workshops (JOWO) - Episode XI: The Sicilian Summer under the Etna, co-located with the 15th International Conference on Formal Ontology in Information Systems (FOIS 2025), September 8-9, 2025, Catania, Italy

✉ mariarosariastufanomelone@cnr.it (M. R. Stufano Melone); stefano.borgo@cnr.it (S. Borgo); domenico.camarda@poliba.it (D. Camarda); s.degiorgis@vu.nl (S. De Giorgis)

ORCID 0000-0002-5846-5738 (M. R. Stufano Melone); 0000-0001-6001-2765 (S. Borgo); 0000-0001-6311-3289 (D. Camarda); 0000-0003-4133-3445 (S. De Giorgis)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This work should be read within our previous efforts on the application of different research paths to provide a rich, unified and consistent framework for the modelling of complex environments like the urban territory [10]. More specifically, we focus on the study of two approaches, namely applied ontology and image schema, to support decision-making in urban and territorial planning [9, 11]. An UDT is a powerful tool to help in managing and monitoring some aspects of cities, but it presents criticalities and limitations of which the research community is not yet well aware [12].

In the literature there are only a few software solutions for managing image schemas that could produce outputs useful for our goal, i.e., knowledge that comes into play during planning decision processes or in scenario testing. Within this research line, there are some works on pipelines for extracting image schema from vision to support robots to operate more efficiently within spatial contexts [5]. In some cases a connection to ontology is proposed but the aim is to provide a method to better represent the semantics. De Giorgis [3] has developed a script to extract and identify image schemas from a text or image, an approach that might be relevant also for architectural and city design features.

Overall, our approach seeks to incorporate nuanced and dynamic elements of spatial experience and everyday urban life, extending beyond conventional sensor data and standard environmental-geographical information. The focal point of our recent work has been the urban square - a complex spatial node characterized by diverse patterns of movement and occupation, ranging from habitual pedestrian routes that have evolved through repeated use, to the qualities of light that characterize the square atmosphere at particular hours of day.

By combining the methodologies of ontological analysis and applied ontology with the reading of places given by image schemas we would like to develop new tools that can manage heterogeneous knowledge. We think that the use of image schemas to “reverse engineering” human ways to inhabit and understand places can lead to develop better representation tools for the urban planner, making her more aware of intrinsic cognitive aspects and dynamics that make spaces and places unique to their inhabitants.

As anticipated, in this paper we focus on urban squares because they are important spaces within cities whose meaning goes clearly beyond their structure, are hard to characterize, and yet reflect the habits and social behaviours of city inhabitants. A city square is often the place where meeting and crossing events occur—one of the semantically densest places in the urban fabric, where diverse meanings, events, and objects create a richer substance [2]. While spatial, structural and functional information about squares can be extracted from a variety of existing traditional resources like city maps and GIS repositories, the information about human perception and cognitive understanding of squares is not readily available and even standard techniques (like questionnaires, behavioural studies and the like) are hard to implement and analyse. For this reason we focus on image schemas, understood as cognitive tools, to establish connections between cognitive representations and lived spatial or functional interactions [7]. Image schemas represent semantic frameworks that develop from repeated perceptual experiences, functioning at a more abstract level than concrete mental imagery. These cognitive structures demonstrate the capacity to organize and pattern unlimited numbers of perceptions, images, and experiences [6, 8]. To increase generality, we elicit image schemas from distinct scenarios represented via photos taken at different times and during different activities. The aim is to acquire descriptions of the very same place how it changes during times and usage. The analysis of these photos is done via an LLM (Perplexity) and the Text2AMR2FRED tool [4]. The use of large language models (LLMs) to articulate the experiential dimensions of urban space constitutes an emerging frontier for capturing and enriching knowledge about places and the events that occur within them. The application of LLMs to the elicitation of image schemas is conceived as a preliminary experiment, aimed at facilitating a broader and more rapid process of elicitation. Although not the sole method available for eliciting image schemas, it represents a potentially significant and promising line of inquiry, and it may complement other elicitation techniques such as computer vision and spatial analysis, which are not addressed in this contribution. FRED (Frame-based Entity-Relation Description) is a semantic tool that transforms natural language text into structured knowledge graphs using RDF/OWL formats. The

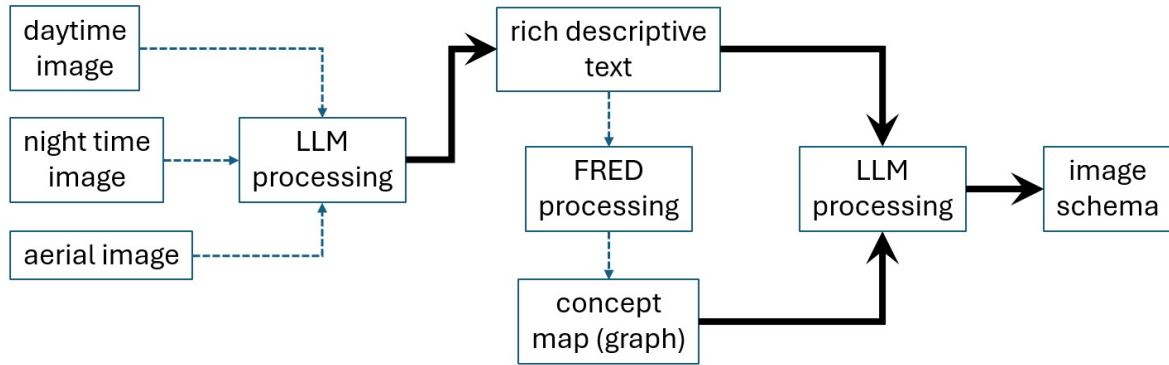


Figure 1: Information flow for the elicitation of image schemas of the Navona and Djemaa el-Fna squares.

AMR-to-FRED translation facilitates knowledge graph enrichment, references WordNet, DBpedia, and DOLCE-Zero, and combines syntactic and semantic parsing with ontologies (such as DBpedia, WordNet, FrameNet) to extract: entities (e.g., people, objects, places), events (e.g., actions), and relations (e.g., who did what, where, and when). FRED is used in fields like the semantic web, AI, digital humanities, and cognitive robotics to enable machines to understand and reason about text in a human-like way.

Following this introduction, §2 briefly describes the tools we use while the following, §3, uses them to study two paradigmatic squares and elicit the image schemas that characterise them. The last section, §4, draws some further conclusions.

2. Setting the approach

Our proposal is to investigate a three-stage procedure to systematically transform visual spatial information into image schemas. This is the strict aspect of interest we intend to explore in this contribution, a partial slice in a broader research scenario. Image schemas can be considered a format for formal knowledge representation, i.e., a way to capture knowledge that is suitable for enriching UDTs (e.g., in the form of knowledge graphs).

The pipeline we use is depicted in Fig. 1. We decided to start from a set of photos of the urban space of interest because these are rich sources of heterogeneous information and overcome, at least to some extent, some constraints of verbal language (like imposing the choice of terminology and enforcing a structural linearity of knowledge), so that these aspects intervene only later in the elicitation process and are mitigated by the integration of knowledge from large language models, knowledge graphs, and ontological structures (even though the latter is not really exploited in this paper). The process begins with the selection of photographic perspectives of the spatial environment of interest. In the next step we submit these to a large language model for the generation of comprehensive textual descriptions. These descriptions are aimed to capture not only the visible architectural and environmental features but also the implicit functional and experiential qualities of the space.

The resulting textual output is processed through the AMR2FRED tool, which converts natural language descriptions into structured knowledge graphs. This transformation represents a crucial transition from the contextual, distributional representations characteristic of neural language models to the explicit, relational and conceptual structures required for formal reasoning. The generated graph provides a structured representation of the spatial relationships, entities, and properties identified in the original descriptions.

The next stage involves the analysis, again via the large language model, of both the textual description and of the knowledge graph to extract relevant image schemas in their role of fundamental spatial and experiential patterns motivated by embodied cognition theory. For instance, the generated description of an instance of “walking” is now formally connected to the source-path-goal image schema, while the

spatial configuration of filled and empty areas is now mapped to container and support image schemas as a cognitive counterpart.

The integration of these ‘ontologized’ (thanks to FRED) image schemas serves a dual function: they provide formal constraints that guide subsequent language model generation while simultaneously grounding the abstract spatial concepts in established patterns of human spatial cognition. This creates a hybrid computational system that combines the contextual understanding capabilities of large language models with the inferential rigour of formal knowledge representation, resulting in a framework capable of representing and reasoning about implicit spatial knowledge through explicit formal structures.



Figure 2: Credits to [15, 16, 17]



Figure 3: Credits to [18, 19, 20]

3. A comparison between Navona and Djemaa el-Fna squares

Below we describe the procedure we followed for the elicitation of image schemas related to the two squares.

The procedure, as illustrated in Fig. 1, consists in providing to the LLM Perplexity¹ three photos (daytime image, nighttime image, aerial image) of each square considered for the comparison: Navona square and Djemaa-el-Fna square. You see here the pictures we used, figure 2 the ones concerning Piazza Navona, in figure 3 the ones about Djemaa-el-Fna. [figure 2] [figure 3]

The descriptive text generated by Perplexity was provided to AMR2FRED, which transformed the text into a graph. Subsequently, both the graphs and the original texts were provided to Perplexity to elicit the image schemas related to each square under examination. The procedure involved comparing the resulting image schemas derived from text and graph representations of both Navona and Djemaa-el-Fna, and ultimately determining which image schemas better characterize each square in order to acquire the needed variety that a general DT for urban squares should contain.

3.1. Image schemas in Navona square

The analysis of the graph produced by FRED regarding the flows and spatial relations within Navona square, allows us to identify different image schemas to support the understanding of space, movement and relations between elements in the urban space. As a result of this step, the main emerging image schemas are:

¹<https://www.perplexity.ai/>

- PATH: the lines drawn in the graph suggest paths and directions, representing the movement through the square.
- CONTAINER: the square appears as a delimited space that collects events, people and movements.
- LINK: connections emerge between distinct points of the square, visible in the lines that join the nodes.
- CENTER-PERIPHERY: where elements are concentrated in central areas and more marginal zones, the distinction between center and periphery emerges and is clear.
- FORCE: curves and directions of the graph suggest dynamic forces, such as flows of people or spatial attractions.
- BOUNDARY: the limits of the square are marked, distinguishing the inside from the outside.

The comparison between the graph and a descriptive text of the square (of a historical, social and architectural nature) makes evident some substantial differences. The graph makes immediately visible the patterns related to movement and spatial structure, such as PATH, AXIS, CONTAINER and BOUNDARY. The text, on the other hand, emphasizes the lived experience, the social function and the narration of the space, bringing out patterns such as CONTAINER, AXIS, CENTER-PERIPHERY, LINK and FORCE, with the PATH being less evident. When the two approaches – visual and textual – are combined, some main image schemas emerge, here ordered by representativeness according to a weighted estimate that gives equal weight (50/50) to their evidence in the graph (in terms of recurrence, spatial centrality and visual structure) and to their relevance in the text (based on frequency, narrative role and descriptive function):

1. PATH – dominant pattern (30%), since it reflects the longitudinal character of the square and the strong presence of crossing flows.
2. CONTAINER (25%) – the square is perceived as a large enclosed space that hosts events and interactions.
3. AXIS (15%) – the central axis, marked by the arrangement of the fountains, is a structuring element.
4. CENTER-PERIPHERY (12%) – the square has a clear center (the Fountain of the Four Rivers) that attracts activity, contrasted with less active edges.
5. LINK (8%) – the relationships between the architectural and social elements define a network of connections.
6. BOUNDARY (5%) – the physical limits, given by the buildings that surround the square, delimit the urban container.
7. FORCE (5%) – the dynamic forces are less constant but present - for example in the movement of the crowd or in the water that invades the square in summer.

Assuming a single image schema could be used to best summarize Navona square, the most likely choice would be PATH, which has an estimated probability of 40%, followed by CONTAINER (30%) and AXIS (15%). This estimate is based on a criterion of relative representativeness, which proportionally combines the visual centrality of the image schema in the graph (structural presence, directionality, frequency) and its semantic relevance in the text (narrative function, recurrence, descriptive role). The other schemas appear less central in comparative terms. A mutual hierarchical evaluation (pairwise) between the image schemas confirms this preference: PATH is dominant compared to all the others. Therefore, in decreasing order of importance, other schemas follow: CONTAINER, AXIS, CENTER-PERIPHERY, LINK, BOUNDARY and finally FORCE. As a final note, the VERTICALITY image schema was not included among the main ones because, although vertical elements are mentioned in the text (the obelisk, the dome of the church), the graph representation and the experience of the square appear predominantly horizontal. There are no dominant vertical differences or developments in the perception of the urban space, which remains strongly axial and traversable along a longitudinal direction.

3.2. Image schemas in Djemaa El-Fna square

The graph of interactions and flows relative to the Djemaa el-Fna square highlights multiple image schemas that can shape the possible experience of the urban space. Among these, the dominant image schema is undoubtedly PATH, highlighted by the multiple lines that represent movements, paths and flows of people and activities. Alongside this, the CONTAINER image schema emerges, which describes the square as an enclosed space, with well-defined areas within it. The connections between elements - stalls, shows, thematic areas - reflect the LINK image schema, which underlines the relationships between actors and spaces. The presence of a particularly active central area, from which the activities branch out towards the outside, recalls the CENTER-PERIPHERY image schema. Finally, the intertwining and simultaneous coexistence of multiple activities and paths brings out the concept of MULTIPLICITY, typical of chaotic and dynamic spaces like this one. Comparing this analysis of the graph with the text analysis significant differences emerge in the perception of the same image schema. In fact, the text transmits information in a narrative and sequential way, forcing the reader to mentally reconstruct the spatial relationships. However, even in the text different image schemas are clearly recognizable. The most evident is once again CONTAINER: the square is described as the heart of the medina, a space that contains activities, people, noises, lights and transformations. CENTER-PERIPHERY follows, since the square is the focal point from which the souks and streets of the city branch out, and around which social and symbolic activities revolve. The PATH image schema is implicit in the movements of visitors who cross the spaces, passing from one function to another, while LINK shows up in the relationships between functions: trade, entertainment, food and rituality. More subtle patterns can then emerge, such as SOURCE-PATH-GOAL, which is recognized in the daily transformation of the square – from a daytime market to an evening theater – and UP-DOWN, suggested by the presence of the minaret that dominates from above. Finally, BALANCE/FORCE can be perceived in the tension between order and chaos, between structure and improvisation. Combining graph and text, it is possible to order the image schemas by degree of representativeness. The most important seems to be CONTAINER, which collects about a third of the occurrences and reflects the essence of the square as a space that welcomes, contains and organizes. PATH and the related SOURCE-PATH-GOAL follow, which together represent a quarter of the spatial dynamics, describing the flow of bodies, goods and meanings. The CENTER-PERIPHERY pattern covers about 18% of the occurrences, testifying to the central position of the square in urban and symbolic life. Less representative but still relevant are LINK, UP-DOWN and BALANCE/FORCE, which complete the layered perception of space. Assuming that only one image schema is to be identified to represent Djemaa el-Fna square, CONTAINER would have the highest probability of being chosen, equal to 40%, because of its double evidence, visual and narrative. PATH/SOURCE-PATH-GOAL follows with 25%, and CENTER-PERIPHERY with 15%. The others would be placed below 10%. A mutual comparison between the image schemas (according to a “one against one” logic) confirms this hierarchy. CONTAINER prevails over all the others due to its spatial and conceptual centrality. PATH is placed second, which is more evident than LINK, CENTER-PERIPHERY and UP-DOWN. The latter, linked to the vertical perception given by the minaret, is present but secondary. BALANCE/FORCE, finally, is the most transversal and least structuring schema. As for the theme of VERTICALITY, it is present but not central. Although the Koutoubia minaret visually dominates the square, its function remains symbolic rather than organizational. Neither the text nor the graph seem to suggest a vertically hierarchical spatial or functional structure. Djemaa el-Fna is perceived and experienced as an essentially horizontal, open, dynamic and reticular space. In short, Djemaa el-Fna appears to be configured as a container space, crossed by multiple paths, the vital center of the city and a point of connection between activities, functions and urban identities. The main image schemas not only help to describe it, but also to understand its symbolic power and its role in the social space.

3.3. Analytical comparison between Navona and Djemaa el-Fna squares

The CONTAINER image schema is strongly represented in both squares, but with different nuances. In Navona square, the notion of container is shown through the clear architectural boundaries and

Table 1

Comparison of the image schemas for Navona and Djemaa el-Fna squares (the percentages are an estimate of the relative representativeness of each image schema with respect to the total of the schemas identified for each square: the higher the percentage, the more central the schema in the cognitive structure of the space; percentages are calculated by the machine).

Image Schema	Navona square	Djemaa el-Fna square
PATH	Central (30%); dominant in the graph	Central (25%); strong in the graph and text
CONTAINER	Very relevant (25%); closed space	Dominant (30%); containing heart of the urban life
AXIS	Strong (15%); longitudinal axis	Absent or secondary; reticular structure
CENTER-PERIPHERY	Important (12%); central node	Relevant (18%); urban center of the medina
LINK	Secondary (8%); connections between elements	Present (15%); relationships btw activities
BOUNDARY	Marginal (5%); physical limits	Not relevant; borders less emphasized
FORCE	Episodic (5%); crowds, water	Present (5%); order/chaos equilibrium
MULTIPLICITY	[not present]	Present; intertwining of activities
SOURCE-PATH-GOAL	Implicit; hinted at in the text	Explicit; day-night transformation
UP-DOWN / VERTICALITY	Only symbolically present	Present (7%); minaret is dominant
BALANCE	[not present]	Present (5%); dynamism and conflict

the role of the square as a closed and organized urban space. The text describes it as a “large urban theater”, accentuating its aesthetic and scenographic function. In Djemaa el-Fna, however, the concept of container takes on a more dynamic and social value: the square is described both in the graph and in the text as a place that welcomes and gathers in a fluid way a multiplicity of activities - from commerce to shows - and is perceived as the beating heart of the medina. For this reason, while in Navona the container is more static and architectural, in Djemaa el-Fna it is alive and transformative. The concept of AXIS is instead clearly present only in Navona square. Here the longitudinal axis is clearly perceptible both in the visual arrangement (graph) and in the textual description, which highlights the ordered sequence of the three main fountains along a central path. In Djemaa el-Fna, on the contrary, this image schema is practically absent. The Moroccan square is not organized along an axis, but shows up as a reticular and polycentric space, in which the activities are distributed without a prevailing orientation. Concerning CENTER-PERIPHERY, both urban spaces present a significant centrality, but with different functions. In Navona square the center is symbolically represented by the Fountain of the Four Rivers, which acts as a visual and architectural node. In the graph, this centrality is underlined by the greater density of connections that start from this point. In Djemaa el-Fna, instead, the centrality is essentially functional and social: the square constitutes the heart of the medina, the nerve center of urban life, from which the streets branch off towards the souks and other neighborhoods. This centrality is less linked to a single element and more to the overall function of the square.

The image schema LINK is present in both squares, but with different intensity. In Navona, the connections are mainly spatial and architectural: they are relationships between fountains, churches and other landmarks, often visually aligned. In Djemaa el-Fna, on the other hand, the links are dense and varied: connections between different functions (market, entertainment, gastronomy), social, cultural and symbolic relationships. In this case, the square acts as a node of exchange and interaction between different languages and practices, not only as a physical space.

The notion of BOUNDARY is more evident in Navona square. Its limits are marked by baroque buildings and historic palaces, which frame the space in a clear way. In the graph, the edges are clearly represented, while the text underlines the defined perimeter of the square. On the contrary, in Djemaa el-Fna, the boundaries are more fluid, permeable and less emphasized. The square merges with the souks and adjacent streets, and there is no clear separation between “inside” and “outside”.

The image schema SOURCE-PATH-GOAL is particularly present in the textual description of Djemaa el-Fna, and almost absent in that of Navona square. The Moroccan square is described as a space that

Table 2

Comparative hierarchical synthesis of the main image schemas of the two squares, ordered by decreasing importance (the distinctive features are accompanied by weighted estimates based equally on the visual evidence in the graph and on the narrative relevance in the text; percentages are calculated by the machine).

	Navona square	Djemaa el-Fna square
1. Dominant schema	PATH (30%)	CONTAINER (30–40%)
2. Second schema	CONTAINER (25%)	PATH / SOURCE-PATH-GOAL (25%)
3. Third schema	AXIS (15%)	CENTER-PERIPHERY (18%)
4. Other significant schema	CENTER-PERIPHERY (12%)	LINK (15%)
5. Verticality	weak, ancillary	present, symbolic (7%)
6. Structure	axial, ordered	reticular, chaotic
7. Temporality	static	transformative (day-night)
8. Context	European, baroque, scenic	Arab-Berber, dynamic, ritual

transforms during the day: a market in the morning, a theater in the evening. This transformation could suggest a functional and temporal path that starts from an initial condition (source), goes through a change (path), and reaches a final goal (goal), such as the experience of a show or a collective ritual. Navona square, on the contrary, maintains a greater stability of function and structure.

As regards the UP–DOWN image schema, linked to verticality, here too a contrast is highlighted. In Djemaa el-Fna the minaret of the Koutoubia mosque visually “dominates” the square, introducing an evident, albeit symbolic, vertical dimension. This aspect is recognized both in the text and in the spatial perception of the place. Navona square, on the other hand, features vertical elements – such as the obelisk of the central fountain or the domes of the churches – but these do not seem to structure the perception or use of space, thus appearing accessories to the prevailing horizontal configuration.

The image schema BALANCE/FORCE is more strongly perceptible in Djemaa el-Fna. The square is a dynamic place, where order and chaos, stability and improvisation coexist. The text highlights the simultaneous presence of organized structures (such as markets) and more spontaneous forms of sociality (shows, crowds). This tension between opposing forces represents a precarious but vital balance, which characterizes the identity of the place. In Navona square, however, this dimension is less marked. The movement seems more regulated, the spatial composition more harmonious and contained.

Finally, the image schema MULTIPLICITY seems to be absent in Navona square, but central in Djemaa el-Fna. The Moroccan square simultaneously hosts a plurality of events, people, languages and gestures. This multiplicity is perceptible both in the graph, which shows a dense and polycentric network, and in the textual description, which describes the square as a kaleidoscope of activities. Navona square, while offering different functions (tourism, art, sociality), appears to be experienced in a more sequential and orderly way, along a traversable axis. Therefore, multiplicity appears to be a structuring quality in Djemaa el-Fna, but only an accessory quality in Navona.

4. Conclusions

This paper presents an approach to the use of image schemas as a methodological contribution to elicit cognitive and human knowledge and make it accessible to UDTs for planning and urban decision-making. The approach is preparatory to the integration of structural and functional information with personal and cognitive information that together characterize urban spaces.

Our experiment concentrated on two different squares and revealed that while both share certain fundamental image schemas (such as container and path), their hierarchical organization differs significantly in ways that reflect deep cultural and functional logic. Navona square emerges as a space primarily structured by the path scheme (30%), probably reflecting the longitudinal character and the Baroque tradition of organized space sequences, with a prominence of the AXIS image schema (15%) which underlines an emphasis on directional clarity and hierarchical organization. The function of the square as a well-defined CONTAINER (25%) suggests architectural traditions that favour the

creation of borders and the scenic framework. On the contrary, Djemaa El-Fna reveals a more characterized space logic characterized by the CONTAINER image schema (30-40%) but in a dynamic and transformative sense-probably reflective urban traditions of adaptable public space, with SOURCE-PATH-GOAL patterns (25%) that capture the temporal transformations of the 'market square' from market to performative space.

The comparative analysis suggests direct implications for planning and decision-making processes, revealing the importance of an urban design sensitive to cultural differences in space cognition and social organization. For example, the presence of the MULTIPLICITY image schema in Djemaa El-Fna which is absent in Navona square can indicate how different urban traditions organize complexity and social interaction. This suggests that planning approaches that traditionally emphasize order and hierarchy can be inadequate for different contexts and communities. The time dimension captured through the SOURCE-PATH-GOAL image schema pattern also suggests that successful public spaces could be designed for adaptive use rather than a fixed function - a particularly relevant principle for today's urban contexts characterized by rapid changes. The intuition we explored in this contribution is this that in the context of Urban Digital Twins (UDTs), Image Schema information could serve as a cognitive-semantic layer that enriches the representation of spatial and experiential aspects of the city. For example, schemas such as CONTAINER or PATH may be used to structure how movement, accessibility, or boundaries are modeled in digital replicas of urban environments. Integrating such schemas could improve the interpretability of simulations, support citizen-centered analyses of urban space, and complement data-driven techniques by grounding them in embodied conceptual structures together with a knowledge analysed and represented via applied ontology. A follow-up could involve the extraction of image schemas directly from the participation of a space's users or the inhabitants of a place (a direct form of knowledge from which it is often difficult to extract consistent knowledge, as mentioned). The research evokes important perspectives for applying the Urban Digital Twin technologies beyond their current limitations in capturing cognitive and social dimensions, offering a path towards more knowledge aware and inclusive urban modelling systems. The methodological potential for simulating future scenarios could allow decision makers to explore not only future physical impacts but also cognitive and social impacts - so matching very crucial emerging needs nowadays. As a model of knowledge management, this hybrid approach looks particularly suitable to be integrated into system architectures to support the interactive and real-time urban decision, allowing systems that adapt dynamically to upcoming events and occurrences while maintaining methodological consistency. Recent contributions demonstrate how to build a novel framework that bridges embodied cognition theory and agent systems by leveraging a formal characterization of image schemas [14]. By customizing LLMs to translate natural language descriptions into formal representations based on these sensorimotor patterns, can be created a neurosymbolic system that grounds an agent's understanding in fundamental conceptual structures [14]. In the future we aim to consolidate this pipeline, which seems to be stable as our simulation with purely textual descriptions suggests, in order to extract combinatorics of image schemas for squares. We could then present these as framework for a cognition-based representation of squares to be included in today's UDT of these areas. As of today, whether this research line can be consolidated and extended to urban areas in general remains to be verified.

Declaration on Generative AI

During the preparation of this work, the author(s) used Claude 4 for: Grammar and spelling check, Paraphrase and reword; and Perplexity for: automatic image description generation. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] M. Batty. Digital twins. *Environment and Planning B: Urban Analytics and City Science*, 45(5):817–820, 2018.
- [2] S. Borgo, D. Borri, D. Camarda, and M. R. Stufano Melone. *An Ontological Analysis of Cities, Smart Cities and Their Components*, volume 36 of *Philosophy of Engineering and Technology*, chapter 18. Springer, Cham, 2021.
- [3] S. De Giorgis. *Ethics in the flesh: formalizing moral values in embodied cognition*. PhD thesis, Università di Bologna, 2023.
- [4] A. Gangemi, V. Presutti, D. Reforgiato Recupero, A. G. Nuzzolese, F. Draicchio, and M. Mongiovi. Semantic web machine reading with fred. *Semantic Web*, 8(6):873–893, 2017.
- [5] M. M. Hedblom. Image schemas: State of the art in spatiotemporal conceptualisation. *Image Schemas and Concept Invention: Cognitive, Logical, and Linguistic Investigations*, pages 33–51, 2020.
- [6] M. M. Hedblom, F. Neuhaus, and T. Mossakowski. The diagrammatic image schema language (DISL). *Spatial Cognition & Computation*, 25(2):138–175, 2025.
- [7] M. M. Hedblom, O. Kutz, and F. Neuhaus. Image schemas in computational conceptual blending. *Cognitive Systems Research*, 39:42–57, 2016.
- [8] G. Lakoff and M. Johnson. *Metaphors we live by*. University of Chicago press, 2008.
- [9] M. R. Stufano Melone, S. Borgo, and D. Camarda. Digital twins of cities vs. digital twins for cities. In A. Marucci, F. Zullo, L. Fiorini, and L. Saganeiti, editors, *Innovation in Urban and Regional Planning*, pages 192–203, Cham, 2024. Springer Nature Switzerland.
- [10] M. R. S. Melone, S. Borgo, and D. Camarda. Image schema and ontology-based rules to support planning activities: a study of the urban square. In *Proceedings of The Eighth Image Schema Day co-located with The 23rd International Conference of the Italian Association for Artificial Intelligence(AI*IA 2024)*, volume 3888. CEUR Workshop Proceedings, 2024.
- [11] M. R. S. Melone, S. Borgo, and D. Camarda. City interactions in urban planning: The square example from an ontological analysis point of view. In *International Conference on Computational Science and Its Applications*, pages 448–455. Springer, 2024.
- [12] M. R. Stufano Melone, S. Borgo, and D. Camarda. Digital twins facing the complexity of the city: Some critical remarks. *Sustainability*, 17(7):3189, 2025.
- [13] D. Wagg, K. Worden, R. Barthorpe, and P. Gardner. Digital twins: state-of-the-art and future directions for modeling and simulation in engineering dynamics applications. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 6(3):030901, 2020.
- [14] F. Olivier and Z. Bouraoui. *Grounding Agent Reasoning in Image Schemas: A Neurosymbolic Approach to Embodied Cognition*. arXiv preprint arXiv:2503.24110, 2025. Available at: <https://arxiv.org/abs/2503.24110>.
- [15] Aerial view of Piazza Navona, *Nome del sito web*, Available at: <https://www.istantidibellezza.it/piazza-navona.html>
- [16] Open space of Piazza Navona, Available at: <https://www.instagram.com/p/CIMJMdGhywW/>
- [17] Nighttime view of Piazza Navona, Available at: <https://www.viator.com/it-IT/Rome/d511-ttd>
- [18] Aerial view of Djemaa-el-fna, Available at: <https://www.marrakesch.com/sehenswurdigkeiten/djemaa-el-fna/>
- [19] Night view of Djemaa-el-fna, Available at: <https://www.marrakesch.com/sehenswurdigkeiten/djemaa-el-fna/>
- [20] Open space of Djemaa-el-fna, Available at: viaggiaescopri.it, <https://www.linkedin.com/pulse/impact-economique-du-coronavirus-quels-enseignements-pour-maktoum/?originalSubdomain=fr>