



HAL
open science

Un modèle hiérarchique pour les trajectoires sémantiques et l'extraction d'événements dans les espaces intérieurs et extérieurs

Hassan Nouredine

► **To cite this version:**

Hassan Nouredine. Un modèle hiérarchique pour les trajectoires sémantiques et l'extraction d'événements dans les espaces intérieurs et extérieurs. Modeling and Simulation. Université de Bretagne occidentale - Brest, 2022. English. NNT : 2022BRES0059 . tel-03888591

HAL Id: tel-03888591

<https://theses.hal.science/tel-03888591>

Submitted on 7 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

ÉCOLE DOCTORALE N° 598
Sciences de la Mer et du Littoral
Spécialité : *Géomatique*

Par

Hassan NOUREDDINE

**A hierarchical model for semantic trajectories and event extraction
in indoor and outdoor spaces**

Thèse présentée et soutenue à Brest, le 16 Septembre 2022
Unité de recherche : Institut de Recherche de l'École Navale (IRENav)

Rapporteurs avant soutenance :

Hua LU Professeur des universités, Université de Roskilde, Danemark
Karine ZEITOUNI Professeure des universités, Université de Versailles-Saint-Quentin-en-Yvelines, France

Composition du Jury :

Président :	Jean-Philippe BABAU	Professeur des universités, Université de Bretagne-Occidentale, France
Examineurs :	Marlène VILLANOVA-OLIVER	Maîtresse de conférences, University of Grenoble Alpes, France
Dir. de thèse :	Christophe CLARAMUNT	Professeur des universités, IRENav, France
Co-dir. de thèse :	Cyril RAY	Maître de conférences, IRENav, France

I dedicate this to my parents

Acknowledgements

My PhD journey has been full of ups and downs, a journey in which I was able to grow as an independent researcher. First, I would like to thank my thesis director Christophe Claramunt, my supervisor Cyril Ray, as well as the Ecole Navale, for offering me this opportunity to carry out this doctoral work.

I would also like to thank all the jury members for agreeing to be part of my PhD committee. Reviewing a PhD thesis is not a simple task, and I appreciate their time and feedback.

I am also grateful to all the researchers with whom we discussed my work during my PhD, especially Imad Afyouni and Alexander Artikis. I appreciate the feedback and support they gave me.

Most sincerely, I thank my family for being all the time with me; despite the distance between us, thank you for providing me with your unconditional love and support.

Finally, I would like to thank the friends and colleagues of the IRENav I have met during these four years, to name a few: Bilal, Joseph, Loic, Younes, José, Samuel, etc. I would also like to thank all the friends from the outside who supported me.

Ahead of all, thank God, without whom I would never have been able to get this far.

Contents

Acknowledgements	3
1 Introduction	10
1.1 Thesis Context	11
1.1.1 Mobility Data Modelling and Representation	12
1.1.2 Real-Time Event Detection From Trajectories	12
1.2 Research Challenges	14
1.3 Experimental Prototyping	15
1.4 Contribution	16
1.5 Dissertation Structure	17
2 Related Work	19
2.1 Indoor and Outdoor Spatial Models For Navigation, Route Planning and Path Search	20
2.1.1 Navigation and Route Planning	20
2.1.2 Path Search	20
2.2 Trajectory Data Models	21
2.2.1 Semantic Trajectories	22
2.2.2 Annotated Semantic Trajectories	22
2.2.3 Multi-granular Semantic Trajectories	23
2.2.4 Indoor and Outdoor Semantic Trajectories	23
2.3 Trajectory Database Engines	24
2.4 Trajectory Streams Processing	25
2.5 Complex Event Recognition Languages	25
2.5.1 Tree-based Models	25
2.5.2 Logic-based Models	26
2.5.3 Automata-based Models	26
2.5.4 Hybrid Models	26
2.6 Trajectories Complex Event Recognition	26
2.7 Synthesis and Positioning	27
3 Semantic Trajectory Data Model For Indoor and Outdoor Spaces	30
3.1 Introduction	31
3.2 Indoor and Outdoor Space Model	33
3.2.1 Spatial Modelling Principles	33
3.2.2 Formal Definitions	34
3.3 Spatio-temporal To Semantic Trajectories	36
3.3.1 Raw Trajectory Components	36

3.3.2	Semantic Trajectory Components	38
3.4	Multiple Views of Semantic Trajectories	41
3.4.1	Hybrid Trajectory Representation	42
3.4.2	Hybrid Trajectory Operations	44
3.4.2.1	Operation 1	44
3.4.2.2	Operation 2	46
3.4.2.3	Operation 3	47
3.4.3	Contextual-based Trajectory Representation	48
3.4.4	Contextual-based Trajectory Operation	49
3.4.4.1	Operation 4	49
3.5	Semantic Trajectory Data Implementation and Processing Experimentation	50
3.5.1	Data Set Description	50
3.5.2	Pre-processing Methodology	50
3.5.3	Semantic Trajectories Data Manipulations	52
3.5.3.1	Graph Queries Examples	53
3.5.3.1.1	Place-based	53
3.5.3.1.2	Trajectory-based	55
3.5.3.1.3	Hierarchy-based	56
3.5.3.2	Graph Analytics Examples	56
3.5.3.3	Performance Evaluation	58
3.5.4	Multiple Views Extraction Implementation and Experimentation	60
3.5.4.1	Neo4j Multiple Views Extension and Experimentation	60
3.5.4.2	Performance Evaluation	65
3.6	Discussion	66
3.7	Conclusion	67
4	Composite Event Extraction From Stream of Semantic Trajectories	70
4.1	Introduction	71
4.2	Stream of Semantic Trajectory	72
4.2.1	Semantic Trajectory Data Structure	72
4.2.2	Trajectory Segments Stream	74
4.3	Event Modelling	74
4.3.1	Basic Event	74
4.3.2	Composite Event	76
4.4	Event Pattern Language for Composite Event Recognition	77
4.4.1	Individual Events	77
4.4.2	Aggregated Events	78
4.5	System Architecture	81
4.6	Complex Event Processing over Semantic Trajectories Experimentation	83
4.6.1	Query Examples	83
4.6.1.1	Query 1: Individual Event Example	83
4.6.1.2	Query 2: Aggregated Event Example	86
4.6.1.3	Query 3: Aggregated Event Example on a Spatial Hierarchy	89
4.6.2	Performance	91
4.6.2.1	Throughput Evaluation	91
4.6.2.2	Latency Evaluation	93
4.7	Conclusion	94

5 Conclusion	96
5.1 Contributions	97
5.2 Limitations	99
5.3 Possible Research Extensions	99
5.4 Future Work	100

List of Figures

1.1	Thesis positioning in terms of the intersection of the main scientific thematics.	11
1.2	Research framework.	13
3.1	Spatial model showing the hierarchical relationships between indoor and outdoor spatial layers.	35
3.2	Trajectory, subtrajectory and continuous subtrajectory examples. . . .	37
3.3	Temporal abstractions from time series data.	39
3.4	Semantic trajectory represented as a graph.	41
3.5	Hybrid semantic trajectory.	43
3.6	Hybrid semantic trajectory (Example 5).	44
3.7	Hybrid semantic trajectory (Example 6).	46
3.8	Hybrid semantic trajectory (Example 7).	47
3.9	Contextual-based semantic trajectory (Example 9).	48
3.10	Pre-processing methodology.	51
3.11	Multi-dimension semantic trajectory graph meta-model implemented in Neo4j.	52
3.12	Visualisation example of a place-based query. The red point represents the place where a trajectory were at a specific time exposed to a high level of NO_2 . The green points represent the places of interest at less than 100m distance from the red point.	54
3.13	Trajectory based query example visualisation. (a) trajectories at less than 500m distance from the place of interest. (b) two example tra- jectories at less than 500m distance from the train station Versailles- Chantiers associated to trajectories exposed to a high level of NO_2 . . .	55
3.14	Execution time for each example iterated five times on three different databases sizes.	59
3.15	Two illustrations of a raw trajectory and its semantic trajectory repre- sentation in Neo4j (semantic dimensions limited to the spatial dimen- sion hierarchy).	61
3.16	Hybrid semantic trajectory path graph (operation (a)). Orange nodes represent <i>places</i> where time intervals are embedded. Coloured nodes represent other contextual dimensions.	63
3.17	Hybrid semantic trajectory path graph (operation (b)). Orange nodes represent <i>places</i> where time intervals are embedded. Blue nodes repre- sent other contextual dimensions. We limit the contextual dimension to NO_2 semantic for clarification.	64

3.18	Contextual-based semantic trajectory path graph (operation (c)). Orange nodes represent the context of interests (<i>coi</i>) where time intervals are embedded and brown nodes represent the spatial dimension (the spatial hierarchies are hidden for a better illustration).	65
3.19	Execution times of the operation examples	66
4.1	Example of semantic trajectory segments data flow with the corresponding spatial hierarchy.	73
4.2	Composite events are defined from basic events and themselves from semantic trajectories.	76
4.3	Streaming implementation architecture.	81
4.4	A visual illustration of the sliding windows manipulation on Individual (a) and Aggregated (b) events pattern stream. (a) shows how the window operation for detecting individual events are applied by grouping each moving entity's segments in one sliding window. (b) shows how the window operations for detecting the aggregated events can be applied using one sliding window grouping segments from multiple moving entities.	82
4.5	Example 3 automaton.	83
4.6	Query result for an individual trajectory event representing a trajectory going from home to office and get exposed to a high level of NO_2 for 15 minutes (<i>Query 1</i>).	85
4.7	<i>stop</i> event automaton.	86
4.8	<i>meet</i> event automaton.	87
4.9	Example of <i>meetBetweenPlacesSequence</i> event automaton.	87
4.10	Query result for an aggregated event representing trajectory commuting from home to office and meet another trajectory at a <i>place</i> for a minimum duration (<i>Query 2</i>).	88
4.11	Query result for <i>sameContextWithDifferentPlaces</i> aggregated event that detects trajectories having sports activities in two different regions at the Town hierarchical level at the same time. The results show two trajectories doing sports activities, the first at "Montigny-le-Bretonneux" town and the second at "Buc" town (<i>Query 3</i>).	90
4.12	Individual and Aggregated complex events examples throughput with different configurations and streaming rates.	92
4.13	Individual and Aggregated complex events examples latency with different configurations and streaming rates.	93

List of Tables

2.1	Semantic trajectory models: categorisation.	24
2.2	Comparison between CER systems applied on trajectories.	27
4.1	Basic events predicates and descriptions.	75
4.2	Examples of individual trajectory events.	79
4.3	Examples of aggregated events.	80
4.4	Apache flink different configuration settings.	91

Chapter 1

Introduction

The aim of this PhD thesis is to provide a modelling and event detection approach to better understand mobility in indoor and outdoor spaces. Considering the diversity of such multi-environment spaces and where mobility occurs raises several data modelling, managing and processing research challenges. This chapter introduces the research context as well as the concepts required to develop the background knowledge that will help the reader understand the remainder of this thesis. Section 1.1 presents the thesis context. We state our research challenges in Section 1.2 then describe the experimental prototyping in Section 1.3. The thesis contributions are summarised in Section 1.4. Finally, Section 1.5 presents how the dissertation is organised.

1.1 Thesis Context

Mobility data are available in many modern application areas, including intelligent transport systems, maritime surveillance and autonomous vehicles, where both the private sector and governments are searching for novel solutions to bring value to the data. Due to advances in the Internet of Things and the widespread use of Global Positional System (GPS) and embedded sensors, applications have emerged to collect geo-referenced data, leading to an explosive growth of location-based information. One class of applications producing this type of data is the Mobile Crowd Sensing (MCS) paradigm, which provides real-time data acquired by multi-sensors and mobile devices [44].

Real-Time MCS has emerged as a valuable solution to collect multiple and heterogeneous time-dependent location data cross-related with contextual data like ambient and environmental data (e.g., pollution, weather), transportation mode (e.g., bus, pedestrian) and human activities (e.g., sport, office work) [37, 59]. Understanding mobility patterns together with associated contextual information, particularly in the urban domain, requires a sound integration of the modelling level within current information infrastructures. Such notions appear as one of the pillars of the “smart city” that might impact large urban monitoring and planning tasks and how these should be designed and implemented in the city nowadays. Furthermore, urban spaces considered by the concept of smart cities include not only outdoor spaces but also indoor spaces as those are places where humans naturally spend considerable time [49].

The goal of this PhD is to provide the necessary constructs for analysing human mobility that occur jointly along time in both indoor and outdoor spaces, as well as to detect events in order to better understand their behaviours in a mobile crowd-sensed environment. The research context for the thesis includes the major thematic issues most relevant to our work: *Mobility Data Modelling and Representation* and *Real-Time Event Detection From Trajectories* (Figure 1.1).

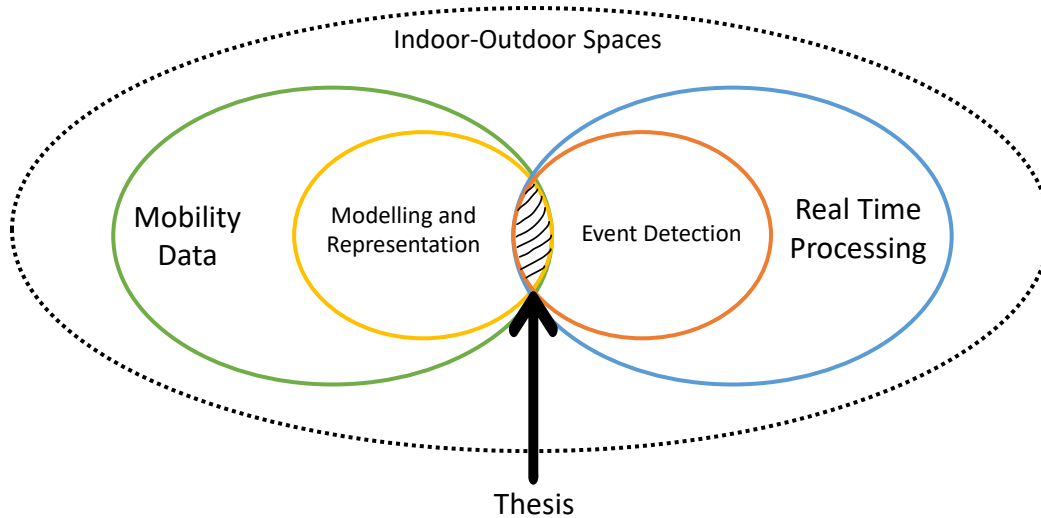


FIGURE 1.1: Thesis positioning in terms of the intersection of the main scientific thematics.

1.1.1 Mobility Data Modelling and Representation

Mobility data implies the localisation of moving objects that change over time [86]. Together with related contextual information, mobility data are noticeably available from humans navigating in both indoor and outdoor spaces. Indoor spaces denote building environments constrained by room and corridor topology [113], while outdoor spaces denote geographical space at large and are generally much less constrained [41]. While most Geographical Information System (GIS) applications have been focusing on outdoor or indoor spaces separately, this is particularly appropriate to consider a unified model for indoor and outdoor spaces where mobility occurs in order to support human trajectories understanding.

Many research studies have focused on modelling and analysing trajectories in a wide range of applications fields, including vehicle [10, 11] and vessel [20] movement, tourist service [18, 52] and air pollution [13, 15]. When studying human trajectories and associated contextual information, analysts rely on raw data generated in both indoor and outdoor spaces without excluding one of these spaces to avoid information loss.

When dealing with mobility data, a reasonable question is whether we really need all the detailed information (raw data) to analyse the data effectively [79]. Over the past few years, a series of works tried to address this question by extracting and managing the necessary semantics from the raw data to create the semantically-annotated trajectories, the so-called semantic trajectories. Semantic trajectories are an alternative representation of raw mobility data that identifies the most appropriate data abstractions, this being of benefit for many urban planning tasks using a holistic view of human mobility. Such a holistic view can be represented by a generic multi-dimensional approach for modelling mobility patterns emerging in both spaces and where trajectories are continuous.

There is a wide range of interests among users looking forward to multi-dimensional approaches for manipulating trajectories emerging in indoor and outdoor spaces. While some might be interested in a straight visualisation of these trajectories, others might either analyse trajectory patterns at the macro level or aggregated views, depending on their interests. Such a generic approach should be flexible enough to represent these trajectories at different levels of granularity and different views.

Research Objective 1

The first objective of this thesis is to provide a semantic trajectory representation of indoor and outdoor individuals' mobility where a unified model for indoor and outdoor spaces is taken into account.

1.1.2 Real-Time Event Detection From Trajectories

Real-time data processing systems became valuable solutions for providing insights on the fly when searching for mobility patterns and events of interest. In particular, detecting complex mobility patterns among crowd-sourced location data and associated contextual information should be of interest for understanding indoor and outdoor mobility and identifying some events of interest. In such cases, the analysis relies on a timely way of processing spatio-temporal data when searching for complex mobility patterns of interest. However, such patterns require multiple foundation components in order to be detected in a real-time system. First, such patterns should be modelled using a formal language that provides abstractions and constructs for identifying primitive and complex events of interest based on spatial, temporal and contextual

criteria. Accordingly, an event-driven architecture is required where a Complex Event Recognition (CER) system is required to detect interesting urban activity patterns. Such patterns occur within a data stream generated by sensors or other devices. CER refers to identifying composite events of interest, which are collections of events that satisfy some pattern [40]. CER languages allow querying for complex patterns that match incoming events based on their content, sequencing and ordering relationships, as well as other spatio-temporal constraints.

Research Objective 2

The second objective of this thesis is to provide complex event modelling, language and processing applied on the individual semantic trajectories streams involved in indoor and outdoor spaces.

While this PhD research might be applied to different application contexts and without loss of generality, it has been carried out using data campaigns produced within the Polluscope project [83], a participative observatory for the surveillance of individual exposure to air pollution relating to health. In the Polluscope project, human mobilities with multiple contextual data are collected in an environmental crowd-sensing context where real-life examples are considered as a proof of concept (e.g., individual pollution exposition in Paris, France). In such a critical phenomenon, it is relevant to study the impacts of pollution on human activities and health.

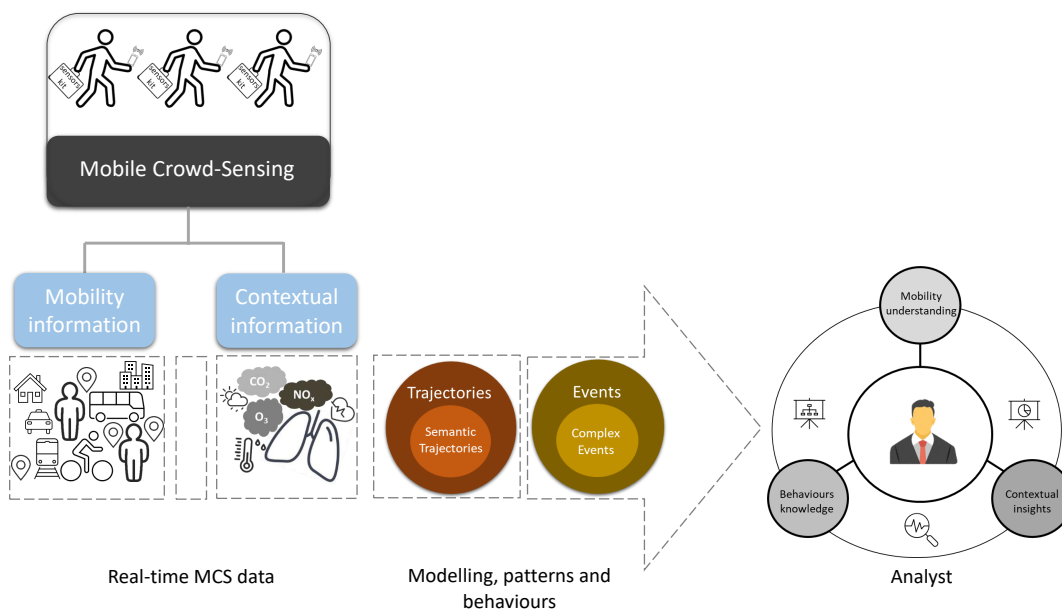


FIGURE 1.2: Research framework.

We therefore present the following thesis statement in this dissertation:

Research Statement

By modelling human mobility with related crowd-sensed contextual information in indoor and outdoor spaces, we are able to provide the appropriate constructs to analyse mobility patterns and behaviours, offline or in real-time, in order to understand urban mobility data for a specific phenomenon and facilitate the analysis of semantic mobility for a large scale of applications (Figure 1.2).

1.2 Research Challenges

The aforementioned research statement encompasses many interesting research challenges. Those addressed by this thesis are discussed below:

Research Challenge 1

How to model indoor and outdoor spaces to support a homogeneous and continuous representation of human mobility?

While most current works oriented to the modelling of human trajectories generally operate in either indoor or outdoor spaces, it clearly appears that there is still a need to consider a unified trajectory model for both indoor and outdoor spaces to infer a deep understanding of the mobility patterns and not excluding indoor spaces from the study as those are often places where humans spend a large part of their time. In addition, a typical person spends an average of 80% of his or her time indoors, at home, at work, in business premises or leisure, on transportation, etc. It will be sure of a value, and for many location-based applications, considering human activities and trajectories as a whole in indoor and outdoor spaces. Analysing trajectories should not only be performed on raw data but also on semantic trajectories abstractions that can homogenise the representation. However, getting from raw data meaningful forms of semantic trajectories in terms of mobility is not a straightforward task, in particular when considering crowd-sourced location data and associated contextual information in a given urban environment that includes indoor and outdoor spaces. To do so, it is necessary to identify the spatial environment that represents the continuous mobility from indoor to outdoor spaces and vice versa in order to ensure a complete representation of human movements and activities that will offer data manipulation and analysis capabilities [49].

Research Challenge 2

How to design a flexible and dynamic multi-dimensional semantic trajectory based on real-time crowd-sourcing data and generic enough to cope with a large range of applications capabilities?

Semantic trajectories have been mainly treated as stops and moves segments [4, 94]. A stop is considered when a given trajectory remains stable at a given location over a minimum period of time, whereas a move indicates a displacement between two consecutive stops. These trajectories need to be enriched and annotated by additional

contextual data in order to provide substantial representation capabilities at the application level [28, 76]. Over the past few years, various studies have addressed these issues but are always based on the concept of stops and moves. However, when the objective is to understand the mobility and related crowd-sensed information, stops and moves models are not fully representative for multi-dimensional trajectories. For example, during a single stop, different contextual dimensions may change of values. In such cases, a stop and moves-based model lack of representation.

Research Challenge 3

How to extract from generic semantic trajectories multiple views at different levels of abstraction according to user-defined criteria?

Users have different interests when interpreting and analysing multi-dimensional trajectories, particularly when they emerge in both indoor and outdoor spaces. Some users simply want to visualise the data, while others demand either finer details or higher-level information and aggregated knowledge. This will promote the emergence of a spatial hierarchy that supports micro to macro data representations and manipulations depending on the user's interests. A generic approach should be flexible enough to represent trajectories at different levels of granularity and views. This raises the need for an abstract and hybrid representation that considers indoor and outdoor spaces at different hierarchical levels and semantic representation. This might be of interest for many urban applications where human movements should be manipulated throughout indoor and outdoor spaces.

Research Challenge 4

How to model complex events using a flexible and expressive language to be detected in a real-time big data processing system among streams of semantic trajectories data?

The multi-dimensional semantic trajectory model represents individuals' continuous movements while integrating multiple contextual data derived from the environment in order to enrich the representation. Such a rich representation helps to reveal more mobility patterns at the processing level and favours the recognition of specific mobility behaviours as well as common patterns with some noticeable contextual situations. An important objective is to identify some complex mobility situations in real-time in order to interact with and notify a moving human on the fly. However, such complex patterns should be first modelled through a formal language that provides the necessary abstractions and constructs for identifying events of interest among indoor and outdoor multi-dimensional semantic trajectories according to some spatial, temporal and contextual criteria.

1.3 Experimental Prototyping

This section briefly describes the experimental prototyping implemented to achieve our research's objectives and evaluate our research approach.

The dataset includes mobility data, environmental data (e.g., pollution, temperature, humidity) and annotations of human activities and behaviours. The data has been collected in Paris region in France. 80+ participants were involved in multiple

data collection campaigns. They were equipped with multiple pollution, temperature and humidity sensors and tablets to manually annotate their trajectories with activities and behaviours.

Two main prototypes have been realised and experimented with a dataset collected within the Polluscope project.

- **Indoor-outdoor semantic trajectories prototype.** In order to evaluate the semantic trajectory modelling approach, we implemented in Python a data pipeline that collects Polluscope crowd-sourced data and enriches them with external spatial data. We have extracted semantic information from each dataset element and then transformed them according to the semantic trajectory conceptual model by a developed algorithm that considers a semantic trajectory segmentation based on at least one value change within a present segment. Finally, the semantic trajectories are stored in Neo4j graph database so that a list of Cypher queries can be applied. We made the semantic trajectory implementation code publicly available on Github [48]. The multiple-views extraction algorithms have been implemented in Java to extend the capabilities of Neo4j querying by multiple procedures. The code is publicly available on [47].
- **Event-driven streaming system prototype.** We have implemented a complex event processing system in Java on Apache Flink using FlinkCEP library in order to evaluate our complex events modelling approach. A list of complex events has been exemplified. A Python script for Apache Kafka streaming has been implemented to simulate different streaming loads to the Flink system in order to monitor the performance on a Prometheus/Grafana dashboard. The prototype's input is a stream of semantic trajectories segments, while the output consists of detected complex events. We made our code publicly available on [46].

1.4 Contribution

The research developed in this thesis provides several contributions to the field of mobility data modelling and processing and geographical information science. The main contributions of this dissertation are:

- We develop a hierarchical spatial data model that encompasses indoor and outdoor spaces that is able to support the spatial annotation of human trajectories in the urban domain (Chapter 3).
- We develop a semantic trajectory data model that provides a unified representation of human mobility in indoor and outdoor spaces. It integrates a hierarchical feature for its spatial dimension since it is based on our indoor and outdoor spatial data model. The semantic trajectory model also flexibly encompasses various contextual dimensions derived from crowd-sourcing data (Chapter 3).
- We implement and evaluate the semantic trajectory model in the context of the Polluscope project (Chapter 3).
- We develop a multi-view approach of semantic trajectories to extract hybrid and contextual trajectories based on spatial and contextual user-defined criteria (Chapter 3).
- We implement and evaluate the multi-view approach in the context of the Polluscope project (Chapter 3).

- We develop a Complex Event Recognition framework for detecting individual and collective complex patterns of urban mobility in real-time based on the Event Pattern Language and implement it on a modern Big Data platform (Chapter 4).
- We implement and evaluate the Complex Event Recognition system in the context of the Polluscope project (Chapter 4).

1.5 Dissertation Structure

The rest of this dissertation is organised as follows:

After this introduction, chapter 2 reviews the existing literature on indoor and outdoor spatial models, semantic trajectories models, trajectories stream processing, complex event recognition languages and trajectories complex event recognition systems.

Chapter 3 presents an indoor and outdoor spatial data model, a semantic trajectory model and a model extension for semantic trajectories multi-view. This chapter also provides experimental evaluations for the overall approach.

Chapter 4 introduces an event-based approach for composite event recognition from streams of semantic trajectories. Experimental evaluations of a developed system are also presented.

Finally, chapter 5 concludes this dissertation by summarising our contributions, limitations and discussing future work.

Chapter 2

Related Work

Due to the evolution of positioning devices and connected sensors available for indoor and outdoor spaces, real-time crowd-sourced location data and collaborative contextual information on the situation and the environment are nowadays widely available to offer many opportunities for analysing moving objects. In light of this, semantic-based representations of trajectories have been addressed over the last years at either the database or/and semantic levels [76]. This thesis aims to understand the urban mobility patterns, together with associated contextual information.

This chapter presents a summary of the related research to this thesis. This chapter is divided into four sections. Section 2.1 discusses some closely related research focusing on representing indoor and outdoor spaces in a unified model. Section 2.2 provides an overview of the literature on trajectory modelling approaches and identifies the limitations of state-of-the-art conceptual models before proposing how to overcome them in the case of continuous indoor-outdoor movement in chapter 3. In Section 2.3, we discuss the well known spatio-temporal database research work. We also address in Section 2.4 some approaches that considered trajectories stream processing to detect some patterns. In Section 2.5, we overview the well-known complex event processing languages in the literature. Then we address in Section 2.6 the complex event recognition work on trajectories. Finally, Section 2.7 summarises the chapter with a critical analysis.

2.1 Indoor and Outdoor Spatial Models For Navigation, Route Planning and Path Search

The space is an essential notion that aims to indicate the physical or imaginary parts of the present environment [124]. Humans refer to portions of space, such as places or locales, rather than universal, unlimited space in their daily lives.

Geographical Information System (GIS) applications have gradually increased from large outdoor environments to small scale indoor environments where indoor and outdoor spaces have been relatively well addressed but separately to a certain degree. Indoor spaces generally denote the space enclosed by a building, while outdoor spaces represent external open spaces [112, 122]. Several data modelling approaches have been searching for a unified representation of trajectories in indoor and outdoor spaces. Challenges behind the design of a unified model for indoor and outdoor spaces have been first discussed by Giudice et al. [41] and Yang et al. [119], where the main similarities and differences between the two are described.

In this section, we present related work that addressed unified indoor and outdoor spatial models and that most of them are often oriented towards navigation, route planning and path search.

2.1.1 Navigation and Route Planning

In a related work, the need for ontological and formal foundations for a navigation model in a unified outdoor and indoor space has been formulated, but so far, no computational implementation has been realised to support the claim [118]. Outdoor GIS models might also provide support for emergency services acting in a 3D indoor environment, as suggested by Tashakkori et al. [99] and Teo et al. [100], but these approaches are mostly oriented towards building information management and do not fully support trajectory representations. In recent works [25, 107, 116], a series of spatial data models have been developed for navigation and route planning in both indoor and outdoor spaces, but these approaches do not provide data representation and manipulation capabilities that can easily support seamless integration of humans trajectories.

Multiple modelling frameworks have been suggested for unifying indoor and outdoor spaces for pedestrian navigation [108]. Zhang et al. [121] introduced a unified graph model where indoor and outdoor spaces are represented with sub-graph structures and presented an automated construction algorithm to extract unified navigation graph from different data formats. Similarly, Claridades and Lee [22] presented a framework to integrate indoor and outdoor network datasets through spatial relationships to enable the connection between indoor and outdoor navigation networks for seamless navigation. However, these approaches do not support trajectory data modelling and queries.

Indoor and outdoor environments have also long attracted interest from robotic applications, but these applications are mostly oriented to individual machine-driven navigation based on sensors and reasoning mechanisms [16, 90].

2.1.2 Path Search

In [50], indoor and outdoor spaces are seamlessly considered for applying shortest path queries. These authors proposed a graph-based model that represents indoor

and outdoor spaces in a single structure and provided shortest path algorithms. However, this model does not integrate a hierarchical component, nor does it provide spatial query capabilities. Indoor and outdoor frameworks have been developed for path discovery where, for instance, an algorithmic approach is proposed to integrate indoor and outdoor spaces to provide a path with minimum outdoor exposure and shortest distance [23]. This work has been extended to provide a context-aware path recommendation system with minimum average congestion and shortest distance [24]. However, and as in previous works, these works do not integrate hierarchical spatial support for trajectory modelling and querying.

Overall, it appears that most, if not all approaches, are oriented towards navigation, routing and path searches. While most of them seamlessly consider indoor and outdoor spaces, there is still a need for a better modelling integration of these two spaces, as well as a better consideration of the multi-level hierarchical properties that arise in many application domains as well as data manipulation capabilities at the query level. This is a required and most expected property for a homogeneous representation of trajectories moving from indoor to outdoor and vice versa.

The main differences between our proposed spatial model and previous ones are that our model provides a unified indoor and outdoor spatial model oriented to the representation of semantic trajectories. It also integrates hierarchical and graph components that support a flexible representation at different levels of abstraction that favours the integration of heterogeneous crowd-sourcing trajectory data, and further data manipulation capabilities at the interface level using graph queries and graph analytics.

2.2 Trajectory Data Models

Every physical object that has its own position that may change over time is said to be a mobile object. Mobility data refers to information on the movement of objects, which comprises, at least, information about the localisation and time. Mobility data could be data coming from a variety of data sensors [96]: GPS devices (e.g., during outdoor displacements), Radio-Frequency Identification (RFID) systems, Wi-Fi access points or Bluetooth devices (e.g., during indoor displacements).

The raw data produced by mobile sensors can be considered as an implicit representation of different trajectories. The raw trajectories are adapted to monitor the successive locations of a moving object, compute statistics, derive patterns and outliers, and thus allow understanding of the mobility phenomenon. However, in order to extract valuable information from trajectories and improve the understanding of a specific domain related to trajectories, raw trajectories could be insufficient. Different properties related to such domain should be considered.

During the past few years, a series of works have introduced several conceptual trajectory data models specific to indoor spaces, while outdoor spaces have been largely and earlier addressed by many spatial database models oriented for instance to the representation of mobility and traffic flows at either urban or regional scales. The objective of the conceptual data models is to provide a high-level data structure that establishes the link between user requirements and the system's design [39, 54].

The following discusses and outlines a series of related works oriented towards either indoor or outdoor semantic trajectory models.

2.2.1 Semantic Trajectories

Several works have introduced semantic-based representations oriented towards the representation of human trajectories (Table 2.1). Parent et al. [76] gives a survey and highlights some of the main principles behind the representation of semantic trajectory modelling and some of the main abstractions used so far (e.g., stop and move abstractions). The first formally defined semantic trajectory at the conceptual level was proposed by Alvares et al. [4] and Spaccapietra et al. [94]. Alvares et al. [4] considered stop and move concepts that qualify semantic trajectories and provides further analysis capabilities close to users' perspectives, and satisfactory processing performance from a computational point of view. They simply identified that a trajectory consist of data in the form of (tid, x, y, t) where tid is the moving object identified, x and y are the spatial coordinates and t is a timestamp. They developed a series of algorithms to extract stops and moves moving patterns. Spaccapietra et al. [94] still considered similar principles with a trajectory modelled as a semantic sequence of stops and moves. The authors defined a trajectory by $trajectory : [t_{begin}, t_{end}] \rightarrow space$ as the user defined record of the evolution of the position (perceived as a point) of an object that is moving in space during a given time interval in order to achieve a given goal.

2.2.2 Annotated Semantic Trajectories

Yan et al. [117] introduced a labelling framework for stop- and move-based trajectories called SeMiTri (Semantic Middleware for Trajectories). SeMiTri's objective is to support semantic trajectories enrichment by exploiting both the geometric properties and background geographic and application data. Their framework annotates trajectories to construct semantic trajectories as derived from GPS raw data. A few attempts have been recently made to enhance semantic trajectory with multi-dimensional information. A semantic conceptual model has been developed by Bogorny et al. [12] to take into account a limited predefined set of semantic properties for trajectory data analysis. Fileto et al. [36] introduced a framework to fill the gap between movement data and formal semantics. The authors provided an ontological model for semantic enrichment and introduced a multilevel hierarchy for representing trajectory in the space. Ruback et al. [88] developed an alternative framework with linked open data based on dynamical enrichment of trajectories using ontology mashups. Ilarri et al. [49] and Ferrero et al. [35] highlighted the need for considering multi-dimensional representation of semantic trajectories. Accordingly, it should be possible to include all aspects information into a unique trajectory representation then allow to commute from one semantic to another according to some query requirements. More recently, Mello et al. [62] extended previous works [12, 36] by introducing the concept of multiple aspect trajectories called master. Master supports trajectory sequences enrichment with different and heterogeneous annotations. At the logical level, Valdés et al. [106] modelled semantic trajectories as symbolic trajectories within the database system SECONDO [45]. Semantic trajectories are segmented by points of interest and by transportation modes, and annotated with labels indexed by time intervals in the form of a set-based model. However, this set-based model does not fully consider hierarchical relationships and associated spatial semantics. Jin and Claramunt [51] introduced a database management approach for representing and analysing human trajectories in an urban environment, but different levels of abstractions are not considered, nor indoor spaces are taken into account. Li et al. [57] proposed a Markov model to enrich indoor mobility data with a semantic region, a time period, and a mobility event. More recently, Vidal-Filho et al. [109] introduced a definition for

documenting and enriching semantic trajectories using Spatial Data Infrastructures. This approach enables automatic and manual annotation of trajectories using meta-data standards. However, this approach does not consider both indoor and outdoor spaces.

2.2.3 Multi-granular Semantic Trajectories

Ilarri et al. [49] raised the need for representing semantic trajectories at different levels of abstraction. Similarly, Pelekis et al. [80] developed a stop and move trajectory model to support different scales and spatio-temporal granularities. Kontarinis et al. [52] designed a specific semantic trajectory model applied to a museum case study that considers an indoor hierarchical space model based on the IndoorGML standard. However, these approaches are oriented towards either indoor or outdoor trajectories. Gómez et al. [42] introduced a semantic trajectory model on top of a graph database to apply different aggregation mechanisms using graph-based queries, their model considered a hierarchical structure to describe different spatial layers. Their study showed that trajectory queries can be expressed more naturally on top of graph-based representations rather than common relational databases.

2.2.4 Indoor and Outdoor Semantic Trajectories

Xu and Güting [115] designed a generic data model to manage moving objects in indoor and outdoor environments. They conceptually model the space extent in which geographical objects define available places for moving objects. However, semantic properties were not fully considered to provide a complete trajectory data representation. While several data manipulation frameworks have been developed regarding the semantic, temporal, and spatial dimensions, however, they do not consider indoor and outdoor spaces together, and trajectories were predefined and extracted from a given urban data set without further manipulation capabilities.

Overall, it appears that most previous works searched for an integrated indoor and outdoor representation, while others take into account a multi-level and flexible semantic representation in which human trajectories can be derived from crowd-sourcing data. Most existing formalisms are supported by raw data sequences that represent trajectory data annotated by contextual information (e.g., weather conditions, pollution exposition along a path).

Semantic trajectory models	Stops and Moves	Indoor	Outdoor	Indoor and Outdoor	Hierarchical space
Alvares et al. [4], Spaccapetra et al. [94], Yan et al. [117], Bogorny et al. [12], Pelekis et al. [80], Ruback et al. [88], Jin and Claramunt [51], Vidal-Filho et al. [109]	✓	✗	✓	✗	✗
Valdés et al. [106], Mello et al. [62]	✗	✗	✓	✗	✗
Fileto et al. [36], Gómez et al. [42]	✓	✗	✓	✗	✓
Kontarinis et al. [52]	✓	✓	✗	✗	✓
Xu and Güting [115]	✓	✓	✓	✓	✗
Noureddine et al. [70], Noureddine et al. [67]	✗	✓	✓	✓	✓

TABLE 2.1: Semantic trajectory models: categorisation.

2.3 Trajectory Database Engines

The research on spatio-temporal databases has separately studied the spatial and the temporal aspects. Research on spatial databases has focused on supporting the modelling and querying of geometries associated with objects in a database. Most commercial and open-source database management systems provide appropriate data management and querying mechanisms for static spatial data conforming to Open Geospatial Consortium (OGC) standards [73]. For example, Oracle [75] and PostgreSQL [84] have been extended with spatial modules, Oracle Spatial [53] and PostGIS [97], respectively. On the other hand, temporal databases have focused on extending the database systems to handle time-series data more efficiently. For instance, TimescaleDB [102] is an open-source time-series extension for PostgreSQL.

The general idea of spatio-temporal databases is to carry out the management of spatio-temporal data by combining spatial and temporal data management into a single framework. A few attempts have been made to achieve an appropriate interaction between temporal and spatial data. Güting et al. [45] developed the first moving objects database engine that appeared in the literature called SECONDO. SECONDO is adjusted to be extended by algebra modules for non-standard applications. Another attempt developed by Pelekis et al. [81] is Hermes. Hermes is a system extension aiming to support continuously moving objects' modelling and querying. Hermes defines a trajectory data type and a collection of operations in order to provide trajectory functionality to Oracle and PostgreSQL. Recently, Zimányi et al. [123] developed MobilityDB that extends the type system of PostgreSQL and PostGIS with time-varying data types for representing moving object data and conforming to the OGC Moving Features Access specification [72]. MobilityDB builds on existing operations and effectively enhances them into temporal operations.

During the last few years, the community has shown interest in modelling and querying trajectories in graph databases, particularly for managing semantically enriched trajectories [32, 55, 98]. Graph databases facilitate the exploration of complex structures and dynamic relationships for connected trajectory data since they are based on graph models that have a typical way of representing networks and connections between entities. Graph-based models are well known for materialising complex relationships, including hierarchical spatial data, temporal data and thematic enrichment data. Gómez et al. [42] discussed how graph databases (Neo4j [64] in their case)

are used to manage semantic trajectory data. Their study has shown that trajectory queries are more naturally expressed as graphs using a graph query language. The study also showed that the graph database queries are 1.2 to 7 times faster than the relational ones in their case.

2.4 Trajectory Streams Processing

A recent body of work addressed trajectory-based and event queries on data stream management systems. The main objective is to identify basic and composite events that embed some application-defined situations. Vieira et al. [110] introduced an algorithm that can be applied in a streaming environment to discover flock patterns among trajectories online. The algorithm assumes that moving objects are close together if a disk with a given radius covers all moving objects in the pattern. Roganovic et al. [87] developed an indoor-based system for patterns detection among symbolic trajectory data streams. Masciari et al. [60] introduced a sequence pattern mining approach for trajectory streams. It is based on a partitioning strategy to reduce the trajectory size and present the trajectories as strings. They applied a sliding windows combined with a counting algorithm. Boutsis et al. [14] proposed an online clustering technique over positioning data streams to identify real-life events when they occur. Salmon and Ray [89] developed a spatio-temporal stream-based architecture that deals with archived and streamed maritime trajectories data. Patroumpas et al. [78] presented a framework to identify mobility events online among aircraft's trajectories streams, where only important events can be retained. Their prototype has been implemented using the DataStream API of Apache Flink. Chen et al. [19] and Fang et al. [34] also investigated a system to support real-time co-movement pattern detection over streaming trajectories where clustering and pattern enumeration are considered, respectively. Tritsarolis et al. [104] introduced an online graph-based mining algorithm to discover co-movement behaviour by monitoring the activity of multiple clusters through time and space.

While several trajectories stream processing frameworks have been successfully applied to detect some events and patterns of interest, however, they do not consider formal languages to fully represent and manipulate them. Therefore, their querying capabilities are very limited and not expressive enough. Composite events are also not taken into account then, limiting the range of possible analysis.

2.5 Complex Event Recognition Languages

During the last decade, a variety of complex event recognition models and languages have been proposed. They can be classified into three types: tree-based structures, logic-based rules and automata-based. Giatrakos et al. [40] gives a survey and highlights the main features behind the formal methods applied to these complex event models.

2.5.1 Tree-based Models

Tree-based models define a tree of event operators to specify complex event patterns. For example, ZStream [61] employs a tree-based definition of complex patterns where operators (sequencing, negation, conjunction, disjunction and Kleen closure) connect primitive or complex events to form new complex events.

2.5.2 Logic-based Models

Logic-based models employ temporal formalism that relies on logic rules and inference to detect complex events. Artikis et al. [8] gives a survey on logic-based event models. One kind of logic-based model is the Chronicle Recognition System (CRS) as [30] and that defines events using logic predicates linked together by time constraints and whose occurrence may depend on the context. The second kind of logic-based event model is the Event Calculus as in RTEC [7] that supports event-based reasoning while incorporating a sliding windowing mechanism.

2.5.3 Automata-based Models

Automata-based models provide pattern automata languages. For example, SASE [114], SASE+ [29], SASE++ [120], Siddhi [93] and Wayeb [2] translate patterns into non deterministic automatas. They have valuable closure properties and expressiveness. Recently, automata-based models have progressed to a good level of maturity, such as FlinkCEP [5] library that made its way into the Apache Flink [6] [17] system.

2.5.4 Hybrid Models

Other event-based languages consider hybrid approaches like Esper [33] and uses trees for the core of its functionality (e.g., filtering, windowing and aggregations) and non-deterministic automatas for their pattern matching functionality. It defines a rich declarative language for rule specification called event processing language and includes all the operators of SQL adding ad-hoc constructs for windows definition.

Overall, it appears that the CER models have common goals, but they differ in their architecture, pattern language and processing mechanisms. The Event Pattern Language (EPL) described by Gehani et al. [38] and Motakis and Zaniolo [63] provide relatively complete logical constructs for composite event specifications; however, to the best of our knowledge, the EPL has not been implemented yet in a CER context.

2.6 Trajectories Complex Event Recognition

Big data platforms have simplified the design and implementation of distributed processing pipelines, such as Storm [103] and Spark [95]. However, these platforms cannot enable CER on their own. Flink [17] and, because of its native FlinkCEP library, can incorporate CER operators in the Flink program.

A few attempts have addressed composite event detection on spatio-temporal trajectories (Table 2.2). Schwiderski-Grosche et al. [91] suggested a composite event language for spatio-temporal reasoning over mobility data with pre-defined semantics. A series of spatial operators associate events to pre-defined areas and whether different events occur at given times and/or locations. Terroso-Saenz et al. [101] introduced CEP-traj, an event-based system that integrates trajectory streams and applies a series of steps covering data cleaning, segmentation and compression to ease pattern detection. CEP-traj has been developed using Esper to detect some behavioural patterns of interest. Composite Event Recognition (CER) systems based on RTEC

CER systems	Event model	Input	Contextual data	Indoor and outdoor	Hierarchical space	Stream processing	Basic and composite events	Scalable
Schwiderski-Grosche et al. [91]	<i>not implemented</i>	n/a	n/a	n/a	n/a	n/a	✓	n/a
Terroso-Saenz et al. [101]	Esper	raw trajectories	✗	✗	✗	✓	✗	✗
Artikis et al. [9], Patroumpas et al. [77], Pitsikalis et al. [82]	RTEC	raw trajectories	✗	✗	✗	✓	✓	✓
Tsilionis et al. [105]	RTEC	raw trajectories	✓	✗	✗	✓	✓	✓
Ntoulias et al. [71]	Wayeb	raw trajectories	✗	✗	✗	✓	✓	✓
Our approach system	EPL	semantic trajectories	✓	✓	✓	✓	✓	✓

TABLE 2.2: Comparison between CER systems applied on trajectories.

have been used in various application domains such as maritime monitoring [77] [82] and fleet management [105] [9] for maritime and urban surveillance, respectively. Artikis et al. [9] presented a CER system based on RTEC and that detects complex events from urban traffic data in order to identify incidents in an urban network. Still based on RTEC, Patroumpas et al. [77] developed a maritime surveillance system for trajectory simplification and that consumes positional streams of ship messages and detects complex events along its trajectory. This system reduces the trajectory by a few detected critical points. Pitsikalis et al. [82] improved the work in [77] by constructing effective patterns of maritime activities, implemented using the RTEC language. Also based on RTEC, Tsilionis et al. [105] proposed an online CEP system that employs weather and points of interest data enrichment to detect events over moving vehicles.

More recently, Ntoulias et al. [71] presented a scalable CEP system based on Wayeb for processing big mobility data streams, experimented for fleet management and maritime monitoring applications

In conclusion, CER should support the composition of primary events towards composite events and should be flexible enough to support different composition mechanisms to fulfil the necessity of a language providing logical constructs for modelling complex events using a wide range of operators.

2.7 Synthesis and Positioning

In this chapter, we have reviewed previous research that has inspired much of the work presented within this dissertation. We covered two main areas. The first area includes spatial modelling, trajectory data modelling and trajectory databases. The second area includes trajectory stream processing and complex event recognition. We present below the critical analysis in each of the covered areas while comparing related work to our work:

Our work in chapter 3 introduces the main principles of a semantic trajectory model in indoor and outdoor spaces. Based on these modelling principles, we first

provide formal support for the semantic-based modelling approach. Second, flexible logical data manipulation is defined at different levels of abstraction and implemented, making a difference between graph queries and graph analytics. And finally, a computational implementation on top of the graph database Neo4j. In contrast to related work, our semantic trajectory formalism is flexible enough to derive different trajectories according to some given contextual properties associated with raw trajectory data as they emerge in space and time. Our approach encompasses the multiple hierarchical granularities of urban spaces and supports a finer representation of movement semantics in both indoor and outdoor environments while maintaining multiple dimensions of semantic associated data. This semantic trajectory model is not completely based on stop and move abstractions, but a peculiarity is that trajectories can be dynamically redefined using some semantic, spatial and temporal criteria.

This PhD research takes advantage of EPL's expressiveness and flexibility to model complex events in the urban mobility domain. The purpose of this PhD research is not to demonstrate that EPL is inherently superior to the previously mentioned CER models but to show that EPL is a high-level language that has a good level of expressiveness with a relatively complete set of composition operators to represent complex events. We apply EPL to a spatio-temporal context and implement it in an architecture that allows considering hierarchical and recursive characteristics for the definition of complex events using pre-defined basic events. EPL language allows the construction of composite events using a wide variety of user-defined basic events, and are not limited to, spatial, contextual and temporal criteria. The spatial criteria embed indoor and outdoor specifications in addition to the hierarchical aspect defined in [67]. The contextual criteria are flexible enough to embed multiple contextual information that arises with urban human mobility. The temporal criteria cover Allen's Interval Algebra [3], duration and overlapping specifications.

Most of the previous CER systems consider raw trajectories as internal data repositories either without or with very limited external information, such as on weather and points of interest, while neglecting the importance of identifying the semantics of the space where the trajectories occur or the contextual information that might be associated along a trajectory. Such semantic information could be very interesting to ease the definition of complex event semantics and detection.

The main difference between our proposed system in chapter 4 and the previous ones is that our system considers semantic trajectories in a well-defined structure with formal support presented in [67] [70] that evolved in indoor and outdoor spaces as input while maintaining multiple and not limited contextual information that arises with urban mobility. Based on such trajectory modelling, our approach supports a flexible composition of events with a flexible definition of a wide range of basic events depending on spatial, temporal and contextual criteria to construct trajectories' complex events.

Our work in chapter 4 introduces a modelling approach that extracts basic events from semantic trajectory data stream while considering spatial, temporal and contextual criteria over different levels of hierarchy. A formal representation of events and composite events is extended from the EPL language. It is implemented and evaluated using a data stream processing platform Apache Flink [6] using the built-in support for CER via FlinkCEP [5].

Chapter 3

Semantic Trajectory Data Model For Indoor and Outdoor Spaces

With other related crowd-sensed information, location data are noticeably available for humans navigating in both indoor and outdoor spaces. Considering the diversity of such multi-environment spaces, and where mobility occurs, offers many opportunities for analysing humans' trajectories. Indoor and outdoor mobility analysis raises several data modelling, managing and processing research challenges. This not only implies developing appropriate database architecture for large streamed data but also identifying the most appropriate data abstractions to model these human trajectories on the semantic level.

Modelling indoor and outdoor mobility requires a holistic view of the spatial, temporal and semantic dimensions. This requires a generic and multi-dimensional representation that encompasses spatial, temporal and contextual criteria.

This chapter presents (i) an indoor and outdoor spatial data model that supports a homogeneous representation of semantic trajectories, (ii) a multi-layered graph representation for semantic trajectories in indoor and outdoor spaces and (iii) multiple-views representation for semantic trajectories. (iv) The overall model is implemented, experimented and evaluated.

The content of this chapter is mainly based on our previous publications in the International Conference on Mobile Data Management 2020 (MDM) [66, 70], the Transactions in GIS Journal [67], the International Conference on Advances in Geographic Information Systems 2021 (SIGSPATIAL) [69] and the International Symposium on Web and Wireless Geographical Information Systems 2022 (W2GIS) [68].

3.1 Introduction

Exploring crowd-sensed location data and associated contextual information at a large scale in a given urban environment should help to reveal mobility patterns and outliers, this being of benefit for many urban planning tasks. Not only these patterns should be modelled and studied at different scales, but also using a holistic view in which human mobility should be considered in both outdoor and indoor spaces. This is a necessary assumption made in this chapter as a large part of human mobility arises in large indoor spaces such as commercial or administrative buildings, not to mention the fact that humans spend most of their life in indoor spaces.

An indoor space denotes a building environment roughly modelled as a set of rooms and corridors, and where connections between these components are considered as topological relations that form a structural graph representation. While outdoor spaces denote geographical spaces at large, they are generally much less constrained than indoor spaces as indeed a building generally has a well-designed set of rooms and corridors that denote some functional relations.

In order to represent human movements emerging in indoor and outdoor spaces, a unified spatial model is required to homogeneously identify the indoor and outdoor places where they occur. Human mobility can be roughly described as a spatio-temporal and semantic trajectory through a range of indoor and/or outdoor spaces, each of which has its own characteristics [49]. In the urban domain, indoor spaces do not exist as isolated frames of references, they constitute an implicit part of the entire urban geographical environment as outdoor spaces do. Humans continuously move between their living and external places (e.g., work, leisure), experiencing smooth movement from indoor spaces to boundaries to outdoor spaces and back. This lets us assume a seamless link between those two fully integrated spaces.

While human movements in outdoor spaces can be easily tracked using GPS-based devices, there is a need in indoor spaces for some specific location-aware devices such as WiFi or RFID sensors [96]. Data integration between trajectory data from multiple devices, spatial data and other data can build more comprehensive city profiles [111]. When considering indoor and outdoor spaces within a sort of integrated framework, the diversity and heterogeneity of scales and location-based devices raise several data modelling, managing and processing research challenges.

The overall objective of this chapter is divided into two major goals:

Goal 1: The first goal of this chapter aims to develop a generic framework for location-based applications and services that consider human behaviours and trajectories in indoor and outdoor as a whole and at a finer level of granularity in time as well as scale in space. In such applications, data scientists and decision-makers rely mainly on precise and massive trajectory data to understand some emerging phenomena in space and time, and should not exclude indoor spaces from the analysis to build more comprehensive user-profiles and then valuable patterns. This goal embeds three research challenges:

Research Challenge 1

How to model indoor and outdoor spaces to support a homogeneous and continuous representation of human mobility?

Identifying the spatial environment where mobility occurs is an essential task in order to ensure a complete representation of human movements in indoor and outdoor. To

address this challenge, we introduce a hierarchical spatial model that encompasses indoor and outdoor spaces. This model supports human mobility and offers further data manipulation and analysis capabilities (cf. Section 3.2).

Research Challenge 2

How to design a flexible and dynamic multi-dimensional semantic trajectory based on real-time crowd-sourcing data and generic enough to cope with a large range of applications capabilities?

Representing human trajectories using multiple annotations can provide substantial representation capabilities at the application level. However, stop and move models cannot fully represent a multi-contextual representation of trajectories when the objective is to understand the movement and related crowd-sensed information, which are not completely predefined. A sound graph-based representation should exhibit the most appropriate stop instances and movements, but according to incoming crowd-sourcing data, this being not completely known a priori. To address this challenge, we develop a flexible semantic trajectory model that integrates multiple data sources in indoor and outdoor spaces and where trajectories are dynamically generated (cf. Section 3.3).

Research Challenge 3

How to extract from generic semantic trajectories multiple views at different levels of abstraction according to user-defined criteria?

There is a wide range of interests among users looking forward to multi-dimensional approaches for manipulating trajectories emerging in indoor and outdoor spaces. While some might be interested in a straight visualisation of these trajectories, others might either analyse trajectory patterns at the macro-level or aggregated views, depending on their interests. Such a generic approach should be flexible enough to represent these trajectories at different levels of granularity and different views. This raises the need for an abstract and hybrid representation that considers indoor and outdoor spaces at different levels of granularity and with a hierarchical and semantic representation. To address this challenge, we develop a formal and logical extension of the semantic trajectory model to extract semantic trajectories at different levels of abstraction according to different spatial and contextual user interests (cf. Section 3.4).

Goal 2: The second goal of this chapter is to define a methodology to implement and experiment the formal models designed at the conceptual level with real data. It also aims to validate the model. This goal encompasses four tasks are as follows:

- Study the feasibility and the performance of the semantic trajectory modelling approach on a graph-based database system using a large and real urban trajectory dataset (cf. Section 3.5).
- Explore to which degree the application framework can reveal some movement patterns in the urban domain while considering the spatial, temporal and thematic criteria (cf. Section 3.5).
- Extend the graph-based database system in order to extract multiple views of semantic trajectories according to users' interests (cf. Section 3.5.4).

- Investigate the feasibility and performance of multi-view extraction mechanisms using large and real data (cf. Section 3.5.4).

The model is applied to real data collected from the Polluscope project [83], where human trajectories with multiple contextual data are collected in an environmental crowd-sensing context. Such trajectories are characterised by external data and geo-located using OpenStreetMap [74]. Semantic trajectories are stored in a property graph database Neo4j [64] and manipulated using a high-level query language and algorithms [42]. A computational implementation on the graph database supported by a series of graph queries and graph analytics examples with performance and robustness evaluations are presented. Then, an extension implementation of the graph database is presented to support the multiple views extraction exemplified by a series of graph-based processing operations and performance evaluations that show its flexibility when implemented on top of a real data context.

3.2 Indoor and Outdoor Space Model

The modelling approach described in this section provides the foundations on which the spatial representation of our semantic trajectory model is based on. The principles and definitions of the indoor and outdoor spatial model are hereafter introduced.

3.2.1 Spatial Modelling Principles

We introduce a spatial model that brings together indoor and outdoor spaces into a unified and multi-granular spatial representation. This model should be multi-granular since it is defined at a lower abstraction level from which coarser levels can be derived appropriately when necessary, and covering both indoor and outdoor spaces. Choosing an appropriate level of abstraction will depend on the application context and user preferences, but by considering the lowest level of abstraction, the objective is to favour different kinds of applications and services. Implicitly, this multi-layered model encompasses various levels from micro to macro, from continuous to discrete spatial structures at the representation level in order to optimise the data manipulation and storage levels.

This representation also takes into account the broad trends of positioning techniques. Since raw mobility data can be received from different types of sensors [96], different data sensors might exist for representing a given trajectory (RFID, Wi-Fi, Bluetooth, etc. for indoor mobility and GPS for outdoor mobility). This leads us to normalise the representation of mobility across different locations and spatial entities in indoor and outdoor. Hence, a homogeneous representation of indoor and outdoor spaces is considered in order to support the smooth mobility of humans from indoor to outdoor spaces and vice versa.

This modelling approach can be materialised as a place-based representation where each place represents a space, either indoor or outdoor, characterised by a list of properties including location data, and where spatial hierarchies are considered between entities in order to provide a sound representation that also considers connections between such entities.

We define a graph-based representation that supports structural properties (i.e., node-to-node relationships) at different levels of granularity while maintaining geometric properties. This representation encompasses the mobility of humans navigating in continuity in these indoor and outdoor places/spaces.

A peculiarity of the hierarchical modelling approach is that it allows the manipulation of human trajectories at different levels of abstraction. Additional data manipulation capabilities can also be applied at the query level using usual semantic, spatial and temporal operators, as this will be further illustrated in the experiments.

3.2.2 Formal Definitions

Let us introduce a series of formal definitions that will successively introduce the notion of the spatial layer that denote a set of entities at a given level of abstraction. The spatial layer acts as a given representation of either an indoor or outdoor space for a given level of abstraction. Therefore, a hierarchy embeds the different layers of both the indoor or outdoor space. Specific constraints are defined to ensure the continuity between indoor and outdoor spaces. Finally, a model of graph is introduced to materialise continuity properties and manipulation capabilities between indoor and outdoor spaces.

Definition 1 (Spatial Layer). A spatial layer denoted by \mathbb{L} is defined as a set of geographical features (i.e., points, polylines, polygons) representing a class of real-world entities at a given level of abstraction.

Definition 2 (Indoor Space). Let us consider an indoor space denoted by $\mathbb{I} = \bigcup_{i=1 \dots |LI|} LI_i$ made up of a set of hierarchically ordering of nested layers which represent the successive levels of abstraction of the indoor space. This is accordingly denoted by $LI_{|LI|} \subset LI_{|LI|-1} \subset \dots \subset LI_1$ as indoor layers. The coarser level of abstraction is LI_1 while the finest level of abstraction is $LI_{|LI|}$. Furthermore, $\forall s'_{i-1} \in LI_{i-1}, \exists s'_i \in LI_i | s'_i \subset s'_{i-1}$.

Definition 3 (Outdoor Space). Let us consider an outdoor space denoted by $\mathbb{O} = \bigcup_{i=1 \dots |LO|} LO_i$ made up of a set of hierarchically ordering of nested layers which represent the successive levels of abstraction of the outdoor space. This is accordingly denoted by $LO_{|LO|} \subset LO_{|LO|-1} \subset \dots \subset LO_1$ as outdoor layers. The coarser level of abstraction is LO_1 while the finest level of abstraction is $LO_{|LO|}$. Furthermore, $\forall s''_{i-1} \in LO_{i-1}, \exists s''_i \in LO_i | s''_i \subset s''_{i-1}$.

Definition 4 (Unified Space). Let us consider a unified spatial representation that encompasses indoor and outdoor spaces, expressed as:

$$\mathbb{S} = \bigcup_{i=1}^k \mathbb{L}_i \quad (3.1)$$

where \mathbb{S} denotes the entire space encapsulating indoor and outdoor spaces layers.

Definition 5 (Indoor-Outdoor Space Continuity). In order to achieve a continuity between indoor and outdoor spaces, there should exist at least one common layer between these two spaces LC such as $LC \in I$ and $LC \in O$ while the other layers should be disjoint. In fact indoor and outdoor spaces should not be completely disjoint as a continuity should cover the elements that belong to both indoor and outdoor spaces. As outdoor and indoor spaces are defined at complementary levels of abstraction and scale, let LC denote the finest layer of outdoor space and the coarser layer of indoor space. Practically there exists at least an instance $i_i \in LI_i$ and an instance $o_i \in LO_i$ that materialises a same spatial entity

Definition 6 (Indoor and Outdoor Graph). A graph-based representation is defined for \mathbb{S} as a layered multi-graph $G = (V, E)$ where

$$V = \bigcup_{i=1}^k S_i = \bigcup_{i=1}^n I_i \cup \bigcup_{i=1}^m O_i \tag{3.2}$$

and

$$E = \bigcup_{i=1}^n E_i \tag{3.3}$$

The graph G is made of different layers of nodes and edges. On the one hand, the edges E_i denote the embedded hierarchy between the different indoor and outdoor layers. On the other hand it also spatially connects the indoor and outdoor spaces using the layer that belongs to both spaces (i.e., the layer building) as denoted in Figure 3.1.

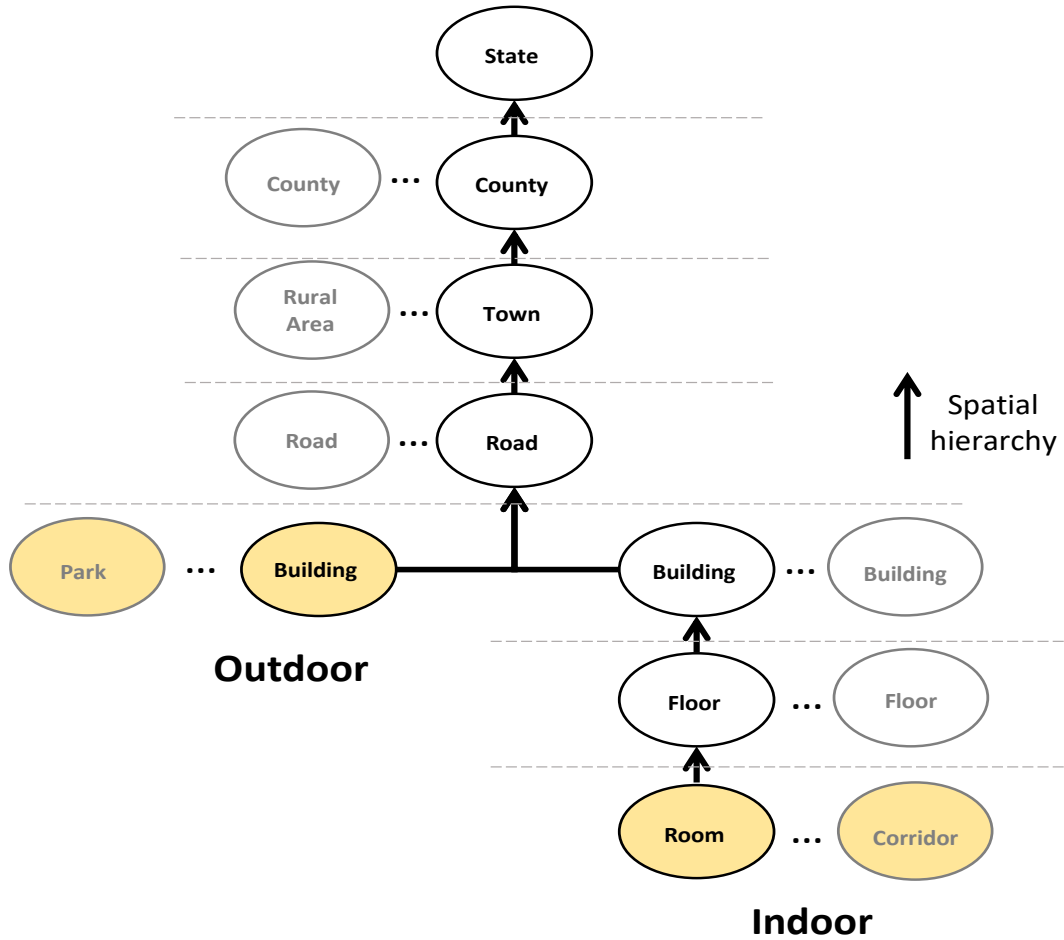


FIGURE 3.1: Spatial model showing the hierarchical relationships between indoor and outdoor spatial layers.

Each node $v_i \in V$ is geometrically embedded into a spatial hierarchy/layer that denotes the spatial structure of the underlying space (e.g., room to floor layers in an indoor space, road to town layers in an outdoor space).

Let $O_{label} = \{O_{State} \cup O_{County} \cup O_{Town} \cup \dots\} \setminus \{O_{Road}\}$ denote a set of labels or symbolic values that represents the nested hierarchy from v_i to v_j such as $v_i, v_j \in O$. And $I_{label} = \{I_{Building} \cup I_{Floor} \cup I_{Room} \cup \dots\}$ is a set of labels or symbolic values that identify the nested hierarchy from v_i to v_j such as $v_i, v_j \in I$. $\{O_{Road}\}$ is a set of labels or symbolic values that identify the nested hierarchy from v_i to v_j such as $v_i \in V$ and $v_j \in O$. Then $S_{label} = O_{label} \cup \{O_{Road}\} \cup I_{label}$ a set of labels.

An edge $e \in E \subseteq V_i \times V_j$ represents a binary nested relationship between two places of different layers where $V_i, V_j \subseteq V$ ($i \neq j$), labelled by the type of relationship linking these two nodes. The labelled edge e is denoted by a tuple $\langle (v_i, v_j), label \rangle$, where $v_i, v_j \in V, v_i \neq v_j$ and $label \subset S_{label}$ is a label describing the hierarchical relation from v_i to v_j . These labels represent relations from refined to coarser grains (e.g., *County to State*) while linking the indoor and the outdoor environment on the *Road* grain level.

The spatial model further supports our trajectory modelling approach, and whose objective is to define the complete spatial environment where human mobility takes place. On top of this model, users can define *places of interest* (POIs) that might take place at any node in the model. Trajectories can then be annotated by POIs based on the spatial model, as developed in the following Section 3.3.

3.3 Spatio-temporal To Semantic Trajectories

A trajectory is considered as a series of consecutive time-dependent records that represents the evolution of the location of an individual in a given space. Similarly to Spaccapietra *et al.* [94], let us consider that a trajectory has two components; The raw/geometric component and the semantic component. The raw/geometric component is defined by a sequence of raw data that records the evolution of trajectory data and cross-related records known as contextual information (e.g., weather conditions or pollution exposition along a path). In contrast, the semantic component is denoted by a time interval sequence annotated by expressive data including contextual information. This section provides a formal definition of a contextualised trajectory and of a contextualised semantic trajectory.

3.3.1 Raw Trajectory Components

Let us first introduce the raw components of a trajectory. A time interval is defined as a convex subset of the time domain, consisting of its start time instant t_1 and end time instant t_2 . In other words, $[t_1, t_2]$ is a time interval where starting time t_1 and ending time t_2 are time instants for $t_1 \leq t_2$. Time point intervals are therefore represented as zero duration intervals where $t_1 = t_2$.

A location p denotes $p = (x, y)$ with x and y sensors logs either in indoor or outdoor spaces (e.g., GPS, RFID, respectively).

Definition 7 (Trajectory). A trajectory is defined as a consecutive temporal sequence of positions over some time intervals $[t_n, t_m]$ denoted as:

$$traj_{id} = \{p(t_n, t_{n+1}), p(t_{n+2}, t_{n+3}), \dots, p(t_{m-1}, t_m)\}. \quad (3.4)$$

$p(t_i, t_{i+1})$ denotes a location valid over a time interval $[t_n, t_m]$ and where $t_n \leq t_i \leq t_m$. If $t_i < t_{i+1}$, $p(t_i, t_{i+1})$ can be a location valid over a time interval (an instant position if $t_i = t_{i+1}$). If no position holds on t_{i+2} , this denotes a gap between $p(t_i, t_{i+1})$ and the next location will be valid at time t_{i+2} .

In order to illustrate our approach, let us consider the following example.

Example 1 : A person travels from home to office over a trajectory located using either a synchronised GPS-based sensor or sensors located at some specific places and is associated with contextual data (e.g., pollutants: NO_2 , PM_{10} , $PM_{2.5}$, PM_1 , BC). This person trajectory can be modelled as
 $traj_{person} = (\dots, p(t_{82}, t_{83}), p(t_{84}, t_{85}), p(t_{86}, t_{87}), p(t_{88}, t_{89}), p(t_{90}, t_{91}), p(t_{92}, t_{93}), p(t_{94}, t_{95}), \dots)$.

Definition 8 (Subtrajectory). Let us consider $traj_{id} \in TRAJ$ for $traj_{id} = \{p(t_n, t_{n+1}), p(t_{n+2}, t_{n+3}), \dots, p(t_{m-1}, t_m)\}$ and $traj'_{id} \in TRAJ$ for $traj'_{id} = \{p'(t_{n'}, t_{n'+1}), p'(t_{n'+2}, t_{n'+3}), \dots, p'(t_{m'-1}, t_{m'})\}$ where $TRAJ$ denotes the set of trajectories.

We say that $traj'$ is a *subtrajectory* of $traj$ denoted by $traj'_{id} \subseteq traj_{id}$ iff $[t_{n'}, t_{m'}] \subseteq [t_n, t_m]$ and $p'(t_{n'}, t_{n'+1}), p'(t_{n'+2}, t_{n'+3}), \dots, p'(t_{m'-1}, t_{m'}) \subseteq p(t_n, t_{n+1}), p(t_{n+2}, t_{n+3}), \dots, p(t_{m-1}, t_m)$ where $[t_n, t_m]$ and $[t_{n'}, t_{m'}]$ are two time intervals.

Definition 9 (Continuous subtrajectory). Let us consider $traj_{id}, traj'_{id} \in TRAJ$. $traj'_{id}$ is denoted as a *continuous subtrajectory* of $traj_{id}$ denoted by $traj'_{id} \sqsubseteq traj_{id}$ iff $traj'_{id} \subseteq traj_{id}$ and $\forall p(t_i, t_{i+1}) \in traj_{id}$ and $[t_i, t_{i+1}] \subseteq [t_{n'}, t_{m'}]$, $p(t_i, t_{i+1}) \in traj'_{id}$. Less formally, this means that there are no elements missing between $t_{n'}$ and $t_{m'}$.

In Figure 3.2, we illustrate the difference between Trajectory, Subtrajectory and Continuous subtrajectory.

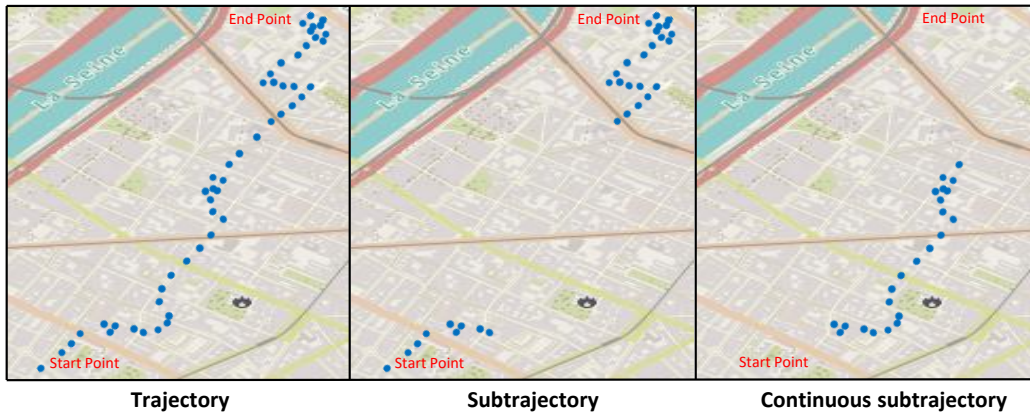


FIGURE 3.2: Trajectory, subtrajectory and continuous subtrajectory examples.

Let us consider contextual information closely related to the notion of trajectory. Contextual information are considered as a set of additional contextual knowledge that can reflect other semantics associated with a specific phenomenon, where $C = \bigcup_{i=1} c_i$ is the set of contexts that denotes the semantic dimensions for each $i \in \mathbb{N}$.

In order to provide a similar formal definition of contextual information than the one of trajectory, let us define the concept of contextual sequence.

Definition 10 (Contextual Sequence). A contextual sequence is defined as time series and closely related to the notion of trajectory as follows:

$$CS_{id} = \{v(t_n, t_{n+1}), v(t_{n+2}, t_{n+3}), \dots, v(t_{m-1}, t_m)\} \quad (3.5)$$

where each $v(t_k, t_{k+1})$ denotes a valid value of c_i over $[t_k, t_{k+1}]$ and where $t_n \leq t_k \leq t_m$.

Example 2 : Let us consider the scenario of Example 1. The contextual sequences cross-related to the person's trajectory as related to a NO_2 sensor.

Hence, $CS_NO2_{person} = (\dots, v(t_{80}, t_{81}), v(t_{82}, t_{83}), v(t_{84}, t_{85}), v(t_{86}, t_{87}), v(t_{88}, t_{89}), v(t_{90}, t_{91}), v(t_{92}, t_{93}), v(t_{94}, t_{95}), \dots)$.

Definition 11 (Contextual subsequence). Let us consider the two contextual sequences $CS_{id} = \{v(t_n, t_{n+1}), v(t_{n+2}, t_{n+3}), \dots, v(t_{m-1}, t_m)\}$ and $CS'_{id} = \{v'(t_{n'}, t_{n'+1}), v'(t_{n'+2}, t_{n'+3}), \dots, v'(t_{m'-1}, t_{m'})\}$ for each $c_i \in C$ and $[t_n, t_m]$, $[t_{n'}, t_{m'}]$ two time intervals.

We say that CS'_{id} is a *subsequence* of CS_{id} , denoted $CS'_{id} \subseteq CS_{id}$ if $[t_{n'}, t_{m'}] \subseteq [t_n, t_m]$ and $\{v'(t_{n'}, t_{n'+1}), v'(t_{n'+2}, t_{n'+3}), \dots, v'(t_{m'-1}, t_{m'})\} \subseteq \{v(t_n, t_{n+1}), v(t_{n+2}, t_{n+3}), \dots, v(t_{m-1}, t_m)\}$.

Definition 12 (Contextual continuous subsequence). Let us consider CS_{id} and CS'_{id} two contextual sequences. We say that CS'_{id} is a *continuous subsequence* of CS_{id} denoted $CS'_{id} \sqsubseteq CS_{id}$, iif $CS'_{id} \subseteq CS_{id}$ and $\forall v(t_i, t_{i+1}) \in CS_{id}$ and $[t_i, t_{i+1}] \subseteq [t_{n'}, t_{m'}]$, $v(t_i, t_{i+1}) \in CS'_{id}$.

Less formally, this means that there are no elements missing between $t_{n'}$ and $t_{m'}$.

3.3.2 Semantic Trajectory Components

The semantic trajectory is a different way of representing the raw location and its associated contextual data. It is made up of different components. Let us now introduce the semantic components of a trajectory.

The first component is the *spatial semantic sequence*, whose objective is to match a given human trajectory to a sequence of places associated over a time interval to a spatial semantic $([t_1, t_2], place)$. This means that this trajectory passes through a place labelled *place* at a start time t_1 and left it at end time t_2 , where $place \in S$. We consider that a change of location of a human being does not necessarily mean a semantic transition. On the contrary, a specific trajectory location does not always belong to a given place. Overall, for a given trajectory, this denotes a sequence of geo-annotated places over some given time intervals in indoor and outdoor spaces.

Contextual semantic sequence is the second component. It aims to extract from quantitative time series a qualitative high-level structure in order to transform crowd-sourcing raw data into contextual state sequences. Let us consider an interval-based temporal abstraction as introduced in [92] to support a qualitative data representation. More specifically, for each context $c \in C$, the *Contextual sequence* CS is

mapped towards an *interval-based* sequence of $([t_1, t_2], sem_v)$ that denotes the *contextual semantic*. Figure 3.3 illustrates examples of extracting qualitative values from quantitative time series data.

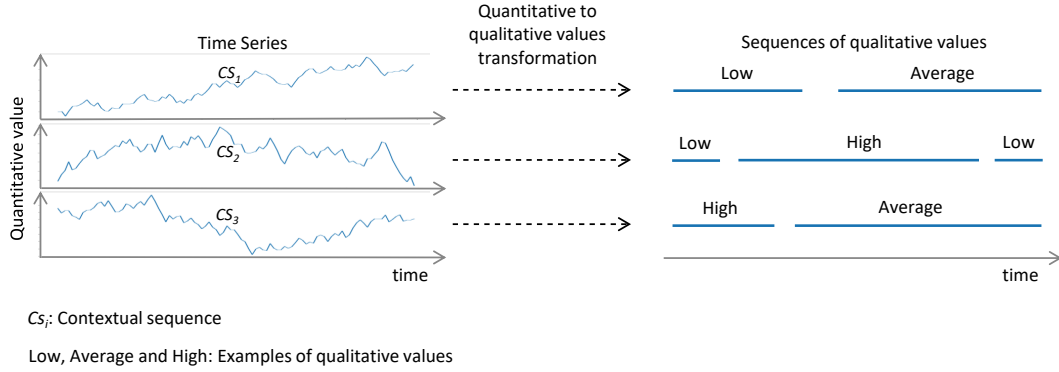


FIGURE 3.3: Temporal abstractions from time series data.

A semantic trajectory represents the continuous move of an individual while integrating contextual information semantics of the indoor and outdoor environments. This reflects an integration of spatial and contextual semantics associated to a given moving individual over some overlapping time intervals. The spatial annotation, either in indoor space or outdoor spaces, is closely associated with the predefined spatial model introduced in section 3.2 and associated with different semantic annotations that embed trajectory cross-related phenomenon.

The semantic trajectory is segmented to define the valid trajectory chunks uniformly over time.

Definition 13 (Semantic Trajectory Segment). Let us define a semantic trajectory $semTraj_{id}$ as a sequence of semantic trajectory segments. A semantic trajectory segment $seg = ([t_s, t_e], Sem_k)_{k \in [1, l]}$ denotes a homogeneous part of a trajectory valid over the time interval $[t_s, t_e]$ where Sem_k represents a list of spatial (*place*) and contextual semantics $c_i \in C$ ordered by time that hold during $[t_s, t_e]$.

Definition 14 (Semantic Trajectory). A semantic trajectory is defined as a sequence of semantic trajectory segments denoted by

$$semTraj_{id} = \{seg_1, seg_2, \dots, seg_n\} \quad (3.6)$$

Let us introduce the following semantic trajectory example.

Example 3 : The semantic trajectory is modelled as: $semTraj_{person} = \{ \dots, ([t_{82}, t_{83}], [space : place_1, NO_2 : Medium]), ([t_{84}, t_{89}], [space : place_1, NO_2 : Low]), ([t_{90}, t_{91}], [space : place_2, NO_2 : Low]), ([t_{92}, t_{95}], [space : place_2, NO_2 : High]), \dots \}$. We limit the contextual dimension to NO_2 semantic for clarification.

A semantic trajectory is more intuitively represented as a graph of annotated time interval sequences since graphs are good at materialising complex relationships (Fig. 3.4).

Definition 15 (Semantic Trajectory Graph). A semantic trajectory graph is defined as $G_{ST} = (V_{ST}, E_{ST})$.

$$V_{ST} = V \cup \bigcup_{i=1}^n V_TI_i \cup \bigcup_{i=1}^m V_C_i \quad (3.7)$$

where V represent the spatial entities, $\bigcup_{i=1}^n V_TI_i$ are the vertices that represent the time intervals while $\bigcup_{i=1}^m V_C_i$ denote the vertices that represent the contextual semantic values.

$$E_{ST} = E \cup \bigcup_{i=1}^k NEXT_i \cup \bigcup_{i=1}^j HAS_VALUE_i \quad (3.8)$$

where E represent the spatial hierarchical relations, $\bigcup_{i=1}^k NEXT_i$ denotes the ordered links between the time intervals $\bigcup_{i=1}^j HAS_VALUE_i$ denote links between time intervals and specific semantic values as illustrated in Figure 3.4. These links are annotated by the specific contextual value they are linked to.

The next edge underlines a moving object passing to a new state upon a semantic value change over at least one dimension (spatial or contextual).

G_{ST} can be decomposed to two graphs (Fig. 3.4), one corresponds to the spatial dimension (upper graph) of the semantic trajectory while the second correspond to the other dimensions of the semantic trajectory (lower graph). The spatial dimension embeds the spatial hierarchy for both indoor and outdoor spaces to secure the semantic mobility continuity in both spaces. It also provides further manipulation at different granularities.

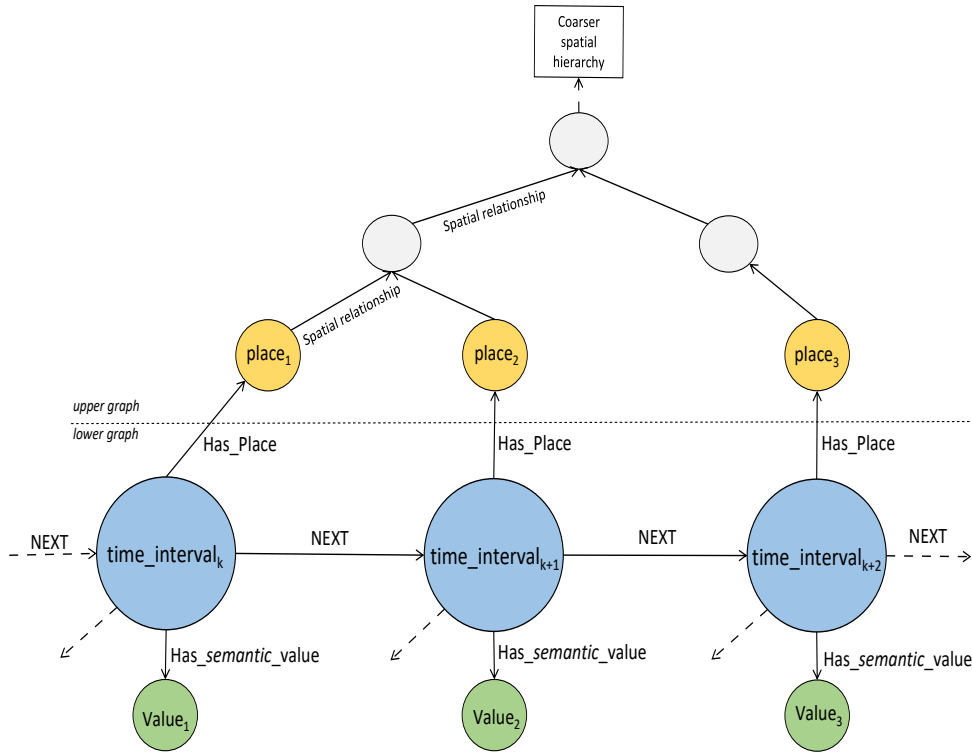


FIGURE 3.4: Semantic trajectory represented as a graph.

3.4 Multiple Views of Semantic Trajectories

Users have a wide range of interests in multi-dimensional approaches for manipulating semantic trajectories emerging in indoor and outdoor spaces. While some users essentially want to visualise the data straightly, others aim at exploring further details, higher-level information and aggregated knowledge. Different levels of abstraction should be considered to provide a flexible enough representation of crowd-sourced semantic trajectories that continuously arise in indoor and outdoor spaces. This should facilitate the derivation of different points of view and interpretations, from micro to macro granularity, and provide analysis at the straight, aggregated or micro-level, depending on the user's interest.

In this section, we develop a derivation view mechanism that extracts, from a generic semantic trajectory, different flexible views derived from the spatial, temporal and contextual semantics.

The view concept is defined as a function that takes into account a user's interest in extracting a trajectory view:

$$SemTrajView: semTrajid \times UserInterests \rightarrow viewTrajid$$

where $viewTraj$ denotes some data extracted from $semTraj$ according to some $UserInterests$ and $UserInterests$ denotes a set of spatial and/or contextual semantic preferences/criteria and is given as:

$$UserInterests = \{Spatial : Layer, Contextual : Semantic\} \quad (3.9)$$

where *Spatial* denotes set of layers (*Layer*) and *Contextual* the associated set of semantics (*Semantic*). For example, *UserInterests* can denote some specific layers (e.g., Town, Road), or/and a semantic value of a contextual dimension (e.g., NO_2).

Let us consider again the Example 3 as presented in Section 3.3.2:

Example 3: The semantic trajectory is modelled as: $semTraj_{person} = \{..., ([t_{82}, t_{83}], [space : place_1, NO_2 : Medium]), ([t_{84}, t_{89}], [space : place_1, NO_2 : Low]), ([t_{90}, t_{91}], [space : place_2, NO_2 : Low]), ([t_{92}, t_{95}], [space : place_2, NO_2 : High]), ...\}$. We limit the contextual dimension to NO_2 semantic for clarification.

Example 4 (cont'd): Let us consider the following dual of *SemTrajView* function:

(2.1) $SemTrajView(semTraj_{person}, \{Spatial : Town; NO_2 : Low\})$ where Town is the granularity target for a *Low* level of NO_2 context

(2.2) $SemTrajView(semTraj_{person}, \{Contextual : NO_2\})$.

We derive two types of views that are a hybrid representation based on either spatial or contextual criteria. The semantic trajectory is a generic representation, while the hybrid semantic trajectory is based on places of interests (cf. Section 3.4.1) and the contextual-based semantic trajectory on a context of interest (cf. Section 3.4.3).

3.4.1 Hybrid Trajectory Representation

Place-based locations are linked to their hierarchical layers, this favouring the specification of a given trajectory at different levels of abstraction and of the hierarchy. For instance, when a user is interested in a finer spatial granularity for some contextual criteria of interests and coarser spatial granularity for others (or vice versa), the previous semantic trajectory is still a flat-like representation of the trajectory and do not provide such flexibility. For that purpose, the semantic trajectory representation is extended by a hybrid semantic trajectory concept whose purpose is to extract from a flat-like representation a hybrid one.

The hybrid semantic trajectory can be expressed at either homogeneous or heterogeneous levels of granularity and according to the user needs. The idea is to provide a flexible representation of the semantics associated with a human trajectory according to different application needs and user interests.

Definition 16 (Hybrid Semantic Trajectory). A hybrid semantic trajectory is denoted by:

$$hybTrajid = \{([t_{s1}, t_{e1}], place_1, CS_1), ([t_{s2}, t_{e2}], place_2, CS_2), \dots, ([t_{sn}, t_{en}], place_n, CS_n)\} \quad (3.10)$$

where $place_i$ belongs to a layer $L \in \mathbb{S}$ either in indoor or outdoor spaces and CS_i is a set of semantic values that hold during $[t_{si}, t_{ei}]$ with $t_{ei} < t_{si+1}$. $place_i$ represents a spatial place of interest that is an instance of a given layer of interest.

For example, $place_i$ can be a floor or a town instance that belongs to the 'floor' or 'town' layers of the indoor or outdoor space, respectively. Therefore, POIs can be expressed at different levels of granularity and abstraction.

For example, at a coarse level of granularity, a human trajectory can be represented from a $town \in L_{Town}$ to another town and at a fine level from a $room \in L_{Room}$ to another room in indoor and from a road or building to another road or building in outdoor. Such layers are chosen according to the users' interests.

The main difference between $hybTraj$ and $semTraj$ is that $place$ in $semTraj$ belongs to the finest layer of granularity in the spatial model, and this for either indoor or outdoor spaces, although in $hybTraj$ can belong to any layer of the spatial model.

Furthermore, users' interests are expressed by some criteria of interest, and where the layer granularity can be adjusted accordingly. Contextual values or spatial categories can be defined accordingly by the hybrid representation. For example, a contextual value and user interest can be $\{temperature : hot\}$ and a spatial category and user interest can be $\{category : restaurant\}$.

According to the user interests, the hybrid semantic trajectory considers one spatial layer for its POIs when one of the user interests is available and a second spatial layer when none of the user interests is valid. In other words, given a level of reference, which is the finer one, it searches for a specified spatial granularity when one of the user interests is valid and another specified spatial granularity when there are no valid user interests.

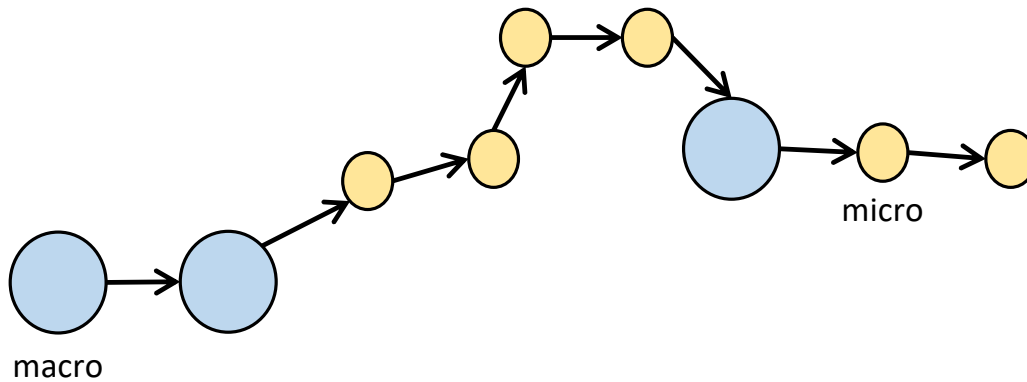


FIGURE 3.5: Hybrid semantic trajectory.

The hybrid trajectory defines a trajectory associated with a hybrid space representation of different granularities according to the indoor and outdoor spaces, and where micro and macro places can be considered and a discrete representation of the spatial dimension (Fig. 3.5). For example, let us consider that the user interest denotes a high level of NO_2 exposition for a $place$ spatial layer of interest and a $town$ spatial layer of interest if user interests are not valid. In this case, the trajectory can be represented from a place (either in indoor or outdoor spaces) when the criteria are found and to a town when the criteria are not found and vice versa. In this example, all POIs belong to $\{place, town\}$ layers.

Example 5 (cont'd): Let us consider the function $SemTrajView$ of the Example 4.(2.1). $viewTraj$ is the equivalent of $hybTraj$ in this type of view.

Let us consider that $place_1, place_2 \subseteq town_1$. In this example, $SemTrajView$ function searches for the spatial granularity from L_{Place} to L_{Town} layer for each $semTraj_{person}$ segment having a *Low* level of NO_2 . Otherwise, spatial granularity takes place at L_{Place} layer. So for the second and the third segments of $semTraj_{person}$, the function searches for L_{Town} granularity when extracting the hybrid trajectory for $hybTraj_{person}$.

A temporal aggregation occurs for the two segments of $[t_{84}, t_{89}]$ and $[t_{90}, t_{91}]$ since these two elements have the same annotations. Hence,
 $hybTraj_{person} = \{ \dots, ([t_{82}, t_{83}], [space : place_1, NO_2 : Medium]), ([t_{84}, t_{91}], [space : town_1, NO_2 : Low]), ([t_{92}, t_{95}], [space : place_2, NO_2 : High]), \dots \}$
 for $place_1, place_2 \in L_{Place}$ and $town_1 \in L_{Town}$.

This example shows a hybrid trajectory from a micro granularity (L_{Place}) to a macro granularity (L_{Town}) and vice versa. Note that L_{Place} includes indoor and outdoor places. $viewTraj_{person}$ is illustrated in Figure 3.6.

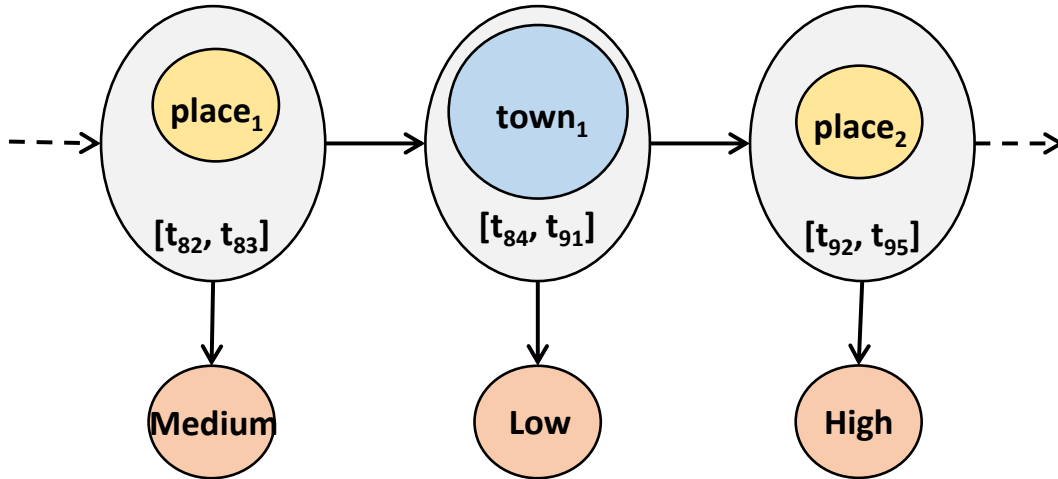


FIGURE 3.6: Hybrid semantic trajectory (Example 5).

3.4.2 Hybrid Trajectory Operations

In order to extract a hybrid representation of some semantic trajectories, let us introduce a series of operations that are embedded in Algorithm 1. Algorithm 1 shows the extraction process from a semantic trajectory to a hybrid trajectory. Algorithm 1 embeds operations presented in the following.

3.4.2.1 Operation 1

The first operation extracts a hybrid trajectory from a semantic trajectory according to layers of interest. It is defined as:

$$f1: semTraj_{id} \times layer_1 \times layer_2 \rightarrow hybTraj_{id}$$

Algorithm 1 Extract hybrid trajectory with spatio-temporal aggregation

```

1: input: semantic trajectory  $semTraj_{id}$ ;  $UserInterests : \{Spatial : Layer, Contextual : Semantic\}$  where  $Layer = \{layer_1 \in \mathbf{S}, layer_2 \in \mathbf{S}\}$  is a set of spatial layers of interest and  $Semantic$  is a set of values of interest.
2: output: hybrid semantic trajectory  $hybTraj_{id}$ 
3:  $hybTraj_{id} \leftarrow \emptyset$ 
4: for each  $([t_i, t_j], Sem_k) \in semTraj_{id}$  do
5:    $place_k \leftarrow \emptyset$ 
6:    $CS_k \leftarrow \emptyset$ 
7:    $place_{k-1} \leftarrow \emptyset$ 
8:    $CS_{k-1} \leftarrow \emptyset$ 
9:   if  $Semantic$  not null then
10:     if  $Sem_k.findContextualInterests(Semantic)$  then  $\triangleright$  //searches if one of the contextual interests is valid
11:        $place_k \leftarrow Sem_k.findSpatialLayer(layer_1)$ 
12:     else
13:        $CS_k \leftarrow Sem_k.findSpatialLayer(layer_2)$   $\triangleright$  //returns a place at a specific granularity
14:     end if
15:     else
16:        $place_k \leftarrow Sem_k.findFinerLayer(layer_1, layer_2)$   $\triangleright$  //one indoor layer and one outdoor layer
17:     end if
18:      $CS_k \leftarrow Sem_k.getContextualDimensions()$ 
19:      $place_{k-1} \leftarrow hybTraj_{id}.getPlace(k-1)$   $\triangleright$  //get the last  $place$  in  $hybTraj_{id}$ 
20:      $CS_{k-1} \leftarrow hybTraj_{id}.getContextualDimensions(k-1)$   $\triangleright$  //get the last  $CS$  in  $hybTraj_{id}$ 
21:     if  $place_{k-1} \neq \emptyset$  AND  $place_{k-1} == place_k$  AND  $CS_{k-1} == CS_k$  then
22:        $hybTraj_{id}.updateEndTimeInterval(k-1, t_j)$   $\triangleright$  //change end time interval of segment k-1 with  $t_j$ 
23:     else
24:        $hybTraj_{id}.addSegment([t_i, t_j], place_k, CS_k)$   $\triangleright$  //add a new segment
25:     end if
26: end for
27: return  $hybTraj_{id}$ 

```

that takes as parameters a list of spatial layers of interests $layer_1 \in \mathbb{I}$ and $layer_2 \in \mathbb{O}$, or, $layer_1 \in \mathbb{O}$ and $layer_2 \in \mathbb{I}$. $f1$ extract from the semantic trajectory the output hybrid semantic trajectory. The list of layers of interest represents the user's interest and refers to the indoor and outdoor spatial models. $\forall([t_i, t_j], Sem_k) \in semTraj_{id}$, $f1$ searches in the hierarchy of the spatial semantic the finest layer between $layer_1$ and $layer_2$ denoted by $place_k$ and associate it with the contextual dimensions semantics denoted by CS_k to get the k element/segment of the hybrid trajectory $([t_i, t_j], place_k, CS_k) \in hybTraj_{id}$.

Example 6 (cont'd): Let us consider the function $f1(semTraj_{person}, \{layer_1 : Floor; layer_2 : Town\})$ with $place1 \subseteq floor_1$ with $floor_1 \in L_{Floor}$ and $place2 \subseteq town_1$ with $town_1 \in L_{Town}$. Hence, $viewTraj_{person} = \{\dots, ([t_{82}, t_{83}], [space : floor_1, NO2 : Medium]), ([t_{84}, t_{89}], [space : floor_1, NO2 : Low]), ([t_{90}, t_{91}], [space : town_1, NO2 : Low]), ([t_{92}, t_{95}], [space : town_1, NO2 : High]), \dots\}$.

This example shows a hybrid trajectory from a micro (L_{Floor}) to a macro granularity (L_{Town}) and vice versa as illustrated in Figure 3.7.

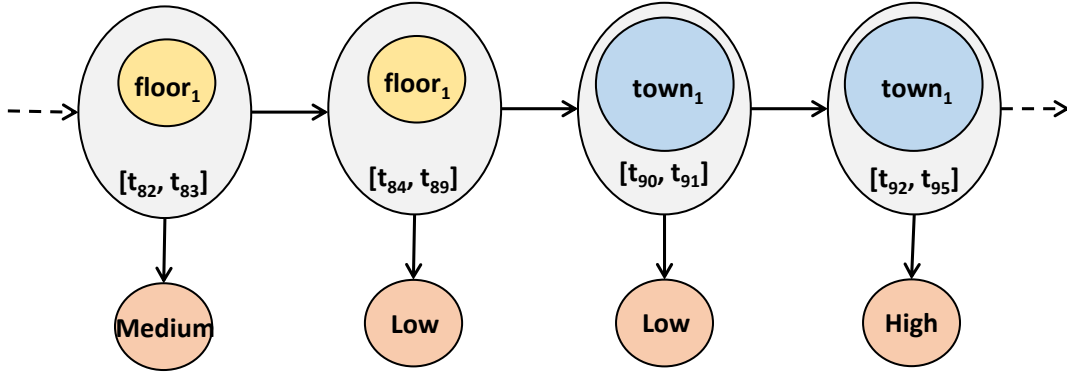


FIGURE 3.7: Hybrid semantic trajectory (Example 6).

One may be interested in extracting a hybrid trajectory representation according to the semantic trajectory dimensions values. The following operation targets the spatial dimension category.

3.4.2.2 Operation 2

The second operation extracts a hybrid trajectory from a semantic trajectory according to spatial layers and semantic values of interest. It is defined as:

$$f2: semTraj_{id} \times layer_1 \times layer_2 \times SC \rightarrow hybTraj_{id}$$

where SC is set of geo-tagged semantic values related to $place$ and $layer_1, layer_2 \in \mathbb{S}$ represent two layers of interest. $\forall([t_i, t_j], Sem_k) \in semTraj_{id}$, if $\exists sc \in SC$ where $sc \in place_k$ for $place_k \in Sem_k$, $f2$ searches for $layer_1$ hierarchy of spatial semantic of $semTraj_{id}$. Otherwise, $f2$ searches for $layer_2$ hierarchy of $semTraj_{id}$ spatial semantic. The found hierarchy layer $place_k$ is associated with the contextual semantics denoted by CS_k to get the k element of the hybrid trajectory $([t_i, t_j], place_k, CS_k) \in hybTraj_{id}$. For example, this operation provides a representation oriented towards a specific category of places (e.g., restaurants, highways).

Example 7 (cont'd): Let us consider the function $f2(semTraj_{person}, \{place_category : restaurant; layer_1 : Place; layer_2 : Town\})$ where $layer_1$ is the granularity target when a $place_category : restaurant$ is valid. Let us consider that $place_1$ is a restaurant. Hence, $viewTraj_{person} = \{\dots, ([t_{82}, t_{83}], [space : place_1, NO_2 : Medium]), ([t_{84}, t_{89}], [space : place_1, NO_2 : Low]), ([t_{90}, t_{91}], [space : town_1, NO_2 : Low]), ([t_{92}, t_{95}], [space : town_1, NO_2 : High]), \dots\}$. This example shows a hybrid trajectory from a micro (L_{Place}) to a macro granularity (L_{Town}) and vice versa as illustrated in Figure 3.8.

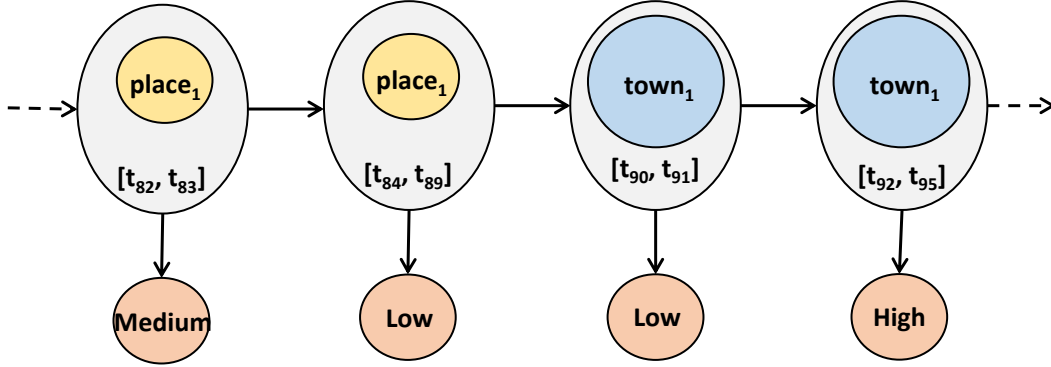


FIGURE 3.8: Hybrid semantic trajectory (Example 7).

3.4.2.3 Operation 3

The third operation focuses on the contextual dimension properties. It extracts a hybrid trajectory from a semantic trajectory according to spatial layers and contextual values of interest. Operation 3 is defined as:

$$f3: semTraj_{id} \times layer_1 \times layer_2 \times CV \rightarrow hybTraj_{id}$$

where CV is a set of contextual semantic values that belongs to C . $\forall([t_i, t_j], Sem_k) \in semTraj_{id}$, if $\exists cv \in CV$, where $cv \in Sem_k$, $f3$ searches for $layer_1$ hierarchy of $place_k$. Otherwise, $f3$ searches for $layer_2$ hierarchy of $place_k$. The resulting hierarchy layer $place_k$ is associated with the contextual dimensions semantics denoted by CS_k to get the k element of the hybrid trajectory $([t_i, t_j], place_k, CS_k) \in hybTraj_{id}$. For example, this operation provides a representation oriented towards specific contextual dimensions values (e.g., high pollution value, in a bus) to extract the hybrid spatial granule.

Example 8 (cont'd): Let us consider a third example of operation with $f3(semTraj_{person}, \{NO_2 : Low; layer_1 : Place; layer_2 : Town\})$. The output is the $hybTraj_{person}$ presented in Example 5.

$layer_1$ and $layer_2$ represent two spatial layers. There are also optional parameters for all operations regarding the time interval and the contextual dimensions to consider in order to constrain the extraction scope if needed.

3.4.3 Contextual-based Trajectory Representation

The semantic trajectory representation supports multiple dimensions, including the spatial and contextual ones. However, and in many cases, a user might require a specific view from such a generic representation, and this by taking into account some contextual criterion. For that purpose, we extend the generic model by the contextual-based semantic trajectory concept that aims to extract from the generic/flat semantic trajectory representation a new one based on one context of interest.

Definition 17 (Contextual-based Semantic Trajectory). A contextual-based semantic trajectory is denoted by:

$$contTraj_{id} = \{([t_{s1}, t_{e1}], coi_1, O_1), ([t_{s2}, t_{e2}], coi_2, O_2), \dots, ([t_{sn}, t_{en}], coi_n, O_n)\} \quad (3.11)$$

coi_i represents the semantic value of the context of interest and it is denoted by $\{c : value\}$. O_i is a set of the other semantic values that holds during $[t_{si}, t_{ei}]$ where $t_{ei} < t_{si+1}$. In this case, O_i represents a set of semantic values, including the spatial semantic value except for the context of interest value.

For this representation, the user interest is considered as a unique contextual dimension/semantic of $c \in C$.

Example 9 (cont'd): Let us consider $SemTrajView$ function of the Example 4.(2.2). $viewTraj$ is the equivalent of $contTraj$ in this type of view. Hence, $contTraj_{person} = \{\dots, ([t_{82}, t_{83}], [space : place_1, NO2 : Medium]), [space : place_1, NO2 : Low]), ([t_{84}, t_{89}], [space : place_2, NO2 : Low]), ([t_{90}, t_{91}], [space : place_2, NO2 : High]), \dots\}$. $contTraj_{person}$ is illustrated in Figure 3.9.

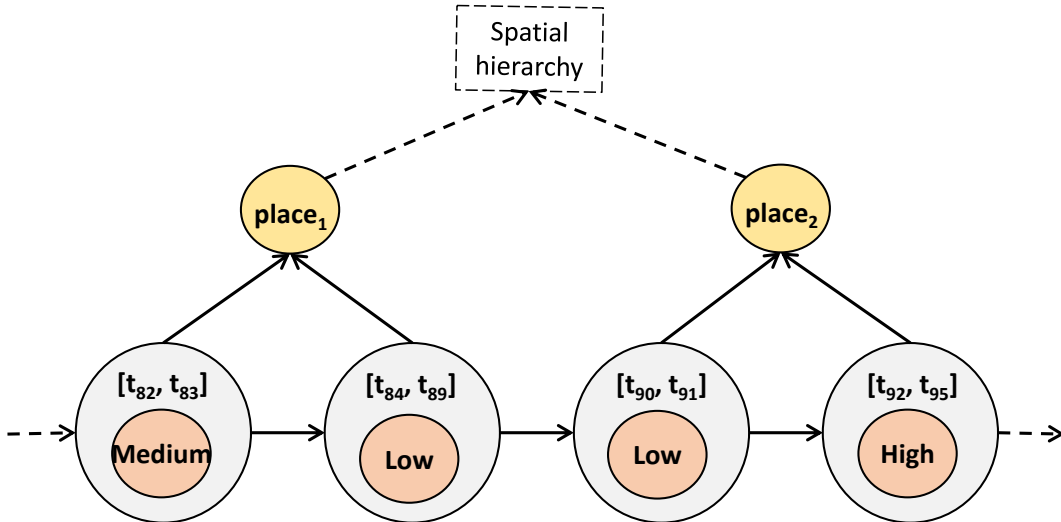


FIGURE 3.9: Contextual-based semantic trajectory (Example 9).

3.4.4 Contextual-based Trajectory Operation

In order to extract the contextual-based representation of semantic trajectories, we introduced the following operator that is embedded in Algorithm 2. Algorithm 2 shows the extraction process from a semantic trajectory to a contextual-based trajectory. It embeds the following operation.

3.4.4.1 Operation 4

The fourth operation is defined as:

$$f4: semTraj_{id} \times c \rightarrow contTraj_{id}$$

that takes as parameter a contextual dimension name $c \in C$ and extract from the semantic trajectory the output contextual-based semantic trajectory. $\forall ([t_i, t_j], Sem_k) \in semTraj_{id}$, $f4$ searches for coi_k and O_k . This operation provides a representation that focuses on a specific contextual dimension to return an output of annotated time interval sequences of the specified context state associated with the remaining dimensions, including the spatial one.

Example 10 (cont'd): We can consider the Example 9 an example of operation 3.4.4.1 with $f4(semTraj_{person}, \{contextname : NO_2\})$.

Algorithm 2 Extract contextual-based trajectory

```

1: input: semantic trajectory  $semTraj_{id}$ , a context of interest name  $c \in C$ 
2: output: contextual-based semantic trajectory  $contTraj_{id}$ 
3:  $contTraj_{id} \leftarrow \emptyset$ 
4: for each  $([t_i, t_j], Sem_k) \in semTraj_{id}$  do
5:    $coi_k \leftarrow \emptyset$ 
6:    $O_k \leftarrow \emptyset$ 
7:    $coi_k \leftarrow Sem_k.getContextualSemantic(c)$   $\triangleright$  //get the semantic value of the
      context of interests  $c$ 
8:    $O_k \leftarrow Sem_k.getOtherSemantics(c)$   $\triangleright$  //get all semantics, including spatial
      semantics, except for  $c$ . Spatial semantics include places with hierarchies
9:    $contTraj_{id}.addSegment([t_i, t_j], coi_k, O_k)$   $\triangleright$  //add a new segment
10: end for
11: return  $contTraj_{id}$ 

```

3.5 Semantic Trajectory Data Implementation and Processing Experimentation

The aim of this section is to provide the implementation methodology to experiment the formally defined models at the conceptual level. This section validates the semantic trajectory model at the data manipulation level and then validates the multiple views extraction from the semantic trajectories.

3.5.1 Data Set Description

This modelling approach has been experimented with real environmental crowd-sensing data collected in Paris region France in Polluscope project [83] where several data collection campaigns (over three campaigns) have been carried out. 86 individuals participated in the campaigns where each one were equipped with three sensors that record ambient air data (i.e., Temperature, Humidity, Particulate Matter: $PM_{2.5}$, PM_{10} , $PM_{1.0}$, NO_2 and Black Carbon) and a tablet with GPS chipsets for geo-location and a self-reporting application to manually annotate the data with activities and pollution-related behaviours [56]. The activities last for a period of time and include transportation mode (e.g., car, bus, metro) as well as indoor (e.g., home, office, restaurant) and outdoor (e.g., park, street) activities. On the other side, behaviours are temporary acts for a short period of time and include actions related to air pollution (e.g., open a window, start cooking, smoking, turn on a chimney). The users are anonymised.

3.5.2 Pre-processing Methodology

The trajectories collected from Polluscope project [83] were outdoor mainly, since they do not contain precise location using the GPS-connected tablet. In order to provide additional indoor trajectories, a synthetic indoor spatial environment was simulated using the VITA simulator as already experimented in a related work [58]. The simulated synthetic indoor trajectories have been generated according to different random-walks and destination targets, from virtual RFID sensors and annotated, and continuously associated to each outdoor trajectory whenever possible and coherent, in order to obtain a continuous indoor/outdoor raw trajectory. A list of real outdoor trajectories of participants working on the same site was selected and merged with synthetic indoor data according to the annotation "office" the participants indicated it using their toolkit.

Data quality issues have not been considered so far, but this does not impact the principles behind our data integration and modelling approach.

Figure 3.10 illustrates an overview of the proposed framework. It is a data pipeline that includes data collection, integration, extraction, and transformation before it is stored in a database. Resolving the quality issues in the data is out of scope in this work. The inputs of the framework are synchronised data flows for each participant. The initial trajectories pre-processing step is to integrate outdoor GPS logs with indoor logs to construct the indoor and outdoor raw trajectories. A geo-tagging process is pursued where geo-tagged information are extracted from *OpenStreetMap* (OSM) [74] using the reverse geocoding. We extend the OSM outdoor data with indoor data and build spatial hierarchies in between. Time interval spatial sequences are extracted to form the spatial semantic sequences.

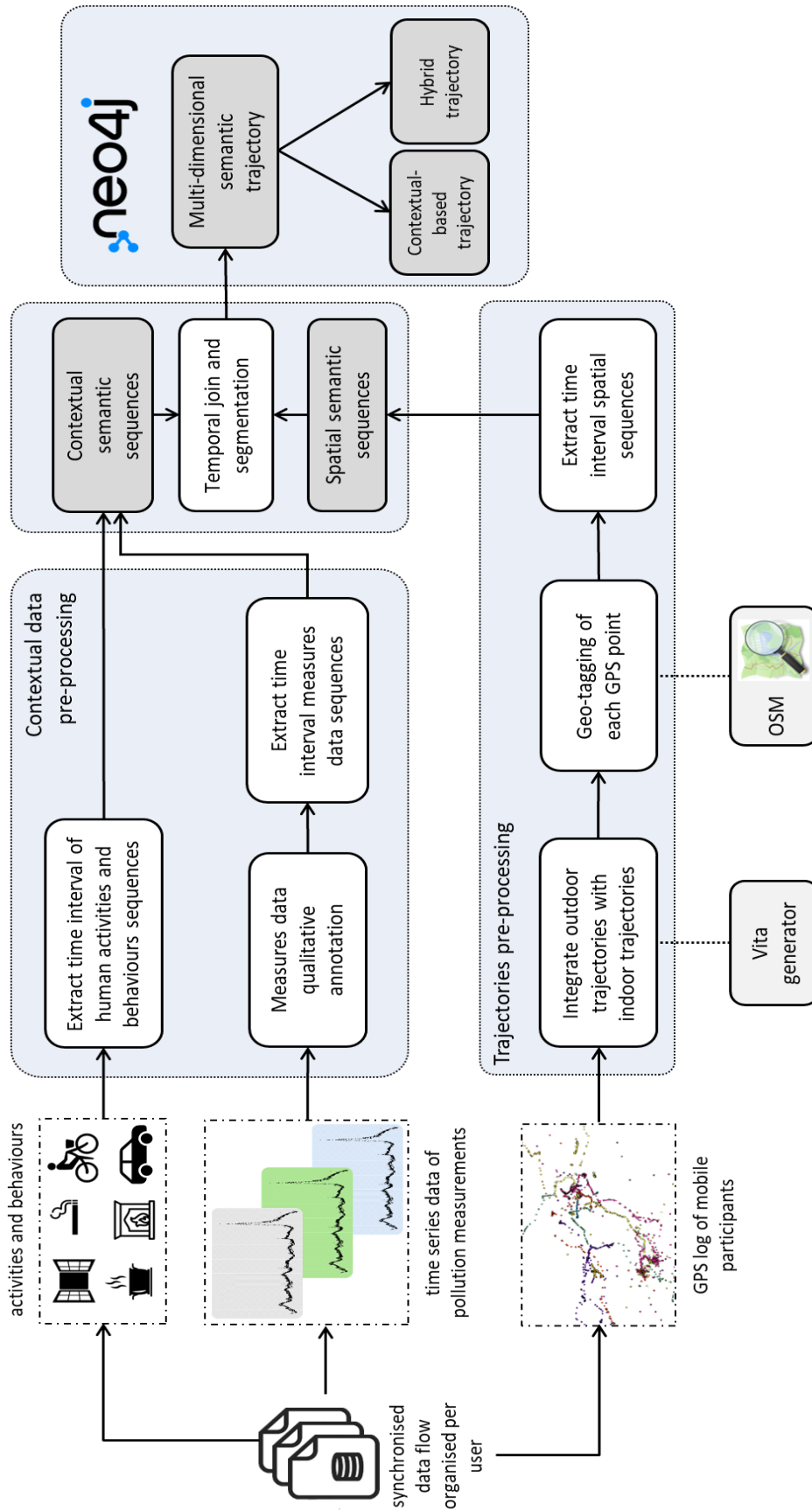


FIGURE 3.10: Pre-processing methodology.

The contextual data pre-processing steps convert time-series measures data to qualitative information, then extract time interval data sequences from these data and the activities and behaviours data to get a contextual semantic sequence for each component of these contextual data. All of these sequences are grouped in one sequence and segmented upon at least one dimension value change to get the multi-dimensional semantic trajectory sequences finally for each user.

The semantic trajectories are extracted and stored in the Neo4j [64] graph database. While the data can be stored using different database models such as relational ones, Neo4j has been logically selected since it is based on a graph-based data model and advanced data manipulation capabilities. It has been also recognised that the representation and processing of trajectory data can be naturally expressed and further manipulated as graphs and using a graph query language [42]. Moreover, the same study showed that most of data manipulations are computationally more efficient with graph databases as compared to relational ones. The main principle behind the representation of trajectories with Neo4j can be illustrated by the meta-model presented in Figure 3.11. Figure 3.11 illustrates the materialised relationships in the Neo4j graph database. Temporal functions are derived from the APOC library [65] to manipulate time interval nodes.

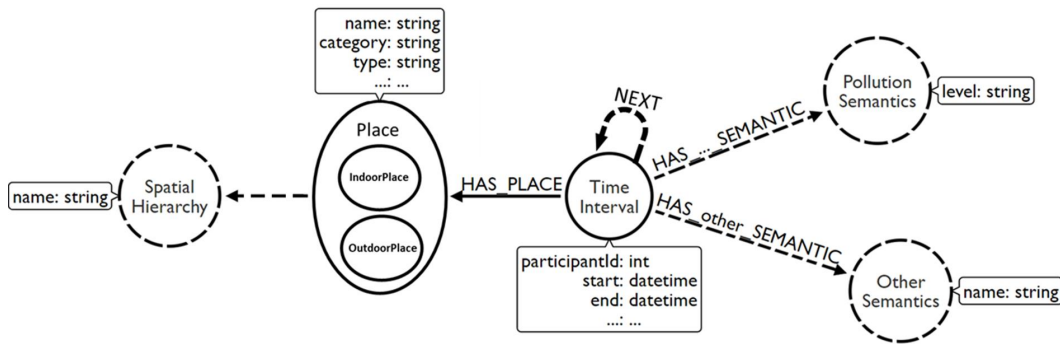


FIGURE 3.11: Multi-dimension semantic trajectory graph meta-model implemented in Neo4j.

3.5.3 Semantic Trajectories Data Manipulations

Trajectory data manipulations should be designed appropriately with specific data manipulation mechanisms at the query level, but also with well-designed processing capabilities. This should be designed according to the spatial, temporal and semantics criteria that are very likely to be part of different query schemes. This will generate some derived trajectories filtered according to the previously mentioned conditions. Such trajectory processing evaluations occur on different processing types that encompass different moving individual behaviours according to the *where*, *when*, *what* and *which* dimensions. The *where* represents the spatial component, while the *when* denotes the temporal one, the *what* the contextual one, and the *which* the individuals. The suggested model have been built to express a wide range of processing operations that manipulate all these orthogonal dimensions. The evaluation of our modelling approach is applied to two main processing categories, namely, graph queries and graph analytics. Let us hereafter introduce these categories:

- **Graph queries** retrieve or extract explicit data and patterns from features that have been modelled as a graph, that is, either from graph components or

properties using procedural query languages. Graph queries have the expressive power to return a subset of instances or a sub-graph according to some given spatial, temporal and semantic criteria derived from common graph data processing systems.

- **Graph analytics** analyse emerging graph properties using graph structural properties and statistics. This generally reflects some trends and patterns that can be computationally extracted using graph-based and statistical metrics and operators. For example, one might derive some statistics on the accumulated times some trajectories have passed within some given nodes of the graph for a given period of time and under some given constraints (e.g., high NO_2 level).

Different query categories can be applied for trajectory data analysis, for instance, searching for patterns, places, contextual information over a given period of time on some given or selected trajectories. In fact, our modelling approach should take into account the multi-dimensional properties of a trajectory representation in order to perform this diversity of queries. The combination of these different data dimensions is applied using the graph query language Cypher [27] that acts both as a support for the specification of data manipulations at the query levels, as well as the embedding of graph analytics. Graph queries and analytics can be applied to trajectories and by satisfying multiple criteria, including time, space and context, and where specific hierarchical embedding is also taken into account.

Moreover, our modelling approach makes a difference between indoor and outdoor spaces using specific labels and where each “*Place*” node is labelled by a second label either “*IndoorPlace*” for indoor or “*OutdoorPlace*” for outdoor spaces. This not only permit to constraint processing operations to indoor or outdoor places, but also the two spaces together as a hybrid environment when appropriate. A high level of semantic expressiveness should be also guaranteed by taking into account the different levels of the spatial hierarchy for both indoor and outdoor spaces.

Let us introduce a collection of queries and analytical examples that illustrate the potential of our modelling approach. The following examples manipulate the spatial, temporal and semantic dimensions. They are logically expressed using the Cypher query language.

3.5.3.1 Graph Queries Examples

We introduce a few examples in which the returned objects must satisfy a series of spatial, temporal and semantic criteria. These queries are organised according to three aspects taking advantage of our model: place-based, trajectory-based and hierarchy-based.

3.5.3.1.1 Place-based This is a typical query category that for instance returns the objects close to a series of places some trajectories passed through under some predefined constraints and over a given time interval. An example of a place-based query is to “*find places of interest within a distance of 100m from trajectories exposed to a high NO_2 level at a specific time*”. This Cypher query is specified below (Example 3.1).

EXAMPLE 3.1

```

MATCH (poi:Place)
MATCH (:No2Semantic {level: 'High'})
    <-[HAS_NO2_SEMANTIC]-(ti:TimeInterval)-[:HAS_PLACE]
    ->(p:Place)
WHERE ti.start = '2019-06-28 12:39:19+02:00'
WITH point({latitude: poi.latitude, longitude:poi.longitude})
    AS pm, point({latitude: p.latitude, longitude:p.longitude})
    AS pfixed, poi, ti
WHERE distance(pfixed, pm) < 100
RETURN poi

```

The first MATCH defines all indoor and outdoor places. The second MATCH returns for each semantic trajectory, the time-constrained semantic trajectory segments with a high level of NO_2 . Each segment is associated with a place and a semantic denoting a high level of NO_2 exposition at a specific time. Distances from objects of interest to these time-constrained places are then derived. The results are finally filtered using an Euclidean distance threshold of 100m. Figure 3.12 visualises the results on the map.

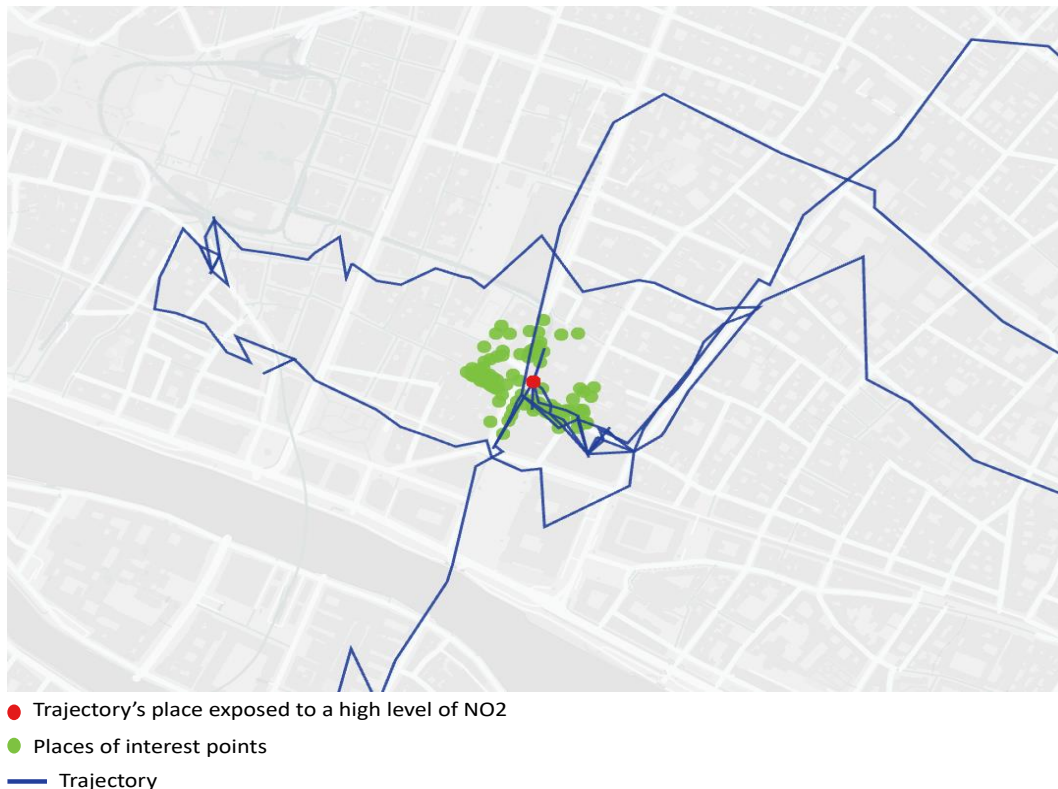


FIGURE 3.12: Visualisation example of a place-based query. The red point represents the place where a trajectory were at a specific time exposed to a high level of NO_2 . The green points represent the places of interest at less than 100m distance from the red point.

3.5.3.1.2 Trajectory-based Queries can also return some trajectories selected according to some given place-based criteria. For instance, a query might return a list of pre-selected trajectories according to some spatial, temporal and semantic criteria close enough to a given place of interest. An example of such a trajectory-based query is to “find all trajectories that were near a place of interest that exhibited a high level of NO_2 exposure over a distance constraint”. This Cypher query is shown below (Example 3.2).

EXAMPLE 3.2

```

MATCH (poi:Place {name:'Versailles-Chantiers'})
MATCH (:No2Semantic {level: 'High'})
  <-[:HAS_NO2_SEMANTIC]-(ti:TimeInterval)-[:HAS_PLACE]
  ->(p:Place)
WITH point({latitude: poi.latitude, longitude:poi.longitude})
  AS pm, point({latitude: p.latitude, longitude:p.longitude})
  AS pfixed, poi, ti, p
WHERE distance(pfixed, pm) < 500
RETURN p, ti

```

The first MATCH clause defines a place of interest with a specific name. The second MATCH clause returns the trajectory segments associated with a high level of NO_2 exposure and related places and over a distance constraint. Finally, this query returns the trajectory segments at a maximum distance of 500 meters as (*place node*, *time interval node*). Figure 3.13 illustrates the final results. Figure 3.13 (a) shows all trajectories exposed to a high NO_2 level at a maximum distance of 500m from the place of interest. Figure 3.13 (b) illustrates two trajectories exposed to a high level of NO_2 denoted by red points.

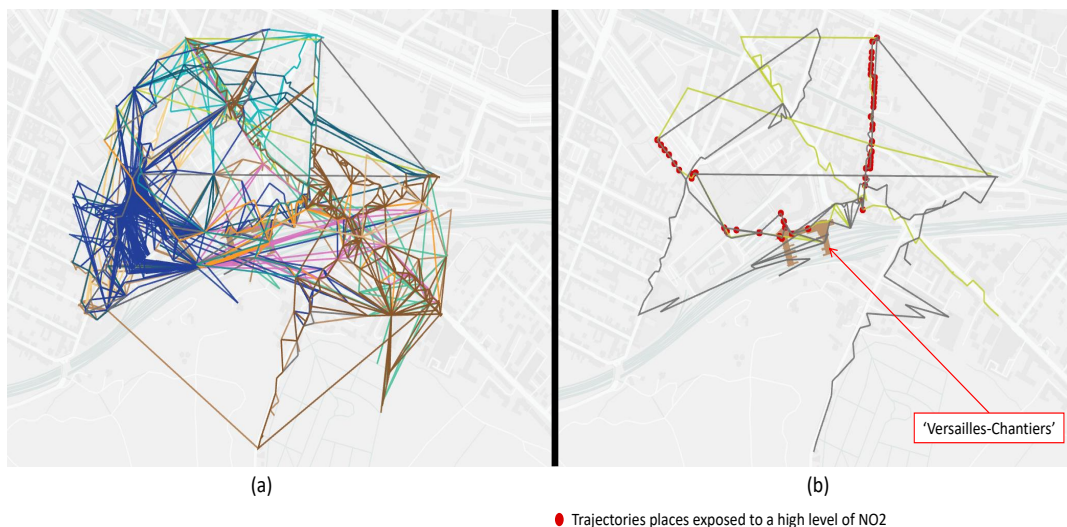


FIGURE 3.13: Trajectory based query example visualisation. (a) trajectories at less than 500m distance from the place of interest. (b) two example trajectories at less than 500m distance from the train station Versailles-Chantiers associated to trajectories exposed to a high level of NO_2 .

Under some specific adaptations it is worth noting that similar queries might be applied to indoor spaces.

3.5.3.1.3 Hierarchy-based This query addresses the hierarchical aspect of the proposed model while considering the multi-dimensional constraints. It takes into account the indoor and outdoor trajectory continuity considered in the proposed model. A hierarchy-based query should constrain a given query by a spatial hierarchy. Given a trajectory and area of interest denoted by a spatial layer, such a query might for instance returns the objects of a given spatial layer and under some semantic, spatial and temporal criteria. An example of a hierarchy-based query (Example 3.3) is to “*find all indoor places of a specific layer, passed through some trajectories during a time interval, that have been exposed to a high level of Black Carbon when they were in an outdoor place of the same specific layer during the same given temporal interval*”.

EXAMPLE 3.3

```

MATCH (t:Town {postcode:'78150'})<-[:HAS_TOWN]-(:Road)
      <-[:HAS_ROAD]-(:OutdoorPlace)
      <-[:HAS_PLACE]-(:ti1:TimeInterval)-[:HAS_BC_SEMANTIC]
      ->(bc:BCSemantic{level:'High'})
WHERE ti1.start > '2019-12-04 00:00:00+01:00'
AND ti1.end < '2019-12-05 00:00:00+01:00'
CALL apoc.path.subgraphNodes(ti1, {relationshipFilter:"NEXT"})
YIELD node AS ti2
MATCH (t)<-[:HAS_TOWN]-(:Road)<-[:HAS_ROAD]-(:Building)
      <-[:HAS_BUILDING]-(:Floor)<-[:HAS_FLOOR]-(:ip:IndoorPlace)
      <-[:HAS_PLACE]-(:ti2)
WHERE ti2.start > '2019-12-04 00:00:00+01:00'
AND ti2.end < '2019-12-05 00:00:00+01:00'
AND ti2.participantID = ti1.participantID
RETURN ip

```

The first MATCH clause sets the indoor semantic trajectory segments associated to a high black carbon level for an identified region and over a given time interval. The `apoc.path.subgraphNodes` function returns all successive segments of each trajectory. The second MATCH clause returns the resulting sequence of semantic trajectory segments associated with an indoor place of a given spatial layer. The query finally returns the indoor places node that match.

3.5.3.2 Graph Analytics Examples

Graph analytics uses graph structural properties to reveal some statistics and trends. One can analyse trajectories satisfying multiple criteria including time, space and context, and where the spatial semantic can also reflect a specific hierarchy. An example of that is to “*compute the number of trajectories that crossed a specific road within a particular county over a time period while exposed to a high level of NO_2* ”. This query is illustrated below (Example 3.4).

EXAMPLE 3.4

```

MATCH (:No2Semantic {level:'High'})
  <-[:HAS_NO2_SEMANTIC]-(ti:TimeInterval)-[:HAS_PLACE]
  ->(p:Place)-[:HAS_ROAD]
  ->(r:Road {name: 'Rue de Versailles'})-[:HAS_TOWN]
  ->(:Town)-[:HAS_COUNTY]->(:County {name:'Hauts-de-Seine'})
WHERE ti.start > '2019-11-01 00:00:00'
AND ti.end < '2019-11-31 00:00:00'
RETURN count(DISTINCT ti.participantID)

```

The MATCH clause defines the criteria applied to the trajectories segments. The results are filtered according to the considered time interval to finally aggregate the distinct participants ids. This example shows how complex queries can be manipulated in a flexible way.

Another example of query analytics is to “*calculate the number of times each participant was exposed to a high level of NO_2 and compute their aggregated duration*”. The query is given below (Example 3.5).

EXAMPLE 3.5

```

MATCH (:No2Semantic {level: 'High'})
  <-[:HAS_NO2_SEMANTIC]-(ti:TimeInterval)
WITH apoc.date.parse(ti.end, 'ms', 'yyyy-MM-dd HH:mm:ss') -
  apoc.date.parse(ti.start, 'ms', 'yyyy-MM-dd HH:mm:ss') AS
  dur, ti
RETURN ti.participantID, COUNT(ti), SUM(dur)

```

The query sets the trajectory segments associated with a high level of NO_2 and calculates the duration (*dur*) for each segment. Then it counts for each participant the segments number and sums their durations.

One can analyse the graph properties according to some temporal constraints. For example, one can “*find the times of the day when trajectories are the most exposed to a high level of NO_2* ”. The query is given below (Example 3.6).

EXAMPLE 3.6

```

MATCH (:No2Semantic {level: 'High'})
  <-[:HAS_NO2_SEMANTIC]-(ti:TimeInterval)
WITH COUNT(ti) AS nbOfTimes,
  apoc.date.fields(ti.start, 'yyyy-MM-dd HH:mm:ssz').hours AS
  hour
RETURN hour, nbOfTimes
ORDER BY nbOfTimes DESC

```

The first MATCH clause defines the trajectory segments associated with a high level of NO_2 pollutant. The query then counts the number of these segments and groups them according to the 'hour' of the start time extracted from the trajectory segments. To extract the 'hour', the APOC library is used. Finally, the query returns, for each hour of the day, the number of times trajectory segments were associated with a high level of NO_2 . In fact, the analytical queries illustrated by examples 3.5 and 3.6 are practically specified to apply to both spaces.

Some of the aggregations performed, especially in the case of Example 3.6, can be considered as OLAP roll-up queries. Moreover, these query capabilities can be further extended by additional OLAP queries to re-fragment, slice or restructure trajectory data.

Overall, the series of graph queries and analytics illustrate the expressive power of the graph-based representation. One of the peculiarities of the approach is that it supports multiple dimensions and granularities, and advanced manipulation capabilities across the graph structural and embedded properties (i.e., node to edge relationships). The computational evaluation shows that the proposed model can express different data manipulation categories generally of interest to mobility applications, and this using the Cypher query language that embeds an expressive query language. Moreover, the whole approach shows the potential of the homogeneous representation of indoor and outdoor spaces since the user can homogeneously apply data manipulation on both spaces.

3.5.3.3 Performance Evaluation

The experimental evaluation has been performed on a Neo4j 4.2.3 Enterprise docker container running on Windows 10. The container resources are 50 CPUs, 86GB of memory and 200GB disk image size. The whole server resources are a processor Intel Xeon Platinum 8180 with 2.50GHz and 56 cores, 128GB of memory and 1TB Avago MR9460-16i SCSI disk.

In order to evaluate the execution time of the different graph query and analytics examples, three different databases have been prepared. The first one contains 10 semantic trajectories (141,898 nodes and 1,336,423 relationships), the second one with 50 semantic trajectories (380,898 nodes and 3,525,846 relationships) and the third one with 86 semantic trajectories (783,166 nodes and 7,433,059 relationships). The database designed with 50 semantic trajectories includes the 10 trajectories of the first database. Similarly, the database of 86 semantic trajectories includes the trajectories of the second database. The six query and analytics examples are executed five times on each database. We noticeable remark the effect of the Neo4j cache regarding differences between the execution time of the first and the following iterations since Neo4j applies various caching mechanisms. In fact, when a given query is executed for the first time it caches nodes and relationships used for subsequent executions.

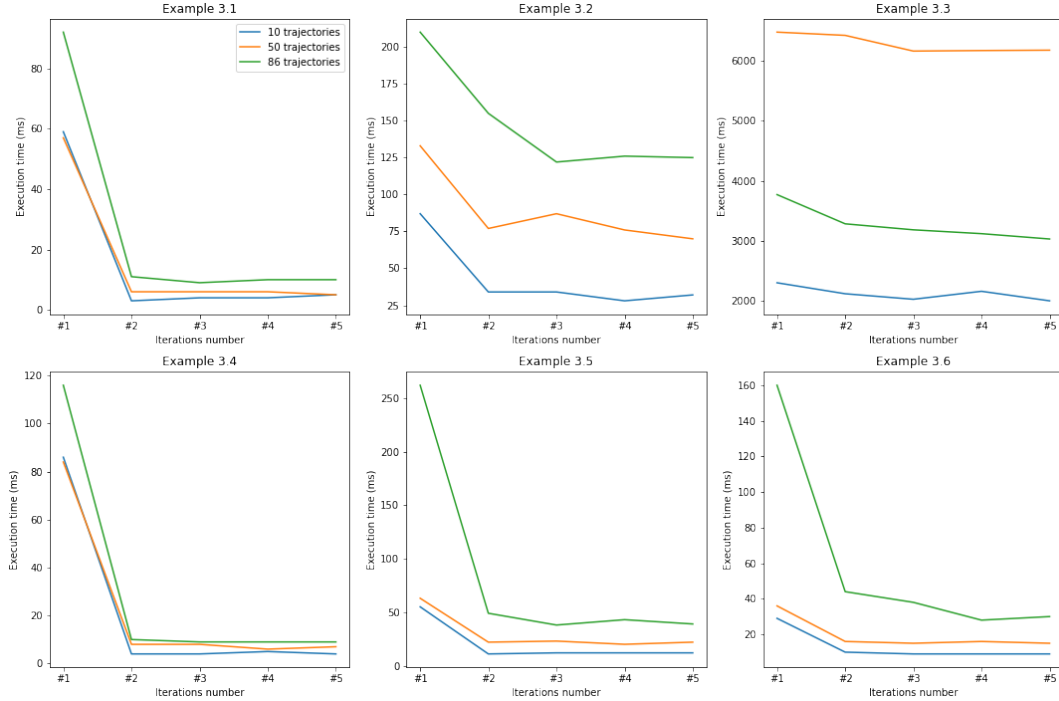


FIGURE 3.14: Execution time for each example iterated five times on three different databases sizes.

Overall, execution times vary from a few milliseconds to a few tenth of milliseconds for all the examples except *example 3.3* that reaches a few thousands of milliseconds (Figure 3.14). The experiments show that computational times increase linearly with the size of the database. We report the mean of four execution time after the first one. For instance, *example 3.1* execution with 10 trajectories takes 4ms, with 50 trajectories takes 5.75ms while it takes 10ms for 86 trajectories. *Example 3.2* execution with 10 trajectories takes 32ms, with 50 trajectories takes 77.5ms, while it takes 132ms for 86 trajectories. *Example 3.3* execution with 10 trajectories takes 2082.75ms, with 50 trajectories takes 2730.25ms while it takes 3151.75ms for 86 trajectories. *Example 3.4* execution with 10 trajectories takes 4.25ms, with 50 trajectories takes 7.25ms, while it takes 9.25ms for 86 trajectories. *Example 3.5* execution with 10 trajectories takes 11.75ms, with 50 trajectories takes 21.75ms, while it takes 42.25ms for 86 trajectories. *Example 3.6* execution with 10 trajectories takes 9.25ms, with 50 trajectories takes 15.5ms, while it takes 35ms for 86 trajectories.

As a summary, execution times remain acceptable despite the fact that spatial, contextual and temporal constraints are applied. Execution times of *example 3.1*, *example 3.4*, *example 3.5* and *example 3.6* are the faster ones. These very fast execution times are due to the fact that these examples manipulate a very limited number of semantic trajectory segments and under limited constraints. The *example 3.2* is slightly slower but still fast. This is because this query does not limit semantic trajectory segments by a temporal constraint, hence more nodes to select and more distances to calculate. However, the *example 3.3* is the slower one. This is due to the fact that it involves a transitive closure between semantic trajectory segments over some time interval nodes linked by NEXT relationships. We limited the hierarchical-based query by a temporal interval and space layer in order to limit the number of transitive closure calculations; if not, the execution time will significantly increase.

The experiments show that most of our queries performed in linear execution

times, this being relatively comparable to current indoor or outdoor related works ([51, 115]) although these approaches are based on different modelling principles, making the comparison not completely relevant.

3.5.4 Multiple Views Extraction Implementation and Experimentation

The multiple views approach has been experimented with real environmental crowd-sensing data described in Section 3.5.1.

Pre-processing steps are discussed in Section 3.5.2 and illustrated in Figure 3.10.

3.5.4.1 Neo4j Multiple Views Extension and Experimentation

Neo4j is an open-source Java-based graph database that allows extending its functionality with user-defined procedures. These procedures can be easily added as plugin packed in a .jar file so that they can be invoked directly from the Cypher query language. The two aforementioned algorithms, Algorithm 1 and Algorithm 2 have been implemented as user-defined procedures to extend Neo4j as illustrated in Figure 3.10.

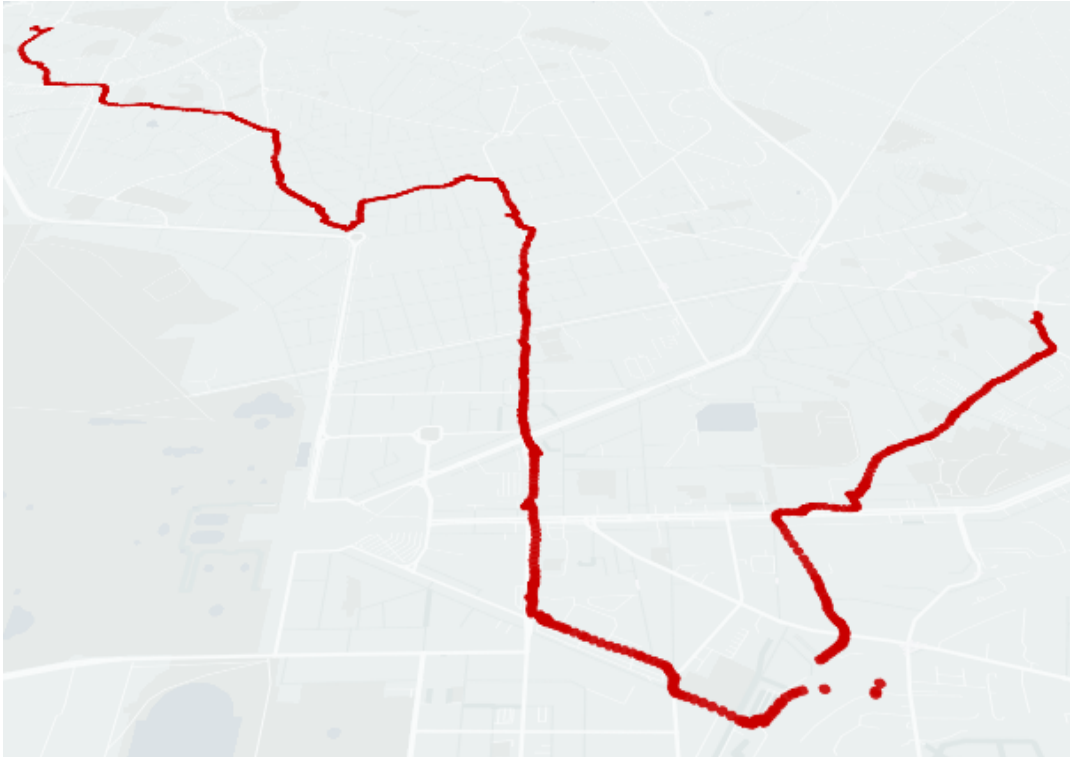
We show experiments on the following three examples of operations:

- (a) $f2: semTraj_{100} \times Road \times Town \times category: tourism \rightarrow hybTraj_{100}$
- (b) $f3: semTraj_{100} \times Place \times County \times NO_2: High \rightarrow hybTraj_{100}$
- (c) $f4: semTraj_{100} \times Activity \rightarrow contTraj_{100}$

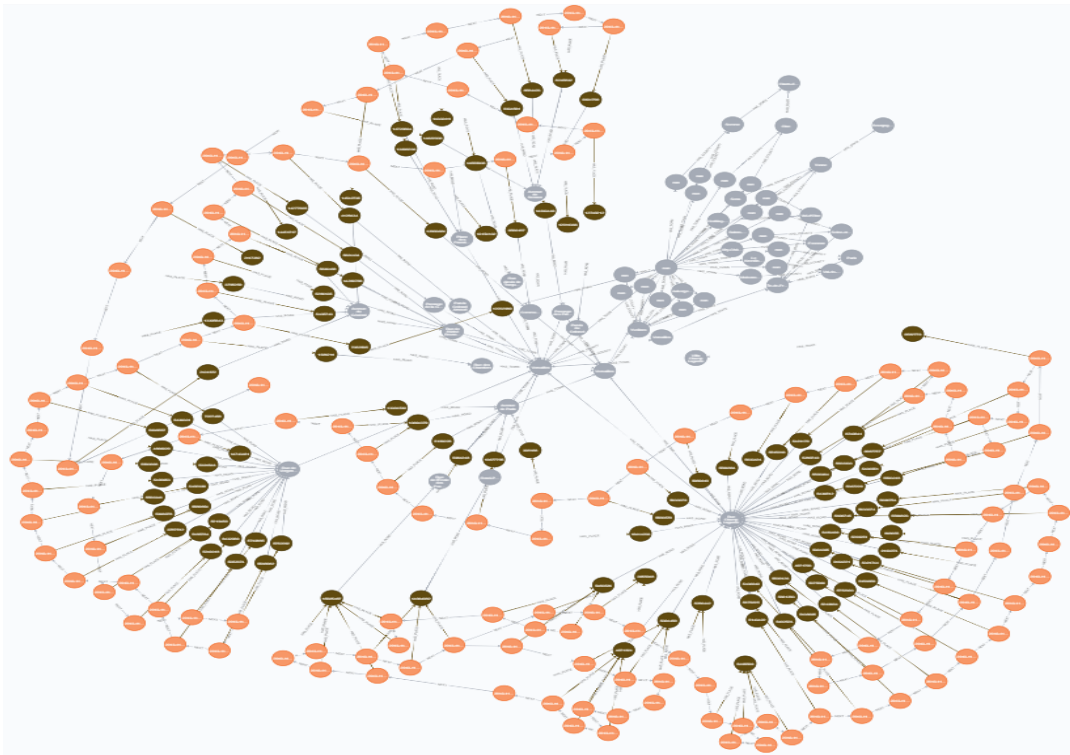
The operation (a) extract hybrid trajectories with ‘Road’ granularity when there is a *tourism* place and a ‘Town’ granularity otherwise. The operation (b) extracts hybrid trajectories with ‘Place’ granularity when there is a High level of NO_2 exposition, a ‘County’ granularity otherwise. The operation (c) extracts from the semantic trajectories the contextual-based trajectories on ‘Activity’ dimension.

In order to show the difference between raw trajectories and the different semantic representations of the trajectories, we discuss in the following the output views of a portion of one trajectory in each representation and after each operation in the Neo4j graph. The experiments are applied on a participant trajectory of 40 minutes going from work office to home. The outputs can be explored interactively at the Neo4j interface level.

Firstly, let us show the raw trajectory portion on the map in Figure 3.15a. The semantic trajectory graph of this portion is illustrated in Figure 3.15b. We limited its semantic dimensions to the spatial dimensions hierarchies. Figure 3.15b shows the semantic trajectory obtained from the raw trajectory illustrated in Figure 3.15a. Orange nodes in this figure represent time intervals, and they are linked to different spatial dimension hierarchy (grey, brown nodes). Interactive exploration of this semantic trajectory allows the user to inspect the sequences places or specific time interval.



(A) Raw trajectory.



(B) Semantic trajectory graph in Neo4j.

FIGURE 3.15: Two illustrations of a raw trajectory and its semantic trajectory representation in Neo4j (semantic dimensions limited to the spatial dimension hierarchy).

Let us apply operation (a) (query in Listing 3.7) to the same part of the trajectory to get the hybrid semantic trajectory illustrated in Figure 3.16.

LISTING 3.7: Query using operation (a) example

```
CALL multiViews.hybridTrajectoryGraph(100,
  {category:'tourism'}, 'Road', 'County')
MATCH (poi:POI {participantID:'100'})
MATCH (poi)-[:HAS_NO2_SEMANTIC]->(no2:No2Semantic)
MATCH (poi)-[:HAS_PM10_SEMANTIC]->(pm10:Pm10Semantic)
MATCH (poi)-[:HAS_PM1_0_SEMANTIC]->(pm1:Pm1_0Semantic)
MATCH (poi)-[:HAS_PM2_5_SEMANTIC]->(pm2:Pm2_5Semantic)
MATCH (poi)-[:HAS_BC_SEMANTIC]->(bc:BCSemantic)
MATCH
  (poi)-[:HAS_TEMPERATURE_SEMANTIC]->(temp:TemperatureSemantic)
MATCH (poi)-[:HAS_HUMIDITY_SEMANTIC]->(hum:HumiditySemantic)
MATCH (poi)-[:HAS_ACTIVITY_SEMANTIC]->(act:ActivitySemantic)
MATCH (poi)-[:HAS_BEHAVIOUR_SEMANTIC]->(eve:BehaviourSemantic)
WHERE poi.start>='2019-11-30 15:36:00+01:00'
AND poi.end<'2019-11-30 16:16:00+01:00'
RETURN *
```

The first clause of Listing 3.7 calls the procedure that extracts the hybrid semantic trajectory for the participant with id ‘100’. The user interests are a ‘Road’ granularity for places with a ‘tourism’ category and a ‘Town’ granularity for others. The query then matches the extracted hybrid trajectory segments associated with multiple contextual semantics and returns a portion between a time interval. It returns the hybrid trajectory of the indicated time interval finally (illustrated in Figure 3.16).

One can notice the semantic movement, represented by orange nodes, from ‘Town’ granularity to a finer place at the ‘Road’ granularity (Villa Champ Lagard) since the participant was in a place categorised by *tourism* in that road. Other nodes colours represent the different contextual semantics associated with the trajectory.

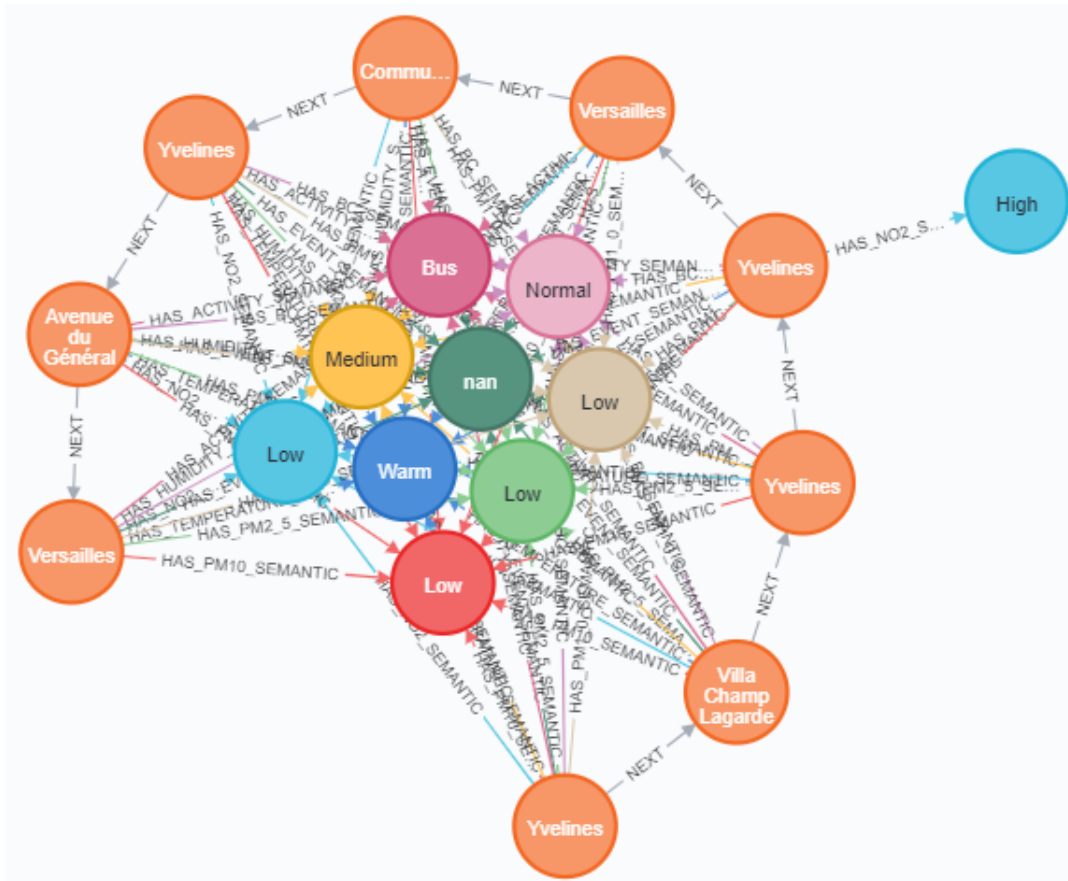


FIGURE 3.16: Hybrid semantic trajectory path graph (operation (a)). Orange nodes represent *places* where time intervals are embedded. Coloured nodes represent other contextual dimensions.

When applying the operation (b) (query in Listing 3.8) on the same portion, one get the hybrid semantic trajectory graph illustrated in Figure 3.17.

LISTING 3.8: Query using operation (b) example

```
CALL multiViews.hybridTrajectoryGraph(100,
  'No2Semantic',{level:'High'}, 'Place', 'County')
MATCH (poi:POI {participantID:'100'})
MATCH (poi)-[:HAS_NO2_SEMANTIC]->(no2:No2Semantic)
WHERE poi.start>='2019-11-30 15:36:00+01:00'
AND poi.end<'2019-11-30 16:16:00+01:00'
RETURN *
```

The first clause of Listing 3.8 extracts the hybrid semantic trajectory for the participant with id '100'. The user interests are a 'Place' granularity for places associated with a 'No2Semantic' with 'High' value and 'County' granularity for other 'No2Semantic' values. The query then matches the extracted POIs with their associated NO_2 semantics and finally returns the extracted hybrid trajectory during a given time interval (illustrated in Figure 3.17).

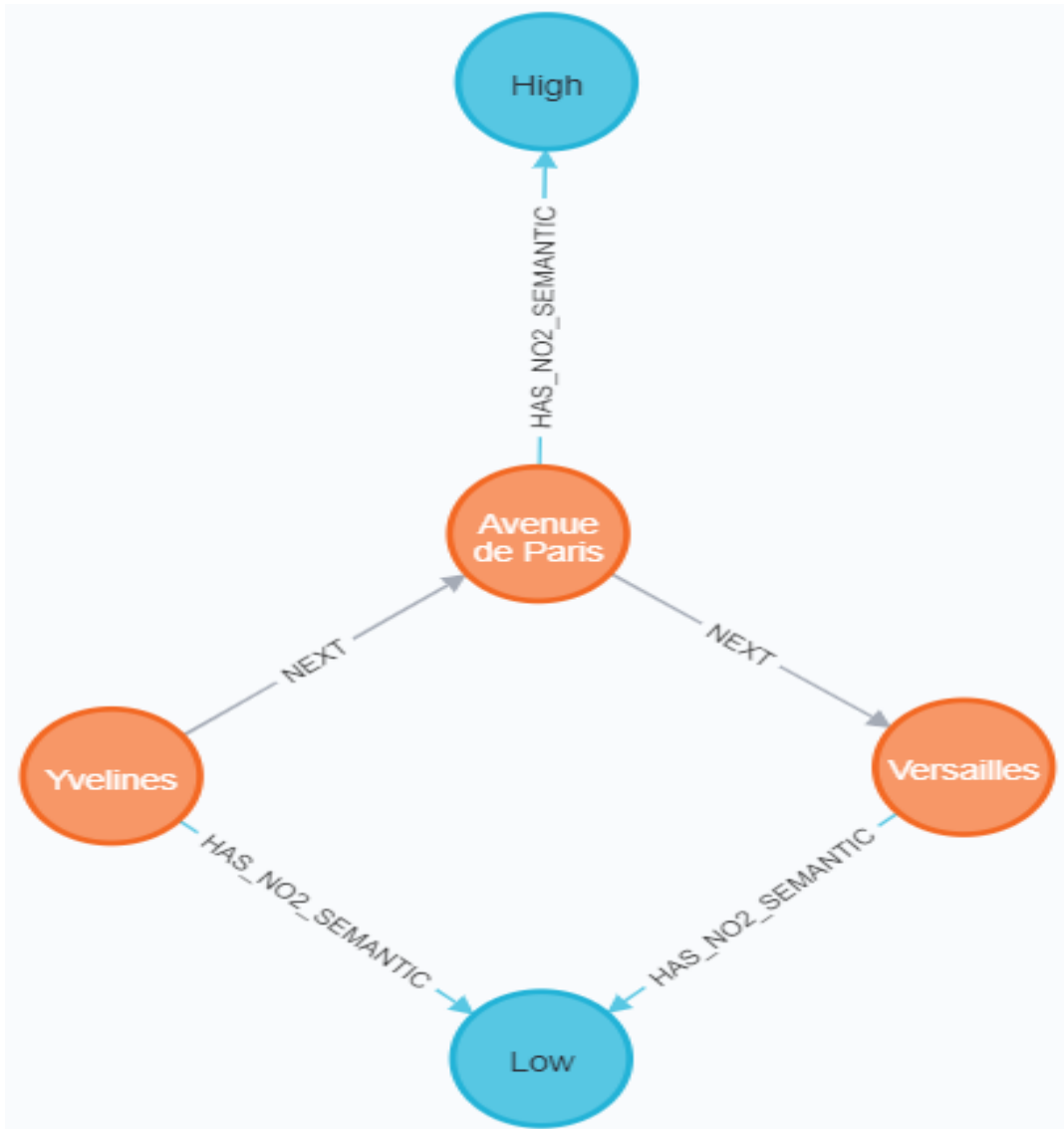


FIGURE 3.17: Hybrid semantic trajectory path graph (operation (b)). Orange nodes represent *places* where time intervals are embedded. Blue nodes represent other contextual dimensions. We limit the contextual dimension to NO_2 semantic for clarification.

We hide all the dimensions except the NO_2 dimension for a better illustration. This highlights how spatial and temporal aggregations impact the trajectory to get a reduced view. This view shows the clear states of the trajectory regarding the user interests, high level of NO_2 at ‘Road’ granularity (Avenue de Paris) and ‘Town’ granularity for other NO_2 levels (Yvelines and Versailles).

Let us apply the operation (c) (query in Listing 3.9) on the same part of the trajectory to obtain the contextual-based trajectory on ‘Activity’ context of interest. The first clause extracts the contextual based semantic trajectory for the participant with id ‘100’ and with ‘ActivitySemantic’ context of interest. Then, it matches the extracted *cois* and finally returns the contextual-based trajectory during an indicated time interval (illustrated in Figure 3.18). Figure 3.18 shows a piece of this part so that we can clearly inspect the *cois* trajectory from ‘Bus’ to ‘Rue’ to ‘Domicile’.

LISTING 3.9: Query using operation (c) example

```

CALL multiViews.contextualBasedTrajectoryGraph(100,
  'ActivitySemantic')
MATCH (coi:COI {participantID:'100'})-[:HAS_PLACE]->(p:Place)
WHERE coi.start>='2019-11-30 15:36:00+01:00'
AND coi.end<'2019-11-30 16:16:00+01:00'
RETURN coi, p

```

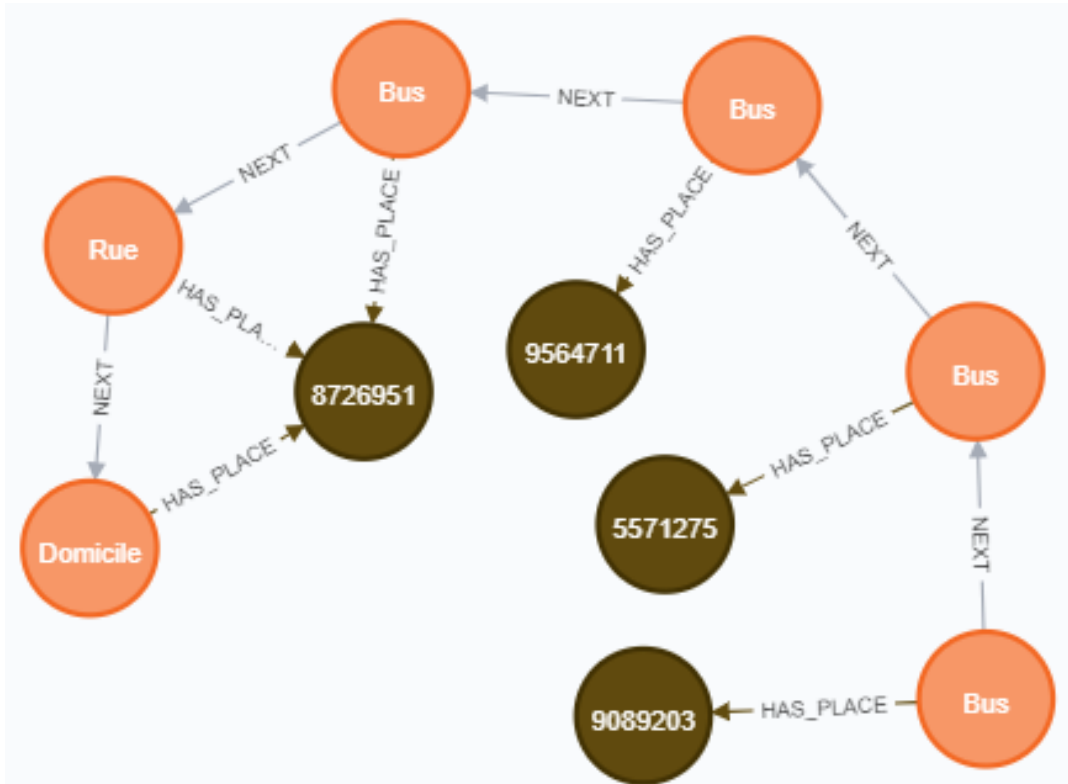


FIGURE 3.18: Contextual-based semantic trajectory path graph (operation (c)). Orange nodes represent the context of interests (*coi*) where time intervals are embedded and brown nodes represent the spatial dimension (the spatial hierarchies are hidden for a better illustration).

3.5.4.2 Performance Evaluation

The experimental evaluation has been performed on Neo4j 4.2.3 (Enterprise) running on Windows 10. The hardware configuration is as follows: 6 cores Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz, the machine has 32 GB in terms of RAM and a fast storage of 239 GB.

In order to evaluate the execution times of the different operations examples, three different databases have been prepared. The first one contains 10 semantic trajectories, the second one 40 semantic trajectories and the third one 86 semantic trajectories. Average execution times are presented in Figure 3.19.

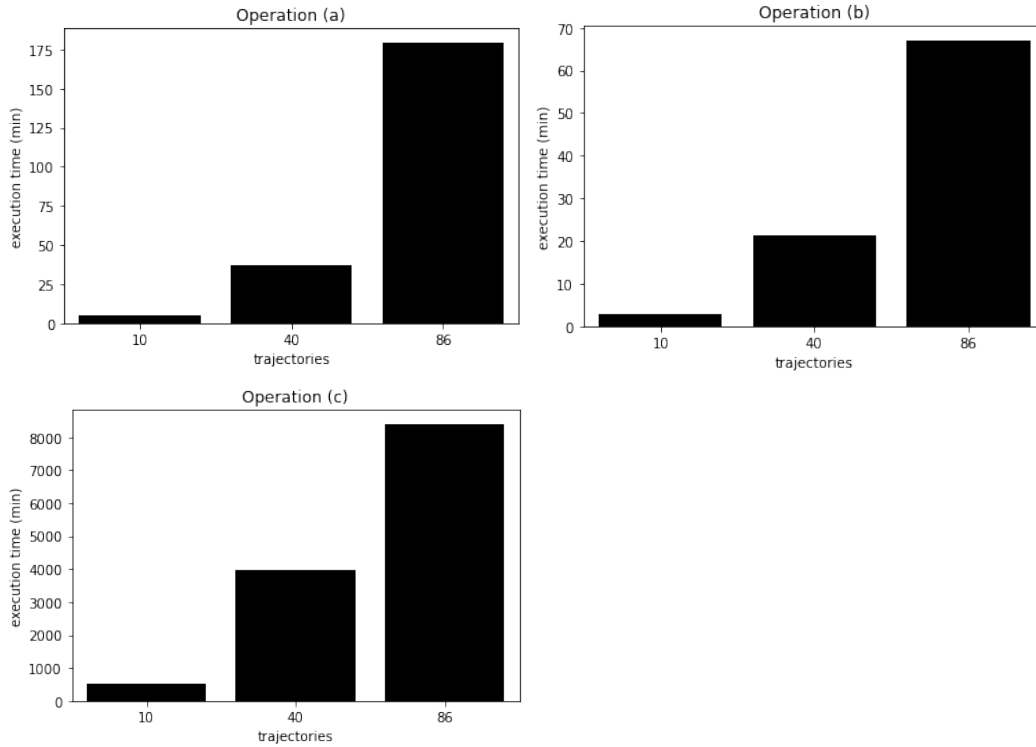


FIGURE 3.19: Execution times of the operation examples

Execution times vary depending on the trajectories length and user interests. One can notice execution times relatively costly and that increase with the number of trajectories; this is due to the fact that these operations extract the entire trajectories and re-manage the representation for each trajectory according to the desired new representation that depends on the user interests. The results of these operations are stored in the graph database in order to apply further queries and analytics. Operation (c) takes additional computational time since there are no temporal and spatial aggregations to highly reduce the amount of nodes and relationships. (c) operation combines each time interval node with the contextual dimension that represents the context of interest operation.

3.6 Discussion

The semantic trajectory model shows how it can successfully integrate multiple crowd-sourced data semantics in a polluted urban environment, including matching the trajectories to place-based spatial annotations and environmental data. The model is flexible enough to embed trajectory data semantics at different granularities. The hierarchical approach denoted by the indoor and outdoor spatial model proves how it enriches the analysis at the query manipulation level. One peculiarity of the model is that it is possible to commute between different semantic dimensions. The experiments on the semantic trajectory manipulations show how the model provides high-level semantic manipulation capabilities at a very accepted run-time. However, the model is experimented with using only one data set. It needs to be experimented with other domain contexts.

The multiple views framework experiments show how this approach supports the generation of different views of the crowd-sourced trajectories in a well-defined

graph structure with associated contextual data. Overall, these trajectory views can be derived according to different spatial, temporal and semantic criteria that provide flexible data manipulation capabilities at different levels of granularity and user interests. However, the intrinsic nature of such a model may involve some loss of information that may occur on some of the represented dimensions. The relevance of the whole data manipulation capabilities and performance figures still have to be experimented in a context of mobility patterns that arise in both outdoor and indoor spaces. There is still a need to experiment with additional query mechanisms and data structures to improve computational times.

As for now, outdoor trajectories only were derived from real data, although indoor trajectories were simulated.

3.7 Conclusion

The research presented in this chapter introduces a semantic trajectory framework applied to indoor and outdoor spaces. Our modelling approach deal with crowd-sourcing environments where multiple dimensional annotations are considered, including a place-based spatial annotation. Indoor and outdoor spaces are homogeneously and formally modelled by a multi-layered graph where trajectories occur and are represented. A semantic trajectory is defined as a sequence of annotated time intervals associated with spatial semantics embedded in the indoor and outdoor hierarchical spatial representations. Our approach is not based on stop and move abstractions but on multi-dimensional and hierarchical abstractions.

This work also provides mechanisms to extract multiple views according to a wide range of user interests, including granularity, spatial category, and contexts. The presented algorithms support from a continuous trajectory representation derivation of even a hybrid spatial and contextual view of trajectories from micro to macro levels and vice-versa.

The framework has been implemented on top of the graph database system Neo4j. A series of graph queries and analytics have been experimentally evaluated. The results show high data manipulation capabilities regarding querying and analytical operations, and where a user can easily commute her/his query expressions between the hierarchical indoor and outdoor layers, the contextual and temporal semantic dimensions. Designing the processing operations using our model shows a wide range of expressiveness using the Cypher query language of Neo4j, where such data manipulations can operate on indoor, outdoor or both spaces. Most of the processing potential of the proposed model can be applied for most common operations involved in the manipulation of trajectory data.

The multiple views approach has been formally defined and implemented on top of the Neo4j graph database. Cypher query language has been extended with different user-defined operations to support the view mechanisms. The experimentation illustrates the whole framework's potential, performance figures and its potential for trajectory data exploration, manipulation, and analysis.

While data heterogeneity is currently inherent to crowd-sourcing data and our flexible modelling approach. Indeed positioning techniques which are likely to be used in outdoor and indoor environments are very likely to rely on different principles (e.g., GPS and RFID, respectively), this generates different data precision and quality in the represented trajectories. However, such differences might be considered as implicit to

the different levels of abstraction and precision generally considered in both outdoor and indoor environments. Overall data quality and uncertainty issues have not been considered so far by our modelling approach. This might be part of our further work and, for example, by integrating uncertainty criteria at the data representation and query levels.

The overall outcome work of this chapter is instrumental in supporting mobility data with related crowd-sensed data offline analysis and exploring humans patterns in indoor and outdoor spaces.

Chapter 4

Composite Event Extraction From Stream of Semantic Trajectories

Real-time mobile crowd-sourcing data available from urban environments offers many opportunities for analysing human trajectories. This increases the need for computing systems and data architectures to process semantic trajectories data in a timely way when searching for complex mobility patterns of interest. Despite the ability to represent spatio-temporal events, there is still a lack of well-defined and flexible semantic trajectories manipulation languages to support abstraction and composition mechanisms for the analysis of urban mobilities.

This chapter develops a modelling framework for complex events applied to our human indoor and outdoor semantic trajectories. The framework is based on an Event Pattern Language that defines complex events from predefined primitive events. Complex events are modelled and defined using spatial, semantic and temporal criteria by a predefined set of composition operators.

The whole approach is experimented by a prototype that integrates semantic trajectory streams in an urban environment and ingests them into an event-based system that implements a list of complex indoor and outdoor mobility events.

4.1 Introduction

Real-time crowd-sensed data available in urban environments offers many opportunities for analysing mobility. Systems processing such data stream can be applied to monitor mobility behaviours. They became valuable solutions for providing insights on the fly when searching for mobility patterns and events of interest. In particular, detecting complex mobility patterns among crowd-sourced location data and associated contextual information in a given urban environment should be of interest to many urban planners and decision-makers. However, such patterns should be first modelled using a formal language that provides abstractions and constructs for identifying primitive and complex events of interest based on spatial, temporal and contextual criteria. This also requires identifying appropriate data structures to annotate and manipulate multi-dimensional trajectories at different levels of abstraction. Moreover, this modelling approach should support the representation of trajectories that emerge in both indoor and outdoor spaces in our case.

A typical human trajectory includes time annotated location data. These trajectories can also be annotated by cross-related contextual information such as ambient and environmental data (e.g., pollution, weather), transportation mode (e.g., bus, pedestrian) and human activities (e.g., sport, office work) [15]. Integrating these annotations can enrich the representation at the modelling level and reveal more mobility patterns at the processing level.

The research presented in this chapter relies on the aforementioned indoor and outdoor semantic trajectory model (cf. Section 3.3) in order to recognise complex events on the fly. The advantage of using our trajectory representation as a data source for the streaming system is that it is flexible enough to derive semantic trajectories according to contextual, spatial and temporal criteria. Moreover, our trajectory model supports a finer representation of continuous movement semantics in indoor and outdoor environments. What enriches the complex event recognition findings is that our semantic trajectory model encompasses multiple hierarchical granularities of the urban space while maintaining multiple dimensions of contextual associated data. Significant patterns would remain hidden without exploiting such enrichment during the complex event recognition process.

Mobility data requires flexible manipulation at the language and processing levels but with two objectives. First, extracting individual events such as “*find people who commute from home to work and are exposed to high levels of pollution*”. Next, when manipulating complex events that denote humans evolving in the urban domain, and to identify composite and collective/aggregated patterns as shown in the following example: “*Find people who commute from home to work and meet new people along their route*” and where multiple trajectories are involved in the query.

The objective of this chapter is of two principal goals:

Goal 1: The first goal of this chapter aims to develop a modelling framework that represents events using a formal language. This research goal embeds the following research challenge:

Research Challenge 4

How to model complex events using a flexible and expressive language to be detected in a real-time big data processing system among streams of semantic trajectories data?

The events defined are of two categories: basic events and complex events, formally defined by a Complex Event Processing (CEP) system. A complex event is defined in the literature as a collection of derived events that satisfy some patterns [7]. The CEP system's objectives are to model and manipulate pre-processed and multi-dimensional semantic trajectories stream made of homogeneous trajectory segments and recognise situations of interest as soon as they occur. The peculiarity of the CEP approach is that composite events of interest are formally defined and derived from basic events and denote some specific situations of interest. Complex events are developed from basic events using a predefined expressive formal language.

Our modelling approach is based on the Event Pattern Language (EPL) introduced in previous work, and that proposes a set of composition operators (e.g., sequence, relaxed sequence, conjunction, disjunction), which has been already extended and applied to geographical applications [63] [21]. It also supports our semantic trajectory representation where events are identified using spatial, semantic and temporal criteria. The objective of the composition language is to aggregate them using event-based operators. This first part of the chapter aims not only to identify basic events defined as primitive predicates but also to aggregate them using a composite-based manipulation language.

Goal 2: The second goal of this chapter aims to define the methods to implement and experiment the conceptually developed formal model with real data. This is also aimed at validating the model. The following are the two tasks that constitute the experimental goal:

- Realise a streaming event-based system implementation that integrates real-time semantic trajectories and embeds formal specifications of urban complex events. These events are categorised as individual and aggregated events that denote a wide range of urban activities.
- Evaluate the system using a demonstrative case study that embeds different composition operators.

4.2 Stream of Semantic Trajectory

4.2.1 Semantic Trajectory Data Structure

In order to represent human trajectories, let us consider a *place-based* model where a *place* denotes a location of interest in either an indoor or outdoor space (cf. section 3.2).

Places are located in *layers* organised hierarchically. Let $\mathbf{S} = \bigcup_{i=1}^k \mathbf{L}_i$ denote the entire space encapsulating indoor and outdoor spaces as a hierarchy of layers \mathbf{L}_i . In fact \mathbf{S} is made of a hierarchy of layers, and then places being defined at the most appropriate layer level.

Places are characterised by thematic properties and their 'location' in the spatial hierarchy. This spatial hierarchy denotes the different levels of abstraction considered

for the outdoor and indoor spaces (e.g., state, county, town, road, building; and building, floor, room, for outdoor and indoor spaces, respectively). This spatial representation further supports a homogeneous representation of semantic trajectories in indoor and outdoor spaces, and whose objective is to define the complete spatial environment where human mobility takes place. On top of this model, users can define places of interest (POIs) that might take place (Figure 4.1) at any granularity in the model.

As mentioned in chapter 3, we define a semantic trajectory (cf. Definition 14), and organised by semantic trajectory segments (cf. Definition 13), as a representation of an individual’s continuous movement while incorporating contextual information semantics that arise in the urban environment. It embeds some homogeneous spatial and contextual semantics associated with a mobility during a time interval. The spatial annotation is coupled with the spatial model and associated with multiple contextual semantics that integrate trajectory cross-related phenomena.

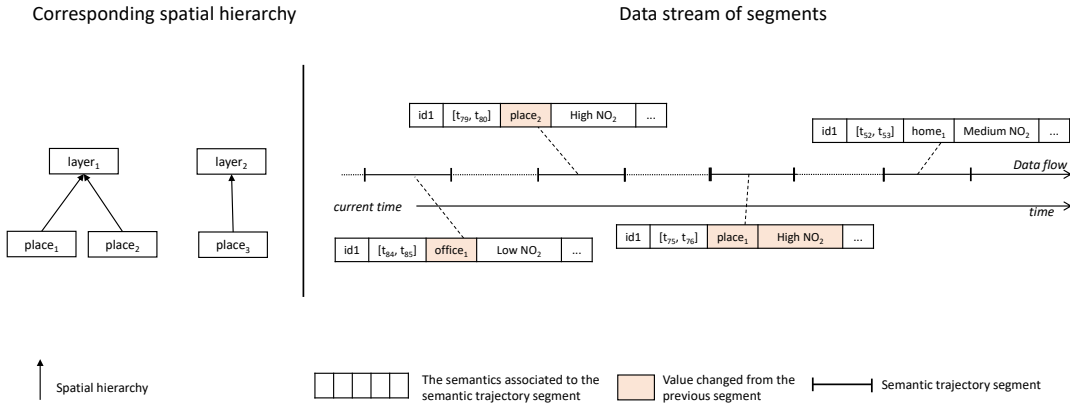


FIGURE 4.1: Example of semantic trajectory segments data flow with the corresponding spatial hierarchy.

In order to illustrate our approach, let us consider the running example as follows:

Example 11 A person having an identification $id1$ travels from home to office over a trajectory associated with contextual data (e.g., pollutants: NO_2 , PM_{10} , $PM_{2.5}$, PM_1 , BC) and located using a GPS device. Without loss of generality, let us limit the contextual dimension to NO_2 semantic to simplify the representation. The semantic trajectory is modelled as: $semTraj_{id1} = \{ \dots, ([t_{52}, t_{53}], [space : home_1, NO_2 : Medium]), \dots, ([t_{75}, t_{76}], [space : place_1, NO_2 : High]), \dots, ([t_{79}, t_{80}], [space : place_2, NO_2 : High]), \dots, ([t_{84}, t_{85}], [space : office_1, NO_2 : Low]), \dots \}$.

Overall, a semantic trajectory is defined as a sequence of annotated time intervals associated with spatial semantics embedded in the indoor and outdoor hierarchical spatial representations (Figure 4.1). A semantic trajectory represents an individual’s continuous movement while integrating multiple contextual data derived from the environment in order to enrich the representation at the modelling level and reveal more mobility patterns at the processing level.

4.2.2 Trajectory Segments Stream

Our model is applied on top of a stream processing approach that supports real-time data integration of mobility behaviours. The aim is to apply the principles behind the trajectory-based representation to data streams that are likely to provide valuable and immediate insights and identification of some situations of interest. At the primitive level, semantic trajectories are derived from streaming data. Figure 4.1 shows an example of a semantic trajectory stream where each semantic trajectory segment is derived as a tuple of attributes denoted by $(a_i)_{id} = ([t_{si}, t_{ei}], Sem_i)_{id}$ for mobile user id . Figure 4.1 shows different trajectory segments data flows and where there is also a lack of data between two consecutive trajectory data segments (illustrated by the dots). We can notice the corresponding hierarchy of the spatial annotation referring to our place-based model. A semantic trajectory stream can be defined as a time-ordered sequence of events denoted by $s = \langle a_1, a_2, \dots, a_n \rangle$.

4.3 Event Modelling

An event is defined as a real-world situation emerging from human mobility and that can be characterised according to spatial, temporal and contextual information (e.g., pollution exposure, activity). In other words, an event is an occurrence happening at a particular location, within a specific time instant or period which holds multiple properties. We consider two kinds of events: so-called basic and composite events. Basic events are derived according to some well-defined predicates that aim to extract some primitive semantic information at the spatial (e.g., a specific location at either fine or coarse granularity) and contextual level (e.g., a specific value at a certain contextual dimension) and where temporal semantics can be associated (e.g., temporal condition, duration or overlaps between different predicates). Composite events are derived according to some predicate-based composition of basic events.

4.3.1 Basic Event

Definition 18 (Basic event). A basic event e is a real-word event/action extracted from the human mobility and can be classified according to spatial, temporal and contextual (e.g., pollution exposure, activity) characteristics. A basic event can be considered as an occurrence happening at a particular *place*, within a specific time instant or period which holds multiple semantic properties (Sem_k). Basic events $e(s)$ denote primitive situations derived from semantic trajectory segments by predicates $p(e)$ according to spatial, temporal and contextual criteria.

Let us introduce some basic events considered in our approach in Table 4.1. Table 4.1 presents basic events predicates that refer to some semantic trajectory segments of interest organised according to some spatial, temporal and contextual semantics. A basic event should identify some specific mobility properties identified at a given level of granularity in the hierarchy of layers, and according to some places of interest. When considering a given phenomenon, a basic event can denote some specific attributes of interest associated to trajectory segments (e.g., pollution level, temperature, activity) and valid for a given interval of time. As the predicate is applied to the incoming streaming of trajectory data the time is the one of the current stream. The following Example 12 illustrates some basic events predicates.

Criteria	Basic event (e) predicates p	Description
Spatial	$position(id, place)$	Given a trajectory id and a given $place$, the predicate returns true if this trajectory is located at this $place$.
	$position(id, layer)$	Given a trajectory id and a given $layer$, the predicate returns true if this trajectory is located at this $layer$.
	$position(id, placeProperty)$	Given a trajectory id and a given $placeProperty$, the predicate returns true if this trajectory is located at a place having these $placeProperty$ (i.e., spatial properties e.g., within a range, type restaurant, type office, etc.).
Contextual	$context(id, context, value)$	Given a trajectory id and a given context $context$ and a value $value$ for this context, the predicate returns true if this trajectory is having the specific $value$ for the specific $context$.
Temporal	$time(p(e), condition)$	Given a temporal condition $condition$ (e.g., before time, after time, duration, between start and end time (time interval)) and given a spatial or contextual predicate $p(e)$, this predicate returns true while the temporal condition $condition$ of $p(e)$ is satisfied. Such condition is defined according to Allen's interval algebra [3].
	$duration(p, window, period)$	Given a temporal window, a time $period$ and any predicate p , this predicate returns true if the overall duration of $p(e)$ occurrences for the recent time $window$ is equal to $period$.
	$timeOverlap(p(e_1), p(e_2), period)$	Given a temporal $period$ and two given spatial or contextual predicates ($p(e_1)$ and $p(e_2)$) this predicate returns true if the overlapped time between $p(e_1)$ and $p(e_2)$ is equal or greater than $period$.

TABLE 4.1: Basic events predicates and descriptions.

Example 12 Let us consider the scenario of Example 11. From the streamed trajectory of $id1$, basic events can be extracted from the segment at spatial and contextual-based levels. From segment $a_1 = ([t_{75}, t_{76}], [space : place_1, NO_2 : High])$, we can extract the following basic events predicates:

- at spatial level: $p(e_1) = position(id1, place_1)$
- at contextual level: $p(e_2) = context(id1, NO_2, High)$

where e_1 and e_2 are basic events and $p(e_1)$ and $p(e_2)$ are their predicates, respectively, as illustrated in Table 4.1. This shows the flexibility of the approach where basic events can be defined either explicitly or using specific event predicates. Moreover, basic events can be manipulated at the temporal level and compared using temporal operations such as:

- $p(e_3) = time(p(e_1), during[t_{75}, t_{76}])$

- $p(e_4) = \text{time}(p(e_2), \text{precedes } t_{77})$ for $t_{77} > t_{76}$.

where e_3 and e_4 are temporal basic events and $p(e_3)$ and $p(e_4)$ are their predicates, respectively.

From segments a_1 and $a_2 = ([t_{79}, t_{80}], [\text{space} : \text{place}_2, \text{NO}_2 : \text{High}])$, we can extract the following temporal event predicate $p(e_5) = \text{duration}(p(e_2), 1 \text{ hour}, 15 \text{ minutes})$ where 1 *hour* is the window and 15 *minutes* is the period.

Composite (or complex) events are a set of patterns of interest that define relations among input derived events that usually impose temporal constraints on its sub-events. Composite events embed basic events to more complex situations that present an interest from an application point of view and should help to identify complex event patterns. Composite events are appropriate mechanisms to search for regular or irregular series of events identified using appropriate composition operators. When considering trajectory segments, composite events should help to extract complex patterns that emerge from streaming data. In short, a composite event is aggregation of basic events associated by composition mechanisms.

4.3.2 Composite Event

Definition 19 (Composite event). A composite event is defined as $CE = \bigcup_{i=1}^k E_i$ where E_i are basic or composite events. E_i are temporally constrained by $C(E_i, E_{i+1}) \forall E_i$ for $i \in [1, k-1]$ where C is a temporal constraint function defined by the event language (cf. Section 4.4).

A composite event is defined by a composition of basic events using some predefined operators according to the timing and logical relationships between the basic events. A composite event represent a continuous query over the streaming trajectories segments that meet mobility behaviours of interests. Let us remark that there is no assumption on the user *id* for the E_i . Composite events may combine basic events issued from one or more mobile users. Composite events are exemplified in Examples 13 and 14.

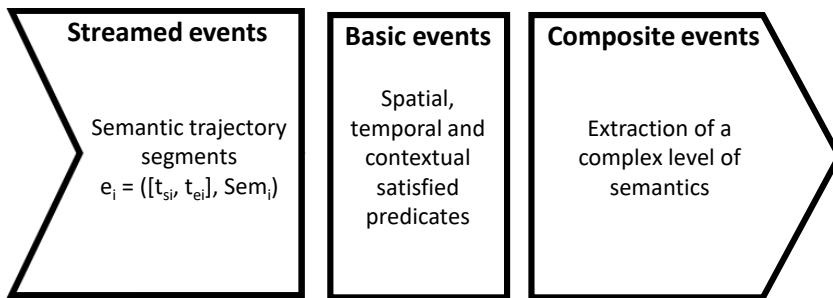


FIGURE 4.2: Composite events are defined from basic events and themselves from semantic trajectories.

Overall, this approach introduces the formal constructs that scale from the streaming events of semantic trajectory segments to the extraction of complex levels of semantics defined by the *composite events* passing through some primitive information denoted by the *basic events* as shown in Figure 4.2.

Composite events are defined from basic events and themselves from semantic trajectories. They must be defined in a formal language.

4.4 Event Pattern Language for Composite Event Recognition

A knowledge representation of composite event patterns requires a formal language to understand and describe their semantics. Composite events can be constructed using sequences, repetitions, conjunctions, disjunctions and negations of other events. As our objective is not to develop a new event-based language, and without loss of generality, we based our modelling approach on the Event Pattern Language (EPL) [38] [63] that provides logical constructs for complex events representation. It supports composite events consisting of an occurrence of a set of basic events that satisfies a pattern modelled as a sequence, repetition, conjunction, disjunction, negation and more derived constructs like relaxed sequence and iteration. The EPL expressive power has been already applied for modelling complex spatio-temporal phenomena [21]. The EPL is a high-level language, comprehensible and intuitive to manipulate and express complex events. Developing basic events for the EPL language is flexible and not limited to a specific domain. The simplicity and expressiveness of EPL give it its strength, with the ability to apply events composition from basic events extracted from predicates with a language that applies a wide range of operators.

Let us consider the basic operations of the Event Pattern Language:

- P is a basic or a composite event predicate (cf. Tables 4.2 and 4.3).
- $\neg P$ is the absence of P .
- $*$: P is a sequence of zero or more consecutive occurrences of P .
- (P_1, P_2, \dots, P_n) is a sequence of events consisting of an occurrence of P_1 , immediately followed by an occurrence of P_2 , ..., immediately followed by an occurrence of P_n .
- $(P_1 \& P_2 \& \dots \& P_n)$ is a conjunction of events occurring simultaneously (at a same instant or period).
- $\{P_1, P_2, \dots, P_n\}$ is a disjunction of events. It happens when at least one event among P_1, P_2, \dots, P_n occurs.
- $[P_1, P_2, \dots, P_n]$ is a relaxed sequence consisting of an occurrence of P_1 , followed later by an occurrence of P_2 , ..., followed by an occurrence of P_n . It is logically equivalent to $(P_1, * : any, P_2, \dots, * : any, P_n)$, where $* : any$ means a sequence of zero or more other events.

In the following, we discuss two types of composite events, called the individual events and the aggregated events.

4.4.1 Individual Events

The individual events are the events where only one moving entity is involved.

Definition 20 (Individual Events). An individual event is a composite event defined by $IE = \bigcup_{i=1}^k E_i$ where E_i are basic or composite events such as $\forall id_l, id_j \in E_i, id_l = id_j$ for $l, j \in [1, k]$. id_m is a mobile user identification for $m \in [1, k]$.

We define and categorise a list of individual events examples in Table 4.2. These events are represented according to the presented EPL language. For example, IE_2 defines an individual event for a trajectory having two contexts that occur at the same time for a minimum duration. The EPL representation of a IE_2 event searches for a time interval overlap of two basic events detected for the same trajectory using the conjunction operation. Example 13 presents a relaxed sequence of an individual event example.

Example 13 Let us consider the example IE_3 of Table 4.2. The pattern that defines this example can be represented by $contextBetweenPlacesSequence(id1, home_1, office_2, NO_2, High)$ and its EPL formalisation is modelled by the following relaxed sequence $[position(id1, home_1), duration((context(id1, NO_2, High) \& \neg position(id1, home_1)), 1hour, 15minutes), position(id1, office_1)]$ and supposing that the duration of $[t_{75}, t_{76}]$ is greater than 15 minutes.

4.4.2 Aggregated Events

Urban domain involves a huge number of human trajectories where composite events can emerge from relations and patterns of interest between different trajectories. Aggregated events are events where multiple moving entities are involved.

Definition 21 (Aggregated Events). An aggregated event is a composite event defined by $AE = \bigcup_{i=1}^k E_i$ where E_i are basic or composite events such as $\exists id_l \neq id_j | id_l \in E_l$ and $id_j \in E_j$ for $l, j \in [1, k]$.

We define and categorise a list of aggregated events examples in Table 4.3. The events are represented according to the presented EPL language.

For example, AE_3 defines an aggregated event for trajectories presented at different places but having the same context (e.g., same sports activity). The EPL representation of a AE_3 event example searches for a conjunction of $position$ basic events of different $places$ and same $context$ basic events for each trajectory. Example 14 presents a relaxed sequence of an aggregated event example.

Example 14 Let us consider another person having identification $id2$ and having the following semantic trajectory: $semTraj_{id2} = \{\dots, ([t_{67}, t_{68}], [space : place_1, NO_2 : High]), \dots\}$. In this example, we consider the aggregated event example AE_2 of Table 4.3. The pattern that defines this example can be represented by $meetBetweenPlacesSequence(id1, id2, home_1, office_2)$ and its EPL formalisation can be modelled by $[position(id1, home_1), meet(\{id1, id2\}, place_1), position(id1, office_1)]$.

Individual trajectory events				
#	Definition	Predicate	Query Example	Example in EPL
IE_1	Trajectory stop. A stop is defined when a trajectory stays at a place for a given duration (time). The place belongs to a given hierarchy.	$stop(id, place, duration)$ $stop(id, layer, duration)$	Find trajectories that stay at a place for a minimum time interval (e.g., station, office, town, road)	$time(position(id, Coffee\ shop), 2hours)$ where $Coffee\ shop$ is a place. $time(position(id, Versailles), 30minutes)$ where $Versailles$ is a town identified in the hierarchy.
IE_2	A trajectory that meets a conjunction of criteria for a given time. The occurrence may have a minimum duration.	$twoContextOccurrence(id, context_1, value_1, context_2, value_2, intervalOverlap)$	Find the trajectories that have a sports activity with a high level of pollution exposure.	$(context(id, Activity, Run) \& context(id, NO_2, High))$ where Run and $High$ are contextual values for $Activity$ and NO_2 criteria, respectively.
IE_3	A trajectory having a sequence of two places for a minimum duration and a context occurrence while moving through the places sequence.	$contextBetweenPlacesSequence(id, place_1, place_2, context, value)$	Find trajectories going from home to office and getting a high level of pollution exposition for 15 minutes.	$[position(id, home_1), duration(context(id, NO_2, High) \& \neg position(id, home_1)), 1hour, 15minutes), position(id, office)]$ where $home_1$ is a place and $office$ is a place property.
IE_4	End a trajectory stop at a specific place for a specific time.	$leavePlaceWithSpecificTime(id, place_1, condition)$	Find the trajectories leaving the office before 17h.	$(position(id, office), time(\neg position(id, office), precedes 17h))$ where $office$ is a place property.

TABLE 4.2: Examples of individual trajectory events.

Aggregated trajectories events				
#	Definition	Predicate	Query Example	Example in EPL
AE_1	Trajectories that stop at a same place and for a minimum duration.	$meet(\{id_i\}, place)$ $meet(\{id_i\}, layer)$	Find the trajectories/persons that meet at the same place for a minimum time interval (the place can belong to different hierarchies)	$(stop(id_1, Coffee), stop(id_2, shop, 30minutes) \& stop(id_2, Coffee shop, 30 minutes) \& \dots \& stop(id_n, Coffee shop, 30 minutes))$ where id_1, id_2, \dots, id_n are different mobilities ids and $Coffee shop$ is a place. Same to $layer$.
AE_2	Trajectory having a sequence of two similar categories of places while meeting over a given place for a minimum duration	$meetBetweenPlacesSequence(id_1, id_2, place_1, place_2)$	Find the trajectories going from home to office and meet other trajectories/persons at a $place_1$ for a minimum time interval where $place_1$ is a place.	$[position(id_1, home_1), meet(\{id_1, id_2\}, place_1), position(id_1, office)]$ where id_1, id_2 are different mobilities ids, $home_1$ is a place and $office$ is a place property.
AE_3	Trajectories having specific contexts and present at specific places for a given time criteria.	$sameContextWithDifferentPlaces(\{id_i\}, context, value)$	Find the trajectories having sports activity simultaneously but with different regions (or places)	$[position(id_1, place_1) \& context(id_1, Activity, Sport) \& position(id_2, place_2), \& context(id_2, Activity, Sport) \& \dots \& position(id_n, place_n) \& context(id_n, Activity, Sport)]$ where id_1, id_2, \dots, id_n are different mobilities ids and $place_1, place_2, \dots, place_n$ are different places.
AE_4	Trajectory having a specific number of occurrences while meeting over a given period.	$minimumMeet(id, \{id_i\}, dur)$	Find the trajectories that meet a minimum of 10 people within 1 hour	$duration([meet(id_1 \cup \{id_2\}, place_1), meet(id_1 \cup \{id_3\}, place_2), \dots, meet(id_1 \cup \{id_{11}\}, place_{10}), 1 \text{ hour}, 1 \text{ hour})$ where $id_1, id_2, \dots, id_{11}$ are different mobilities ids and $place_1, place_2, \dots, place_{10}$ are different places.

TABLE 4.3: Examples of aggregated events.

4.5 System Architecture

The event language presented at the conceptual level must be implemented in a CEP system. To enable a scalable CEP system, we need to build a clustered architecture consisting of multiple processing units that can work in parallel. Each of these processing units has CPUs and threads that can run independently on different parts of a given task before merging the partial outputs of each unit to get a single result. Apache Flink is one of the modern streaming Big Data platforms that provide the requirements to scale-out such a system. Our streaming system has been implemented using Apache Kafka and Apache Flink clusters as shown in Figure 4.3.

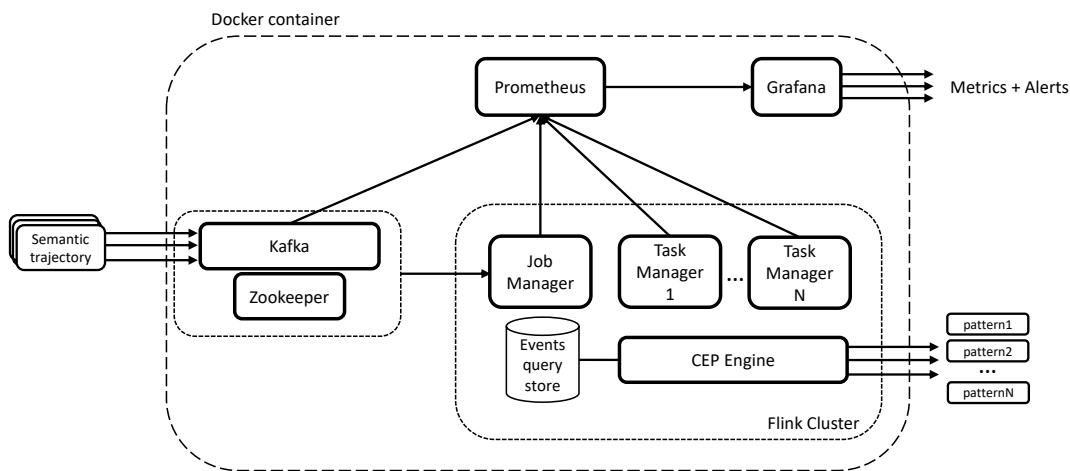


FIGURE 4.3: Streaming implementation architecture.

Apache Kafka is a publish-subscribe based fault-tolerant messaging system. It is used for building real-time data distributed pipelines channels (called topics).

Apache Flink is a distributed processing engine designed to run stateful streaming applications at any scale. The Flink Cluster is composed of at least one Master (Job Manager) that coordinates the distributed execution and a number of Worker nodes (Task Managers) responsible for the physical execution of tasks. Each Worker (JVM process) has a number of task slots (at least one) that are the Threads within the respective JVM process where each Flink operator or instance of an operator is assigned.

Flink provides a library for Complex Event Recognition (CER) and processing called FlinkCEP. FlinkCEP is an automata-based CER engine built on top of Flink and it offers a wide range of features providing substantial flexibility for defining patterns [40] since it employs non-deterministic automata equipped with predicates on their transitions, which matches relatively well the principles of our event modelling approach. FlinkCEP integrates CER operators into the Flink program and translates them into physical tasks like any other Flink operator. We have used FlinkCEP to develop the CEP engine that implements the individual and the aggregated events. The semantic trajectories are then streamed into a topic using Kafka and consumed at the Flink cluster level.

The language of FlinkCEP supports multiple contiguity conditions that are relatively related to the sequences defined by the EPL language. The (i) “*strict*” contiguity where matching events strictly appear one after the other without any non-matching events in-between is related to the sequence of events in EPL. The (ii) “*relaxed*” contiguity that ignores non-matching events appearing in-between the matching ones and the (iii) “*non-deterministic relaxe*” contiguity that allows non-deterministic actions between matching events are related to the relaxed sequence of EPL.

One other advantage of FlinkCEP is that patterns are compositional; for example, basic or composite events patterns can be defined once then used to define other events patterns (individual and aggregated events patterns). FlinkCEP accommodates a wide range of operators and selection policies.

In the stream processing fashion, a particular operator called window is usually utilised to confine the search space [26]. The event detection can quickly become unmanageable, especially when relaxed pattern selection are used. Windows are applied per pattern and their function is to restrict up to a certain point in the past the streamed events (a_i) to be considered. For a time-based window, the windowing function is used to restrict the time difference between the last and first events in a match. For a count-based window, the windowing function is used to restrict the number of events considered between the last and first events in a match. In FlinkCEP, windows are logical views, which means that streamed events are not buffered in batches, but they are directly forwarded to the operators of a pattern. All the streamed semantic trajectories segments of different mobile users are grouped in one logical window to detect an aggregated event (Figure 4.4b).

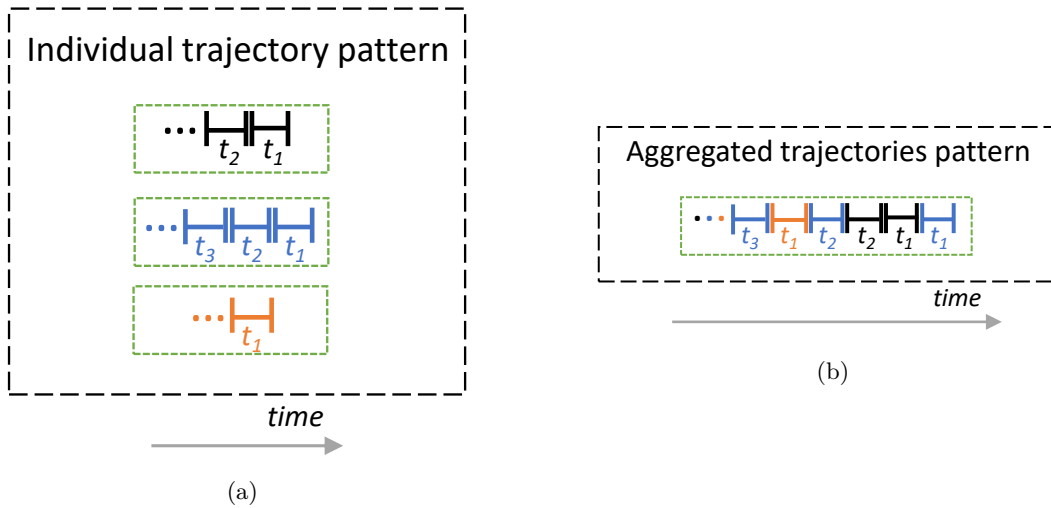


FIGURE 4.4: A visual illustration of the sliding windows manipulation on Individual (a) and Aggregated (b) events pattern stream. (a) shows how the window operation for detecting individual events are applied by grouping each moving entity’s segments in one sliding window. (b) shows how the window operations for detecting the aggregated events can be applied using one sliding window grouping segments from multiple moving entities.

Figure 4.4 shows the difference between an individual and an aggregated event on the sliding window manipulation level. In the case of an individual event (Figure 4.4a), each sliding window is applied on semantic trajectory segments of a unique stream of

semantic trajectory. Conversely, in the case of the aggregated event (Figure 4.4b), the sliding window is applied on semantic trajectory segments of multiple trajectories.

In order to monitor the overall system, we used Prometheus [85] and Grafana [43]. Prometheus records and saves metrics as time-series data, which means that metrics data is saved alongside the timestamp at which it was captured. Grafana provides charts, graphs and alerts about the system while running.

4.6 Complex Event Processing over Semantic Trajectories Experimentation

In this section, we present a list of composite events query examples. We discuss how they can be interpreted as automata. The performance of a set of individual and aggregated events is then studied.

4.6.1 Query Examples

The events query store illustrated in the Figure 4.3 contains the implementation of the individual and aggregated events examples presented in Tables 4.2 and 4.3. In this section, we present how some query examples similar to example 13 and 14 can be executed on our real data.

4.6.1.1 Query 1: Individual Event Example

The pattern of an individual event can be detected among a sliding window applied on the stream of each unique trajectory data as shown in Figure 4.4a.

An example of an individual event is to *find the trajectories going from “home” to “office” and get exposed to a high level of pollution for 15 minutes.*

The EPL formulation of *Query 1* (cf. Example 13) is denoted by:

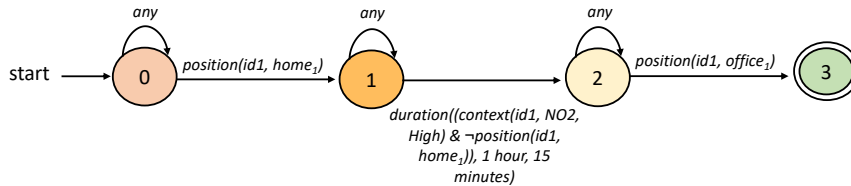
$$\begin{aligned} &contextBetweenPlacesSequence(id1, home_1, office_1, \\ &NO_2, High) : - \\ &[position(id1, home_1), \\ &duration((context(id1, NO_2, High) \\ &\&\sim position(id1, home_1)), 1hour, 15minutes), \\ &position(id1, office_1)] \end{aligned}$$


FIGURE 4.5: Example 3 automaton.

The EPL of *Query 1* is compiled to the non-deterministic automaton of Figure 4.5. This explains the transition between the states of the relaxed sequence. The

CER system, at first, detects the position of *id1* at a *home* place. Once the first state is detected, the system tries to detect the second base event denoted by state (2), by calculating the overall duration of the high level of NO_2 exposure during a 1-hour window, while taking into consideration that the position of *id1* is not at *home*₁ by applying a conjunction operation for $context(id1, NO_2, High)$ and $\neg position(id1, home_1)$ predicates. Once the duration is equal to or above 15 minutes, the system does the transition for the final state 3 to detect, which is a localisation at an *office* place. When the final state is detected, the overall *contextBetweenPlacesSequence* event pattern is detected.

Figure 4.6 illustrates the results of *Query 1* individual event pattern on a singular trajectory sample. The green dot represents the localisation at a *home* place, the red dots represent the places where the trajectory were exposed to a high level of NO_2 while going from home to office, and the yellow dot represents the localisation at an *office* place.

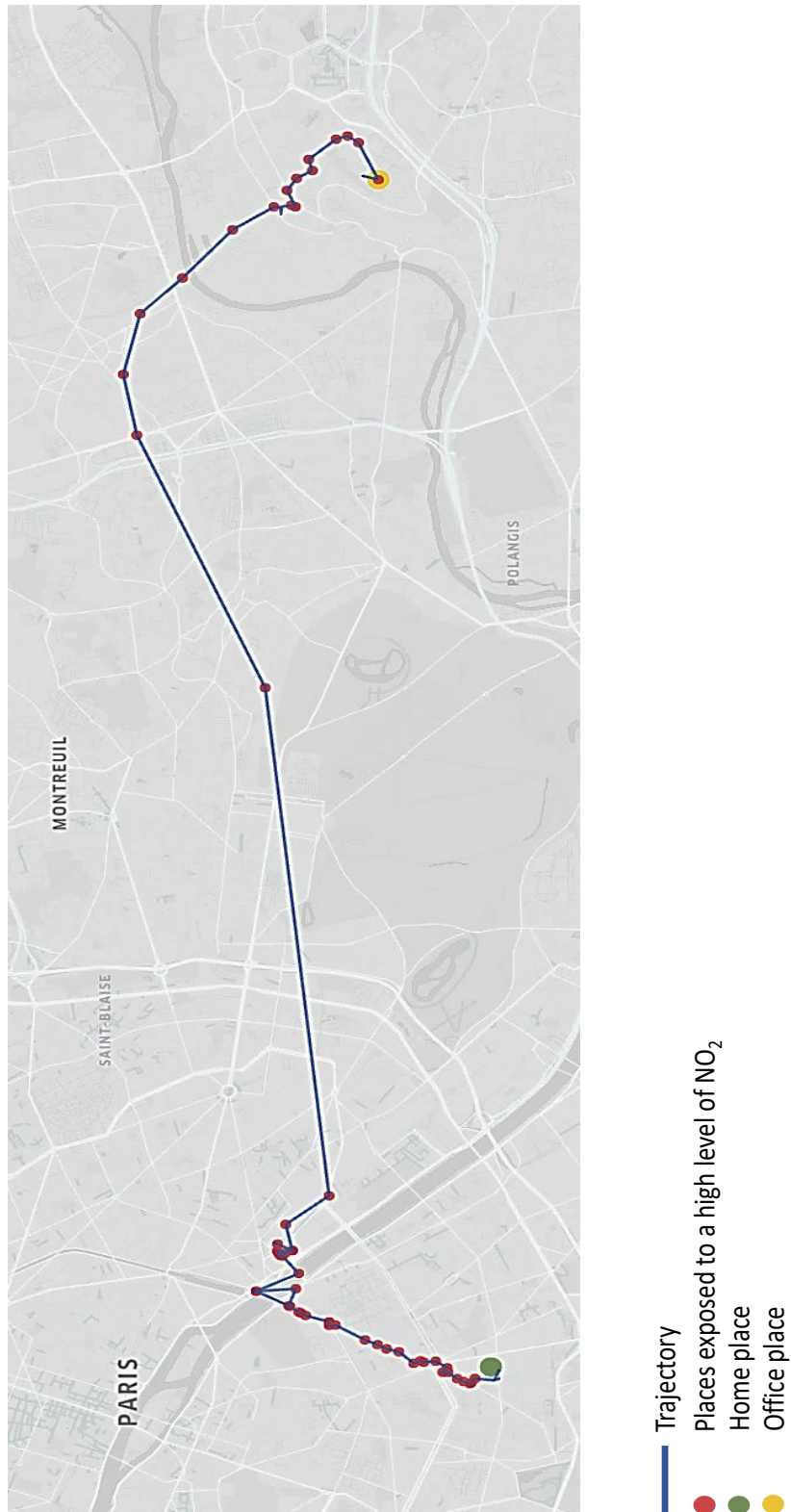


FIGURE 4.6: Query result for an individual trajectory event representing a trajectory going from home to office and get exposed to a high level of NO_2 for 15 minutes (*Query 1*).

4.6.1.2 Query 2: Aggregated Event Example

The pattern of an aggregated event can be detected among a sliding window applied on a stream of multiple trajectories data as shown in Figure 4.4b.

An example of an aggregated event is to *find the trajectories going from "home" to "office" and meet another trajectory/person at a $place_1$ for a minimum time interval where $place_1$ is a place different than home place and not having an office type.*

Let us consider *Query 2* example (cf. Example 14) and denoted by:

$$\begin{aligned} \text{meetBetweenPlacesSequence}(id1, id2, home_2, office_2) : - \\ [position(id1, home_2), meet(\{id1, id2\}, place_2), \\ position(id1, office_2)] \end{aligned}$$

where $id1$ and $id2$ are any mobile users such as $id1 \neq id2$.

Our model considers a composite event approach, and since events patterns in FlinkCEP are also compositional, this supports implementation at multiple steps for complex events if needed. The *meetBetweenPlacesSequence* event example, considers a *meet* event, hence, a *meet* event implementation is required. Therefore, a *meet* event considers a *stop* event, then, a *stop* event implementation is required.

Let us explain the used compositional event list for *meetBetweenPlacesSequence* event as follows.

The EPL formulation of a *stop* event is denoted by:

$$\begin{aligned} \text{stop}(id1, place_1, 30minutes) : - \\ \text{duration}(position(id1, place_1), 30minutes, 30minutes) \end{aligned}$$

The automaton of a *stop* event is illustrated in Figure 4.7. This denotes the detection of the stay of $id1$ at $place_1$ for at least 30 minutes. To define this event, the *duration* basic event identifies the stay of $id1$ at $place_1$ during a 30 minutes period for the recent 30 minutes window.

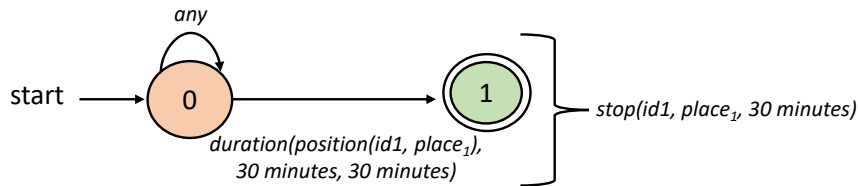


FIGURE 4.7: *stop* event automaton.

The second used compositional event is the *meet* event, its EPL formulation is denoted by:

$$\begin{aligned} \text{meet}(id1, id2, place_1) : - \\ \text{timeOverlap}(\text{stop}(id1, place_1, 30\text{minutes}), \\ \text{stop}(id2, place_1, 30\text{minutes}), 15\text{minutes}) \end{aligned}$$

The automaton of the *meet* event is illustrated in Figure 4.8. This denotes the detection of a joint *stop* event for *id1* and *id2* at *place₁* overlapped in time for a minimum of 15 minutes.

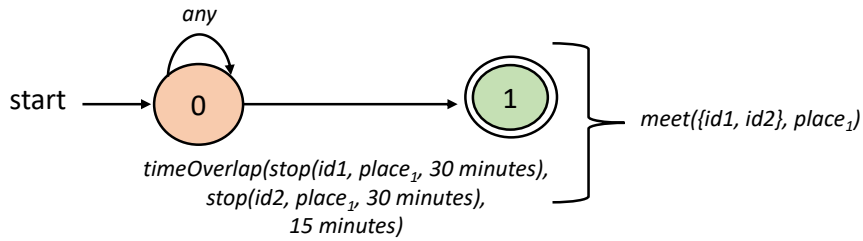


FIGURE 4.8: *meet* event automaton.

Finally, the *meetBetweenPlacesSequence* event automaton is illustrated in Figure 4.9. Figure 4.9 illustrates the transition between the states of a relaxed sequence. The CER system, at first, detects the position of *id1* at a *home* place. The second state to detect is the *meet* state explained above. Once the *meet* state is detected, the CER system detects final state 3 represented by the position of *id2* at an *office* place.

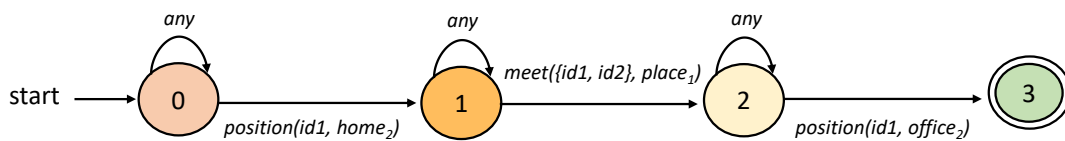


FIGURE 4.9: Example of *meetBetweenPlacesSequence* event automaton.

Figure 4.10 illustrates the results of *Query 2* denoted by *meetBetweenPlacesSequence* example of aggregated event pattern on two trajectories (cf. Example 14).

It is important to remark that the formulation of an event must be appropriately defined, especially for the numerical values, otherwise, the event will not be correctly detected.

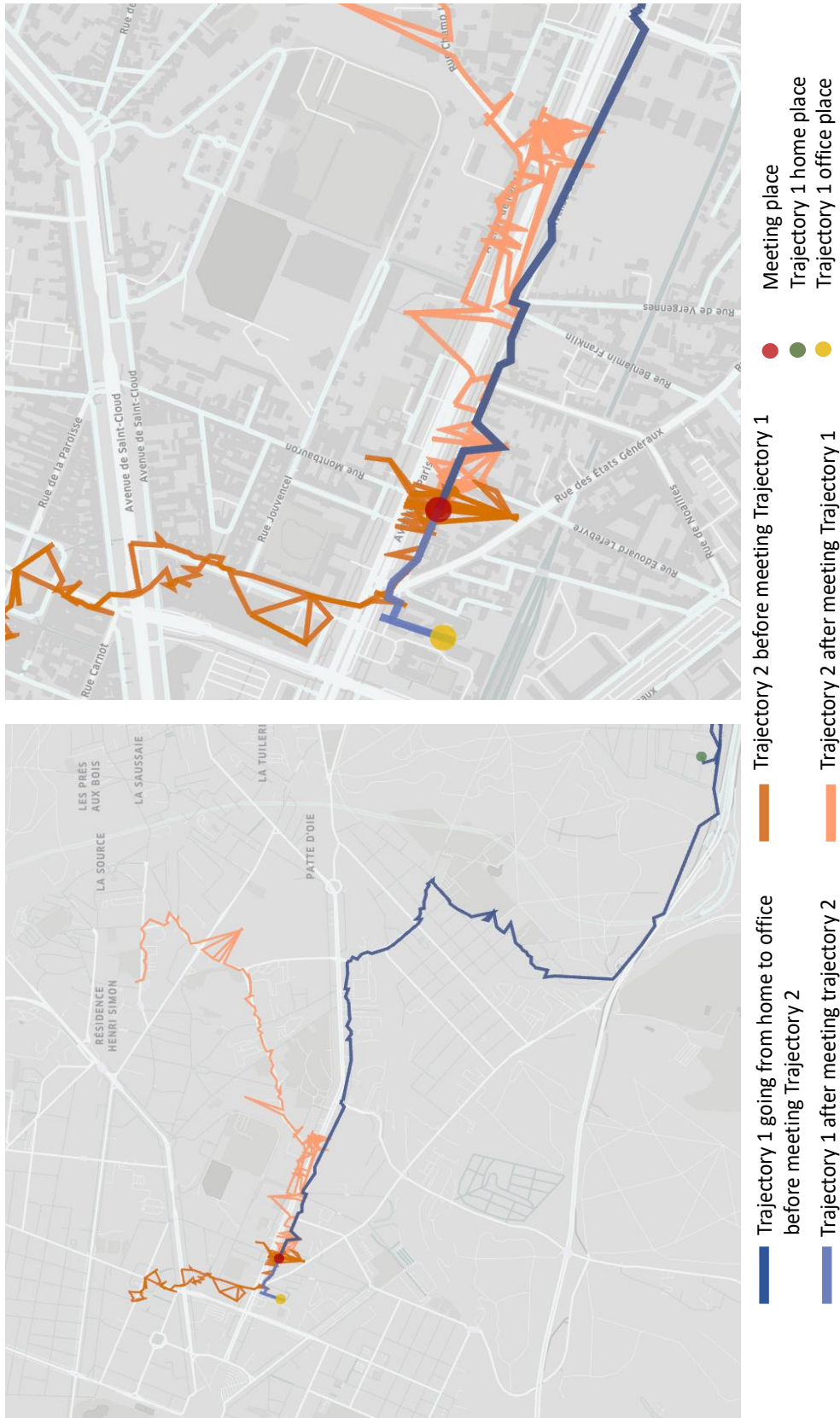


FIGURE 4.10: Query result for an aggregated event representing trajectory commuting from home to office and meet another trajectory at a *place* for a minimum duration (*Query 2*).

4.6.1.3 Query 3: Aggregated Event Example on a Spatial Hierarchy

One can constrain some events by a spatial hierarchy. Let us consider the following query example to *find the trajectories having sports activity at same times, but in different regions.*

Let us consider *Query 3* example (*sameContextWithDifferentPlaces* Table 4.3 AE_3) considering a conjunction of basic events occurring simultaneously and given a coarse spatial granularity constraint. The EPL formulation of *Query 3* is denoted by:

$$\text{sameContextWithDifferentPlaces}(id1, id2, Activity, Sport) : - \\ [position(id1, town_1) \& position(id2, town_2)]$$

where $id1$ and $id2$ are any mobile users such as $id1 \neq id2$.

Figure 4.11 shows the results of *Query 3* on the map. Montigny-le-Bretonneux is represented by $town_1$ and Buc is represented by $town_2$. We can also see a portion of each trajectory during the sports activity in the two different towns.

Overall, the series of individual and aggregated events tested in our CEP system demonstrate the expressive ability to represent and process composite events on top of Apache Flink with FlinkCEP. One peculiarity of the approach is that it provides contextual information support enabling extensive manipulation capabilities across multi-dimensional data associated with the mobility data. Another feature is the support of the spatial hierarchy for indoor and outdoor spaces, allowing the user to apply complex event queries homogeneously for both spaces.

Furthermore, the whole approach shows the potential of the EPL language to express a wide range of composite event operators to express a diverse collection of complex event queries.

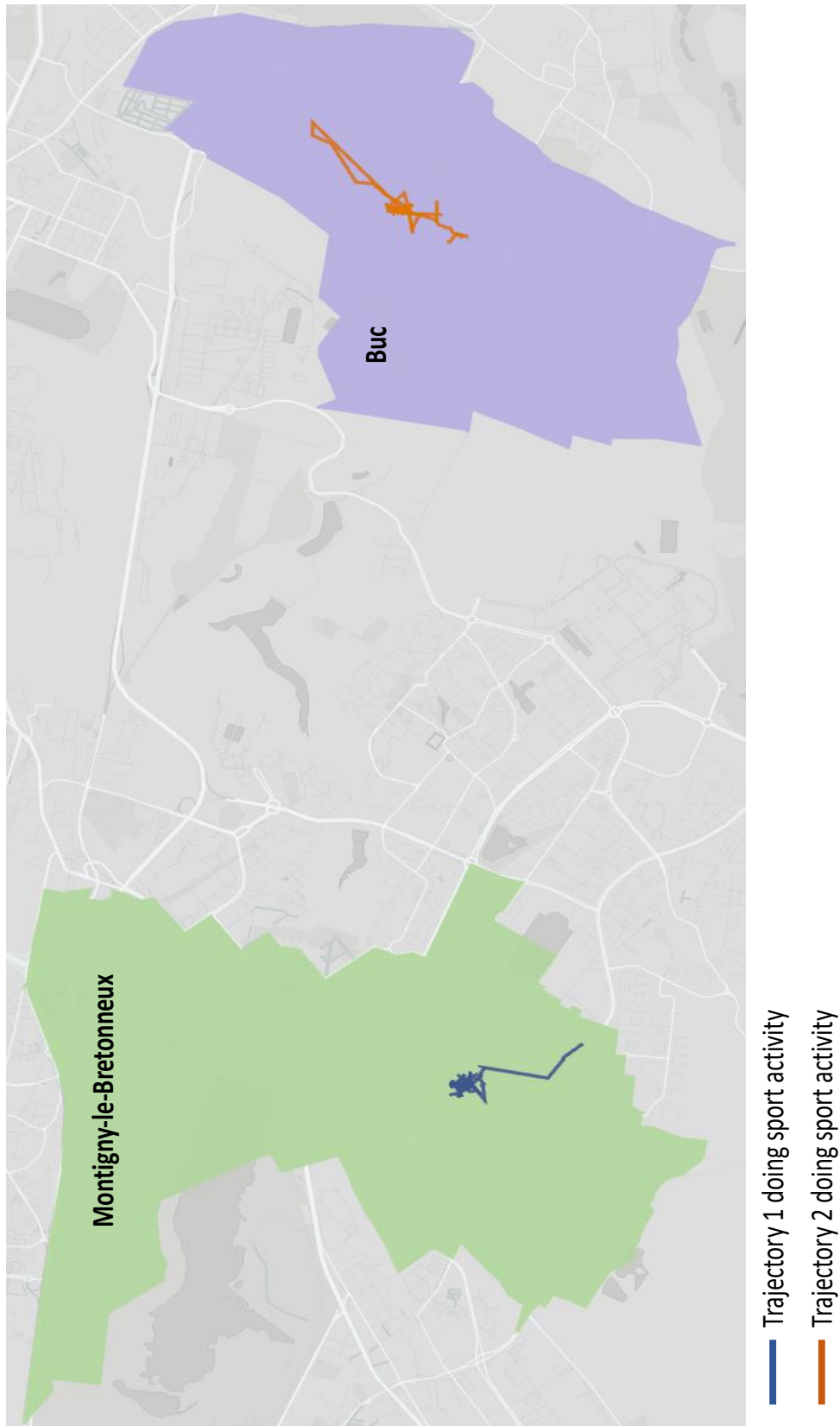


FIGURE 4.11: Query result for *sameContextWithDifferentPlaces* aggregated event that detects trajectories having sports activities in two different regions at the Town hierarchical level at the same time. The results show two trajectories doing sports activities, the first at “Montigny-le-Bretonneux” town and the second at “Buc” town (*Query 3*).

4.6.2 Performance

Our system considers a scalable approach that enables parallel processing. This approach has been experimented with real-time environmental crowd-sensing data and composite events collected in the region of Paris. Data have been described in section 3.5. We present multiple experiments that handle scalability by increasing the amount of working nodes and adding resources to the system. The objective of the parallel CEP system is to increase throughput and achieve low latency. Throughput is defined as the number of event tuples processed per time unit, while latency is defined as the processing time delay between reading from the input stream and writing to the output stream. The experiments are organised to exhibit throughput and latency performance metrics.

The experiments were conducted in Java using the open-source FlinkCEP provided by Flink 1.13.2 in a docker container running on a Windows 10 machine. The hardware configuration is as follows: The docker container resources are 50 CPUs and 86GB of memory. The whole server resources are a processor Intel Xeon Platinum 8180 @ 2.50GHz with 56 cores and 128GB of memory.

The input of our system consists of streams of semantic trajectories segments (*seg*) as defined in Definition 13. The overall semantic trajectories data contains approximately 660,000 records and is 412 MB in the form of CSV files. The semantic trajectory segments were streamed in an Apache Kafka topic of 6 partitions sorted by the date field. The topic was consumed by the Apache Flink cluster.

We ran the experiments on our CER system to detect each of the following individual complex event queries examples IE_1, IE_2, IE_3, IE_4 (the individual complex events examples are defined in Table 4.2) and aggregated complex event queries AE_1, AE_2, AE_3, AE_4 (the aggregated complex events examples are defined in Table 4.3).

Configurations	#JobManager	#TaskManager	#TaskSlots	#Threads
Configuration 1	1	1	1	1
Configuration 2	1	1	2	2
Configuration 3	1	2	2	4
Configuration 4	1	4	2	8
Configuration 5	1	4	4	16

TABLE 4.4: Apache flink different configuration settings.

We have repeated the experiments on five clusters with different configurations of Apache Flink that scale from 1, 2, 4, 8 to 16 threads (Table 4.4) where each TaskSlot is running on one logical processor. Each configuration has been experimented with multiple streaming rates of semantic trajectories segments (*seg*) of 500, 1000, 1500 and 2000 segments per second.

4.6.2.1 Throughput Evaluation

Figure 4.12 reports the system’s throughput for each complex event example obtained from increasing the value of the streaming load for each configuration. We observe that we can achieve a better throughput rate while increasing the number of running

threads. However, this is not the case for AE_1 throughput that drops with the *configuration 5* (16 threads) since this example consumes many memory resources and requires many shared resources to manage the conjunctions between different *stop* events.

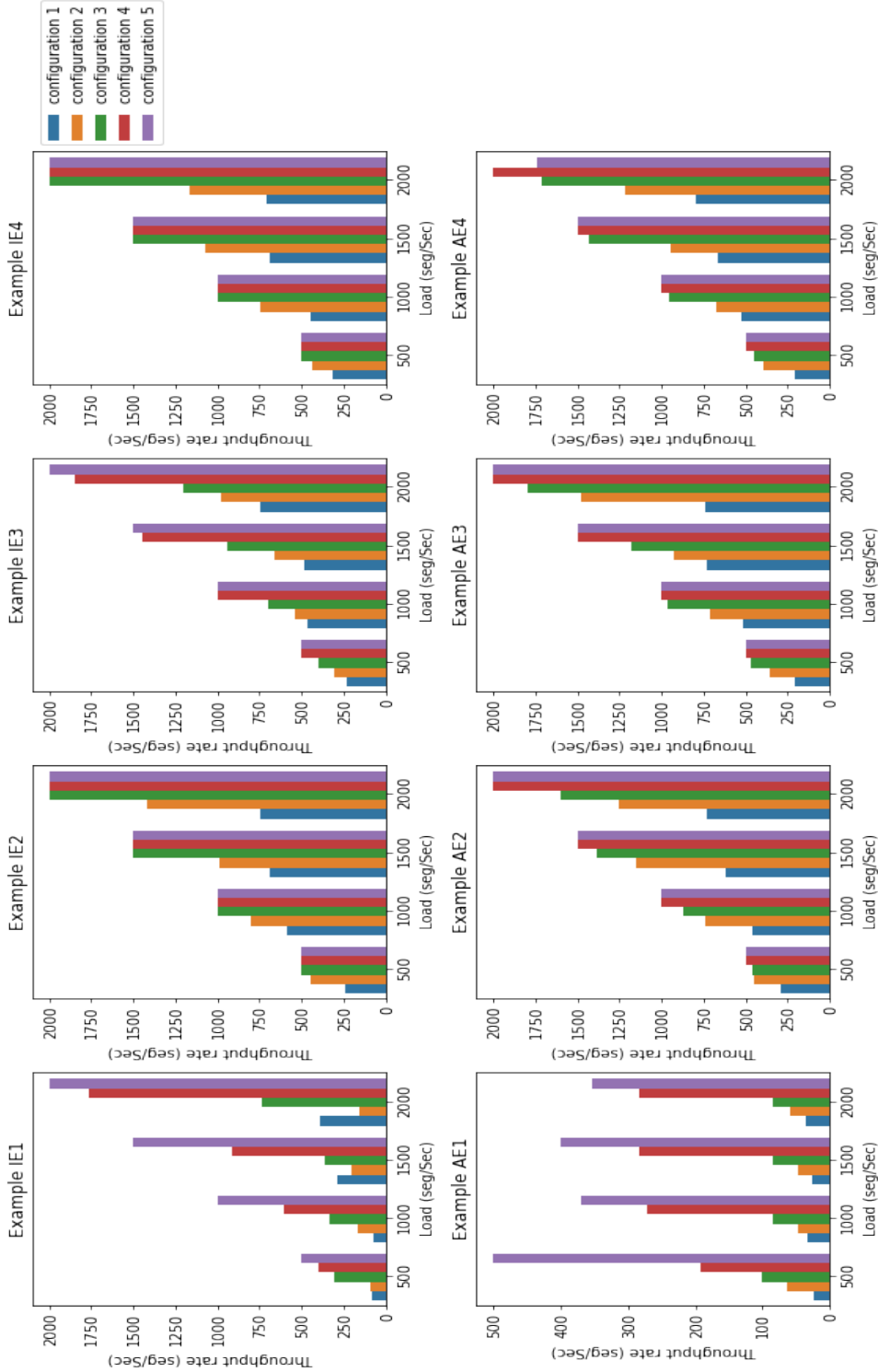


FIGURE 4.12: Individual and Aggregated complex events examples throughput with different configurations and streaming rates.

4.6.2.2 Latency Evaluation

We summarise the complex event examples queries latency in Figure 4.13. We could observe a significant change in the latencies that decrease by increasing the number of working threads.

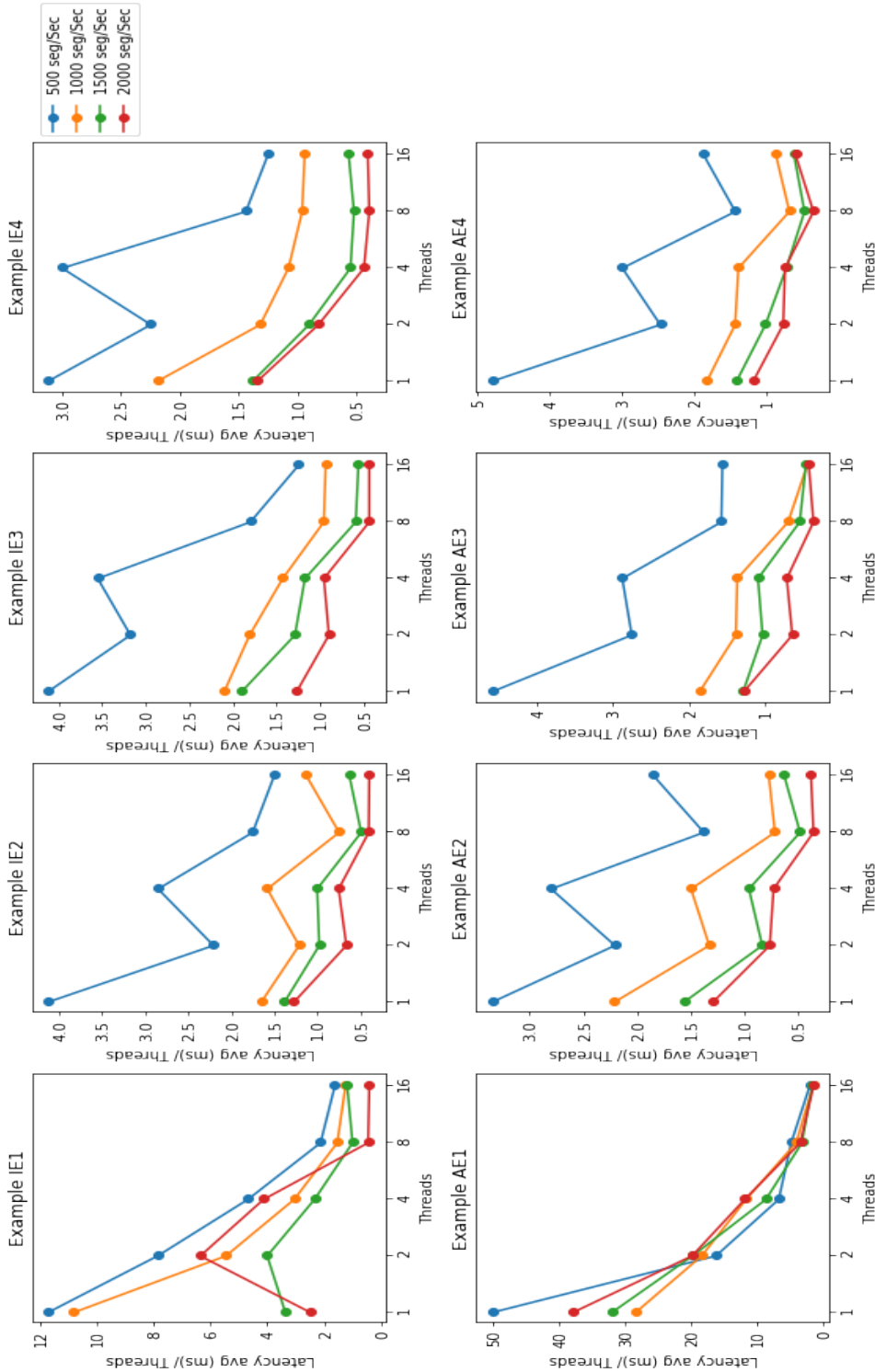


FIGURE 4.13: Individual and Aggregated complex events examples latency with different configurations and streaming rates.

Overall, the performance study shows that our system is capable of detecting individual and aggregated complex events with high throughput and low latency despite the fact that different types of sequences are applied (immediate, relaxed, conjunction and disjunction) while expressing spatial, contextual and temporal basic events predicates.

4.7 Conclusion

This chapter introduces a formal model and language for urban mobility complex event recognition that relies on the Event Pattern Language. We introduced a modelling framework whose peculiarity is to combine a hierarchical representation of semantic trajectories with a manipulation event-based language that provides a series of logical constructs to represent primitive and aggregated events. Semantic trajectory segments are identified and extracted from large data streams and manipulated by a series of predicates that support a wide range of events of interest. One peculiarity of this approach is that it considers indoor and outdoor semantic trajectories defined in a formal structure at a hierarchical level. The semantic trajectories are multi-dimensional and integrate multiple contextual information that arises with urban mobility. Such information enriches the querying process to reveal further interesting events.

The framework is evaluated by a developed prototype using the FlinkCEP engine support on top of Apache Flink. A series of real use cases of individual and aggregated events of interest have been evaluated using real crowd-sensed data collected in an urban environment.

The experimental evaluation demonstrates the capacity to express a wide range of complex events using the EPL model language. A predefined collection of basic events supports a nice composition of complex events. Deriving basic events among streams of semantic trajectories shows its efficiency and facilitates the process of modelling and implementing the query, as well as providing a simple approach to describe contextual and spatial semantics with hierarchical granularity. Finally, the experiments on our system show its scalable capacity to deliver high performance.

Because of the inherent nature of this approach, it is difficult to compare its performance to other existing approaches. However, such a system must be experimented with different data sets. Finally, there is also an interest in implementing the approach using other complex event processing systems for performance comparison.

Chapter 5

Conclusion

At the outset of this dissertation, we introduced the research context, the research statement, and the research objectives of the thesis. The main purpose of this thesis is to provide a modelling framework for analysing indoor and outdoor mobility patterns, offline and in real-time, to understand urban mobility. This chapter summarises the contributions of this thesis and sums up the limitations of our research. Next, we suggest extensions for this research. Finally, future research opportunities are presented through the contributions of this thesis.

5.1 Contributions

Research statement: By modelling human mobility with related crowd-sensed contextual information in indoor and outdoor spaces, we are able to provide the appropriate constructs to analyse mobility patterns and behaviours, offline or in real-time, in order to understand urban mobility data for a specific phenomenon and facilitate the analysis of semantic mobility for a large scale of applications.

The overarching goal of this thesis is to model, manage and process indoor and outdoor urban mobility in order to provide a better understanding of individual patterns and behaviour offline or in real-time. This overarching goal comprises two goals, as presented in chapter 1. In the first part of this thesis, we developed a spatial and trajectory representation in a graph structure and a processing evaluation to achieve the first goal. In order to achieve the second goal, we employed a formal language for complex event modelling and a streaming event-driven system in the second part of this thesis.

To this end, the research work developed in this thesis has led to a series of contributions that provide answers to our research challenges outlined in the introduction chapter:

Research Challenge 1

How to model indoor and outdoor spaces to support a homogeneous and continuous representation of human mobility?

In chapter 3, we first presented an indoor and outdoor spatial model to represent the spatial entities hierarchically and secure a link between both spaces. Indoor and outdoor spaces are homogeneously and formally modelled in order to represent the human movements emerging in these spaces. We define a multi-layered graph representation that supports structural properties while maintaining geometric properties. Our spatial representation can be materialised as place-based modelling where each place represents a space, either in indoor or outdoor, in order to define the complete spatial environment where human mobility takes place. Moreover, the spatial model allows the manipulation of human trajectories at different levels of abstractions; this further additional trajectory querying and analytical capabilities from a user's perspective, where queries may roll up and drill down over the presented spatial structure.

Research Challenge 2

How to design a flexible and dynamic multi-dimensional semantic trajectory based on real-time crowd-sourcing data and generic enough to cope with a large range of applications capabilities?

Chapter 3 presented a semantic trajectory model to represent urban mobility. In order to achieve the continuity of indoor and outdoor mobility, the semantic trajectory representation is spatially annotated by places embedded in our indoor and outdoor spatial model and at different granularities. Moreover, our model is formally defined

by a multi-layered graph as a sequence of annotated time intervals associated with multiple contextual dimensions. Considering a not-limited set of contextual domains in the annotation mechanism provides a good level of flexibility to semantically enrich the trajectory in order to cope with a wide range of application domains. The semantic trajectory model has been implemented in a data pipeline that integrates, enriches, and annotates the collected crowd-sourced data and then stores them in a graph database. The evaluation on queries and analytical level shows the expressive power of our representation where multiple dimensions and granularity are supported in advanced manipulation capabilities, where such data manipulations can operate on indoor, outdoor, or both spaces.

Research Challenge 3

How to extract from generic semantic trajectories multiple views at different levels of abstraction according to user-defined criteria?

Third, we have developed a multi-view extension for the semantic trajectory model in order to extract user-defined representations, namely, hybrid and contextual-based. These trajectory views can be derived based on various spatial, temporal and semantic criteria that allow for flexible data manipulation capabilities at different levels of granularity and user interests. This approach facilitates the derivation of different points of view and interpretations of trajectories from micro to macro levels. The multi-view extension is implemented into a graph database. The results illustrate how this approach supports the extraction from a generic representation, different views of the crowd-sourced trajectories that depend on user-defined semantic and spatial criteria in a well-defined graph structure.

Research Challenge 4

How to model complex events using a flexible and expressive language to be detected in a real-time big data processing system among streams of semantic trajectories data?

In chapter 4, we presented a formal event modelling and language approach for individual and aggregated complex event recognition. The approach relies on the Event Pattern Language (EPL) with a manipulation language that provides a series of logical constructs to represent basic and complex events among crowd-sourced location data and associated contextual information. The modelling framework combines our hierarchical representation of semantic trajectories where multiple contextual information associated with mobility are considered. The EPL provides an expressive definition of derived events that satisfy some complex patterns of interest. The overall framework implementation, based on FlinkCEP, is presented and evaluated. The evaluation using real use cases of individual and aggregated events shows the framework's potential to represent and detect a wide range of complex patterns upon indoor and outdoor mobility. Moreover, we studied the system's performance in terms of throughput and latency in function of the scalability of the distributed system. The results of this approach show how the Event Pattern language can support complex events representation with a good level of expressiveness using a wide range of operations.

Overall, the outcome of the thesis contributes to the scientific and technological flow of indoor and outdoor mobility data engineering and science, where necessary

requirements to analyse and understand indoor and outdoor mobility have been provided. Furthermore, the constructs developed in this thesis support the analysis of mobility data with related crowd-sensed information either in batch or streaming mode in order to explore mobility patterns and behaviours in indoor and outdoor environments.

5.2 Limitations

In chapter 3 and chapter 4 we have presented the limitations related to our work. We can synthesise this to the following six main limitations of the research that we have carried out:

- One of the limitations of the thesis is that it only applies the whole approach to a single context and dataset. Indeed, the work presented in chapter 3 and chapter 4 showed a high potential; however, the model needs to be more experimented on different datasets and contextual domains like CoViD-19 contact tracking domain that require analysing indoor and outdoor mobility.
- The second limitation, and also regarding the dataset, is that our indoor and outdoor mobility data contains real outdoor data and synthetic indoor mobility data. The whole approach needs to be experimented on not only real outdoor data but also to consider real data for both indoor and outdoor mobility.
- The multiple views approach presented in chapter 3 provides flexible data manipulation capabilities at different levels of granularity and user interests; however, there is still a need to experiment with additional query mechanisms and identify the range of possible data structures to improve computational times.
- We restricted our streaming approach presented in chapter 4 to the detection system only by simulating the semantic trajectories stream. This showed the potential of this approach; however, the streaming approach needs to be studied, including the pre-processing and semantic trajectory segmentation tasks in stream mode.
- The event modelling approach is implemented using Apache Flink and Flink-CEP. Our approach could be implemented using other streaming systems for benchmark comparison like RTEC [7].
- We limited the performance study of the event-based system presented in chapter 4 to the throughput and latency. This needs to be extended in order to evaluate the accuracy of the detected events.

5.3 Possible Research Extensions

In chapter 3 and chapter 4, we developed frameworks for analysing indoor and outdoor mobility data in offline and real-time, respectively. Our frameworks considered data heterogeneity that is intrinsic to crowdsourcing data where outdoor and indoor positioning techniques are very likely to rely on distinct principles (e.g., GPS and RFID, respectively). This results in different accuracy and data quality in the represented trajectories. However, these differences may be considered implicit in the various levels of abstraction and accuracy generally considered in indoor and outdoor environments. Our modelling technique has not taken into account data

quality and uncertainty issues so far. This might be part of further work, for example, by integrating uncertainty criteria at the data representation and query levels.

It would be useful to replicate the experiments performed in this thesis in other data contexts. For instance, it is important to evaluate the performance of the graph queries and analytics of chapter 3 and the Complex Events Processing of chapter 4 on different datasets and domain contexts. At the moment, our approaches have only been evaluated with Polluscope data (described in chapter 3).

Data quality and uncertainty issues have not been considered in the work presented in chapters 3 and 4. It will be advantageous to integrate the work on micro-environments detection [1, 31] with our work in order to enhance the data quality on precise micro-environments data.

A possible extension of the work is to study the difference between different processing systems. For instance, for chapter 3, it would be useful to evaluate our semantic trajectory approach using not only Neo4j [64] graph database but also using other relational ones like MobilityDB [123]. And for chapter 4, it would be worth experimenting the Complex Event Recognition (CER) model not only using FlinkCEP [5] but also using other CER tools like Wayeb [2] or RTEC [7]. Different experiments can be applied to compare with other systems' behaviour in terms of performance and expressiveness.

5.4 Future Work

The studies performed in the context of this thesis pave the way for several future work possibilities. We outline avenues of future research below:

Trajectories graph streaming: As future work, we propose developing a framework that generates semantic trajectories graph on the fly according to our semantic trajectory model, followed by a real-time processing engine to query trajectory patterns that emerge in a graph. Graph structures have proved their efficiency for querying complex data and achieving complex relationships between entities. On the other hand, in the recent past, Complex Event Processing (CEP) systems are capable of performing data querying on large-scale data in real-time. However, the former lack in querying structured graph data. We propose a future work towards investigating whether CEP systems can be extended to perform on dynamically graph connected data to apply parallel graph processing on trajectory data.

Trajectories natural query language: Trajectories modelling has been studied widely over the last decade; however, queries in natural languages are still not completely supported in trajectories databases. Developing a natural language for users who are not experts with structured query languages would be very important to ease the querying mechanisms for mobility arising either indoors or outdoor. Such language typically relies on a generic and well-defined conceptual model must be extensively developed in order to cover a wide range of queries.

Similarity measures: As a complement to the current research that focused mostly on the modelling aspects, it would be very useful to conduct data mining research to investigate similar behaviour across the mobility. Our semantic trajectory

model enriches the representation of the classical trajectories with contextual and spatial annotations with different levels of granularity. This allows the semantic comparison between trajectories to find frequent activities and detect atypical movements in order to understand human behaviour. In this context, a major challenge is the comparison of semantic trajectories graphs, looking to extract and learn similar human mobility behaviours. This requires the identification of similarity measures between semantic trajectories sequences where it should take advantage of the multiple contextual annotations and at different granularities.

Trajectories graph analysis: More research is needed to better understand the mobility behaviour by analysing the semantic trajectories graph using the graph-theoretical techniques (e.g., betweenness centrality) in order to monitor individual and collective behaviours and control some contextual strategies. For instance, weighting the semantic trajectories graph can provide additional information for studying an urban system, monitoring its evolution, and finding some correlations.

Bibliography

- [1] Mohammad Abboud, Hafsa El Hafyani, Jingwei Zuo, Karine Zeitouni, and Yéhia Taher. “Micro-environment recognition in the context of environmental crowdsensing”. In: *Workshops of the EDBT/ICDT Joint Conference, EDBT/ICDT-WS*. 2021.
- [2] Elias Alevizos, Alexander Artikis, and Georgios Paliouras. “Wayeb: a tool for complex event forecasting”. In: *arXiv preprint arXiv:1901.01826* (2018). DOI: [10.29007/2s9t](https://doi.org/10.29007/2s9t).
- [3] James F Allen. “Maintaining knowledge about temporal intervals”. In: *Communications of the ACM* 26.11 (1983), pp. 832–843.
- [4] Luis Otavio Alvares, Vania Bogorny, Bart Kuijpers, Jose Antonio Fernandes de Macedo, Bart Moelans, and Alejandro Vaisman. “A model for enriching trajectories with semantic geographical information”. In: *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. 2007, pp. 1–8. DOI: [10.1145/1341012.1341041](https://doi.org/10.1145/1341012.1341041).
- [5] *Apache Flink*. <https://nightlies.apache.org/flink/flink-docs-stable/>.
- [6] *Apache Flink*. <https://flink.apache.org/>.
- [7] Alexander Artikis, Marek Sergot, and Georgios Paliouras. “An event calculus for event recognition”. In: *IEEE Transactions on Knowledge and Data Engineering* 27.4 (2014), pp. 895–908. DOI: [10.1109/TKDE.2014.2356476](https://doi.org/10.1109/TKDE.2014.2356476).
- [8] Alexander Artikis, Anastasios Skarlatidis, François Portet, and Georgios Paliouras. “Logic-based event recognition”. In: *The Knowledge Engineering Review* 27.4 (2012), pp. 469–506. DOI: [10.1017/S0269888912000264](https://doi.org/10.1017/S0269888912000264).
- [9] Alexander Artikis, Matthias Weidlich, Francois Schnitzler, Ioannis Boutsis, Thomas Liebig, Nico Piatkowski, Christian Bockermann, Katharina Morik, Vana Kalogeraki, Jakub Marecek, et al. “Heterogeneous Stream Processing and Crowdsourcing for Urban Traffic Management”. In: *EDBT*. Vol. 14. 2014, pp. 712–723. DOI: [10.5441/002/edbt.2014.77](https://doi.org/10.5441/002/edbt.2014.77).
- [10] Frédéric Bisone. “Extraction de trajectoires sémantiques à partir de données multi-capteurs: application à des véhicules de secours”. PhD thesis. Université de tours, 2021.
- [11] Frédéric Bisone, Thomas Devogele, and Laurent Etienne. “From raw sensor data to semantic trajectories”. In: *Proceedings of the 5th ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management*. 2019, pp. 1–4. DOI: [10.1145/3356998.3365777](https://doi.org/10.1145/3356998.3365777).

- [12] Vania Bogorny, Chiara Renso, Artur Ribeiro de Aquino, Fernando de Lucca Siqueira, and Luis Otavio Alvares. “Constant—a conceptual data model for semantic trajectories of moving objects”. In: *Transactions in GIS* 18.1 (2014), pp. 66–88. DOI: [10.1111/tgis.12011](https://doi.org/10.1111/tgis.12011).
- [13] Matteo Böhm, Mirco Nanni, and Luca Pappalardo. “Improving vehicles’ emissions reduction policies by targeting gross polluters”. In: *arXiv preprint arXiv:2107.03282* (2021). DOI: [10.48550/arXiv.2107.03282](https://doi.org/10.48550/arXiv.2107.03282).
- [14] Ioannis Boutsis, Vana Kalogeraki, and Dimitrios Gunopulos. “Efficient event detection by exploiting crowds”. In: *Proceedings of the 7th ACM international conference on Distributed event-based systems*. 2013, pp. 123–134. DOI: [10.1145/2488222.2488264](https://doi.org/10.1145/2488222.2488264).
- [15] Mariem Brahem, Hafsa EL Hafyani, Souheir Mehanna, Karine Zeitouni, Laurent Yeh, Yehia Taher, Zoubida Kedad, Ahmad Ktaish, Mohamed Chachoua, and Cyril Ray. “Data perspective on environmental mobile crowd sensing”. In: *Intelligent Environmental Data Monitoring for Pollution Management*. Elsevier, 2021, pp. 269–288. DOI: [10.1016/B978-0-12-819671-7.00012-9](https://doi.org/10.1016/B978-0-12-819671-7.00012-9).
- [16] Jorge L Candiotti, Deepan C Kamaraj, Brandon Daveler, Cheng-Shiu Chung, Garrett G Grindle, Rosemarie Cooper, and Rory A Cooper. “Usability evaluation of a novel robotic power wheelchair for indoor and outdoor navigation”. In: *Archives of physical medicine and rehabilitation* 100.4 (2019), pp. 627–637. DOI: [10.1016/j.apmr.2018.07.432](https://doi.org/10.1016/j.apmr.2018.07.432).
- [17] Paris Carbone, Asterios Katsifodimos, Stephan Ewen, Volker Markl, Seif Haridi, and Kostas Tzoumas. “Apache flink: Stream and batch processing in a single engine”. In: *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 36.4 (2015).
- [18] Cécile Cayère, Christian Sallaberry, Cyril Faucher, Marie-Noëlle Bessagnet, Philippe Roose, Maxime Masson, and Jérémy Richard. “Multi-Level and Multiple Aspect Semantic Trajectory Model: Application to the Tourism Domain”. In: *ISPRS International Journal of Geo-Information* 10.9 (2021), p. 592. DOI: [10.3390/ijgi10090592](https://doi.org/10.3390/ijgi10090592).
- [19] Lu Chen, Yunjun Gao, Ziquan Fang, Xiaoye Miao, Christian S Jensen, and Chenjuan Guo. “Real-time distributed co-movement pattern detection on streaming trajectories”. In: *Proceedings of the VLDB Endowment* 12.10 (2019), pp. 1208–1220. DOI: [10.14778/3339490.3339502](https://doi.org/10.14778/3339490.3339502).
- [20] Christophe Claramunt, Cyril Ray, L Salmon, E Camossi, Melita Hadzagic, AL Jouselme, G Andrienko, N Andrienko, Y Theodoridis, and G Vouros. “Maritime data integration and analysis: recent progress and research challenges”. In: *Advances in Database Technology-EDBT 2017* (2017), pp. 192–197. DOI: [10.5441/002/edbt.2017.18](https://doi.org/10.5441/002/edbt.2017.18).
- [21] Christophe Claramunt and Marius Theriault. “Toward semantics for modelling spatio-temporal processes within GIS”. In: *Advances in GIS II*. 1996.
- [22] Alexis Richard C Claridades and Jiyeong Lee. “Defining a Model for Integrating Indoor and Outdoor Network Data to Support Seamless Navigation Applications”. In: *ISPRS International Journal of Geo-Information* 10.8 (2021), p. 565. DOI: [10.3390/ijgi10080565](https://doi.org/10.3390/ijgi10080565).

- [23] Constantinos Costa, Xiaoyu Ge, and Panos K Chrysanthis. “CAPRIO: graph-based integration of indoor and outdoor data for path discovery”. In: *Proceedings of the VLDB Endowment* 12.12 (2019), pp. 1878–1881. DOI: [10.14778/3352063.3352089](https://doi.org/10.14778/3352063.3352089).
- [24] Constantinos Costa, Xiaoyu Ge, Evan McEllhenney, Evan Kebler, Panos K Chrysanthis, and Demetrios Zeinalipour-Yazti. “CAPRIOv2. 0: A Context-Aware Unified Indoor-Outdoor Path Recommendation System”. In: *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2020, pp. 230–231. DOI: [10.1109/MDM48529.2020.00048](https://doi.org/10.1109/MDM48529.2020.00048).
- [25] Daniele Croce, Laura Giarre, Federica Pascucci, Ilenia Tinnirello, Giovanni Ettore Galioto, Domenico Garlisi, and Alice Lo Valvo. “An indoor and outdoor navigation system for visually impaired people”. In: *IEEE Access* 7 (2019), pp. 170406–170418. DOI: [10.1109/ACCESS.2019.2955046](https://doi.org/10.1109/ACCESS.2019.2955046).
- [26] Gianpaolo Cugola and Alessandro Margara. “Processing flows of information: From data stream to complex event processing”. In: *ACM Computing Surveys (CSUR)* 44.3 (2012), pp. 1–62. DOI: [10.1145/2187671.2187677](https://doi.org/10.1145/2187671.2187677).
- [27] *Cypher*. <https://opencypher.org/>.
- [28] Maria Luisa Damiani. “Semantic Trajectories Data Models”. In: *Handbook of Big Geospatial Data*. Ed. by Martin Werner and Yao-Yi Chiang. Cham: Springer International Publishing, 2021, pp. 185–197. ISBN: 978-3-030-55462-0. DOI: [10.1007/978-3-030-55462-0_8](https://doi.org/10.1007/978-3-030-55462-0_8).
- [29] Yanlei Diao, Neil Immerman, and Daniel Gyllstrom. “Sase+: An agile language for kleene closure over event streams”. In: *UMass Technical Report* (2007).
- [30] Christophe Dousson and Pierre Le Maigat. “Chronicle Recognition Improvement Using Temporal Focusing and Hierarchization”. In: *IJCAI*. Vol. 7. 2007, pp. 324–329.
- [31] Hafsa El Hafyani, Mohammad Abboud, Jingwei Zuo, Karine Zeitouni, and Yehia Taher. “Tell me what air you breath, i tell you where you are”. In: *17th International Symposium on Spatial and Temporal Databases*. 2021, pp. 161–165. DOI: [10.1145/3469830.3470914](https://doi.org/10.1145/3469830.3470914).
- [32] Maryam Maslek Elayam, Cyril Ray, and Christophe Claramunt. “Modèle de graphe pour l’analyse des structures de trajectoires maritimes”. In: *Extraction et Gestion des Connaissances: Actes EGC’2021* (2021).
- [33] *Esper*. <http://www.espertech.com/esper>.
- [34] Ziquan Fang, Yunjun Gao, Lu Pan, Lu Chen, Xiaoye Miao, and Christian S Jensen. “Coming: A real-time co-movement mining system for streaming trajectories”. In: *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 2020, pp. 2777–2780. DOI: [10.1145/3318464.3384703](https://doi.org/10.1145/3318464.3384703).
- [35] Carlos Andres Ferrero, Luis Otávio Alvares, and Vania Bogorny. “Multiple aspect trajectory data analysis: research challenges and opportunities.” In: *GeoInfo*. 2016, pp. 56–67.
- [36] Renato Fileto, Cleto May, Chiara Renso, Nikos Pelekis, Douglas Klein, and Yannis Theodoridis. “The Baquara2 knowledge-based framework for semantic enrichment and analysis of movement data”. In: *Data & Knowledge Engineering* 98 (2015), pp. 104–122. DOI: [10.1016/j.datak.2015.07.010](https://doi.org/10.1016/j.datak.2015.07.010).

- [37] Raghu K Ganti, Fan Ye, and Hui Lei. “Mobile crowdsensing: current state and future challenges”. In: *IEEE communications Magazine* 49.11 (2011), pp. 32–39. DOI: [10.1109/MCOM.2011.6069707](https://doi.org/10.1109/MCOM.2011.6069707).
- [38] Narain H Gehani, Hosagrahar V Jagadish, and Oded Shmueli. “Composite event specification in active databases: Model & implementation”. In: *VLDB*. Vol. 92. Citeseer. 1992, pp. 327–338.
- [39] Andrew Gemino and Yair Wand. “Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties”. In: *Data & Knowledge Engineering* 55.3 (2005), pp. 301–326. DOI: [10.1016/j.datak.2004.12.009](https://doi.org/10.1016/j.datak.2004.12.009).
- [40] Nikos Giatrakos, Elias Alevizos, Alexander Artikis, Antonios Deligiannakis, and Minos Garofalakis. “Complex event recognition in the big data era: a survey”. In: *The VLDB Journal* 29.1 (2020), pp. 313–352. DOI: [10.1007/s00778-019-00557-w](https://doi.org/10.1007/s00778-019-00557-w).
- [41] Nicholas A Giudice, Lisa A Walton, and Michael Worboys. “The informatics of indoor and outdoor space: a research agenda”. In: *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*. 2010, pp. 47–53. DOI: [10.1145/1865885.1865897](https://doi.org/10.1145/1865885.1865897).
- [42] Leticia I Gómez, Bart Kuijpers, and Alejandro A Vaisman. “Analytical queries on semantic trajectories using graph databases”. In: *Transactions in GIS* 23.5 (2019), pp. 1078–1101. DOI: [10.1111/tgis.12556](https://doi.org/10.1111/tgis.12556).
- [43] *Grafana*. <https://grafana.com/>.
- [44] Bin Guo, Zhu Wang, Zhiwen Yu, Yu Wang, Neil Y Yen, Runhe Huang, and Xingshe Zhou. “Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm”. In: *ACM computing surveys (CSUR)* 48.1 (2015), pp. 1–31. DOI: [10.1145/2794400](https://doi.org/10.1145/2794400).
- [45] Ralf Hartmut Güting, Victor Almeida, Dirk Ansorge, Thomas Behr, Zhiming Ding, Thomas Hose, Frank Hoffmann, Markus Spiekermann, and Ulrich Telle. “Secondo: An extensible dbms platform for research prototyping and teaching”. In: *21st International Conference on Data Engineering (ICDE’05)*. IEEE. 2005, pp. 1115–1116. DOI: [10.1109/ICDE.2005.129](https://doi.org/10.1109/ICDE.2005.129).
- [46] Hassan Nouredine. *Complex Event Processing from stream of semantic trajectories implementation, 2022*. <https://github.com/hassanouredine/semTrajEvents>.
- [47] Hassan Nouredine. *Implementation of multiple views extraction from indoor and outdoor semantic trajectories, 2021*. <https://github.com/hassanouredine/semantic-trajectory-neo4j-plugin>.
- [48] Hassan Nouredine. *Indoor and outdoor semantic trajectory implementation, 2020*. <https://github.com/hassanouredine/offline-semantic-trajectory>.
- [49] Sergio Ilarri, Dragan Stojanovic, and Cyril Ray. “Semantic management of moving objects: A vision towards smart mobility”. In: *Expert Systems with Applications* 42.3 (2015), pp. 1418–1435. DOI: [10.1016/j.eswa.2014.08.057](https://doi.org/10.1016/j.eswa.2014.08.057).
- [50] Søren Kejser Jensen, Jens Thomas Vejlbj Nielsen, Hua Lu, and Muhammad Aamir Cheema. “Outdoor-indoor space: Unified modeling and shortest path search”. In: *Proceedings of the Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*. 2016, pp. 35–42. DOI: [10.1145/3005422.3005427](https://doi.org/10.1145/3005422.3005427).

- [51] Meihan Jin and Christophe Claramunt. “A semantic model for human mobility in an urban region”. In: *Journal on Data Semantics* 7.3 (2018), pp. 171–187. DOI: [10.1007/s13740-018-0092-4](https://doi.org/10.1007/s13740-018-0092-4).
- [52] Alexandros Kontarinis, Karine Zeitouni, Claudia Marinica, Dan Vodislav, and Dimitris Kotzinos. “Towards a semantic indoor trajectory model: application to museum visits”. In: *GeoInformatica* (2021), pp. 1–42. DOI: [10.1007/s10707-020-00430-x](https://doi.org/10.1007/s10707-020-00430-x).
- [53] Ravi Kothuri and Siva Ravada. “Oracle spatial, geometries”. In: *Encyclopedia of GIS* (2008), pp. 821–826. DOI: [10.1007/978-0-387-35973-1_936](https://doi.org/10.1007/978-0-387-35973-1_936).
- [54] CH Kung and Arne Soelvig. “Activity modeling and behavior modeling”. In: *Proc. of the IFIP WG 8.1 working conference on Information systems design methodologies: improving the practice*. 1986, pp. 145–171.
- [55] Wissame Laddada and Cyril Ray. “Graph-based analysis of maritime patterns of life”. In: *Proceedings of the GAST Workshop, 20th Journées Francophones Extraction et Gestion des Connaissances (EGC)*. 2020, pp. 1–14.
- [56] Baptiste Languille, Valérie Gros, Nicolas Bonnaire, Clément Pommier, Cécile Honoré, Christophe Debert, Laurent Gauvin, Salim Srairi, Isabella Annesi-Maesano, Basile Chaix, et al. “A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science”. In: *Science of the Total Environment* 708 (2020), p. 134698. DOI: [10.1016/j.scitotenv.2019.134698](https://doi.org/10.1016/j.scitotenv.2019.134698).
- [57] Huan Li, Hua Lu, Muhammad Aamir Cheema, Lidan Shou, and Gang Chen. “Indoor mobility semantics annotation using coupled conditional Markov networks”. In: *2020 IEEE 36th international conference on data engineering (ICDE)*. IEEE. 2020, pp. 1441–1452. DOI: [10.1109/ICDE48307.2020.00128](https://doi.org/10.1109/ICDE48307.2020.00128).
- [58] Huan Li, Hua Lu, Xin Chen, Gang Chen, Ke Chen, and Lidan Shou. “Vita: A versatile toolkit for generating indoor mobility data for real-world buildings”. In: *Proceedings of the VLDB Endowment* 9.13 (2016), pp. 1453–1456. DOI: [10.14778/3007263.3007282](https://doi.org/10.14778/3007263.3007282).
- [59] Huadong Ma, Dong Zhao, and Peiyan Yuan. “Opportunities in mobile crowd sensing”. In: *IEEE Communications Magazine* 52.8 (2014), pp. 29–35. DOI: [10.1109/MCOM.2014.6871666](https://doi.org/10.1109/MCOM.2014.6871666).
- [60] Elio Masciari, Gao Shi, and Carlo Zaniolo. “Sequential pattern mining from trajectory data”. In: *Proceedings of the 17th International Database Engineering & Applications Symposium*. 2013, pp. 162–167. DOI: [10.1145/2513591.2513653](https://doi.org/10.1145/2513591.2513653).
- [61] Yuan Mei and Samuel Madden. “Zstream: a cost-based query processor for adaptively detecting composite events”. In: *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. 2009, pp. 193–206. DOI: [10.1145/1559845.1559867](https://doi.org/10.1145/1559845.1559867).
- [62] Ronaldo dos Santos Mello, Vania Bogorny, Luis Otavio Alvares, Luiz Henrique Zambom Santana, Carlos Andres Ferrero, Angelo Augusto Frozza, Geomar Andre Schreiner, and Chiara Renso. “MASTER: A multiple aspect view on trajectories”. In: *Transactions in GIS* 23.4 (2019), pp. 805–822. DOI: [10.1111/tgis.12526](https://doi.org/10.1111/tgis.12526).
- [63] Iakovos Motakis and Carlo Zaniolo. “Composite temporal events in active databases: a formal semantics”. In: *Recent Advances in Temporal Databases*. Springer, 1995, pp. 332–351. DOI: [10.1007/978-1-4471-3033-8_18](https://doi.org/10.1007/978-1-4471-3033-8_18).

- [64] *Neo4j*. <https://neo4j.com/>.
- [65] *Neo4j APOC Library*. <https://neo4j.com/developer/neo4j-apoc/>.
- [66] Hassan Nouredine. “Modelling, Managing and Processing of Mobility Data in Urban Environment”. In: *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2020, pp. 250–251. DOI: [10.1109/MDM48529.2020.00058](https://doi.org/10.1109/MDM48529.2020.00058).
- [67] Hassan Nouredine, Cyril Ray, and Christophe Claramunt. “A hierarchical indoor and outdoor model for semantic trajectories”. In: *Transactions in GIS* 26.1 (2021), pp. 214–235. DOI: [10.1111/tgis.12841](https://doi.org/10.1111/tgis.12841).
- [68] Hassan Nouredine, Cyril Ray, and Christophe Claramunt. “Multiple Views Extraction from Semantic Trajectories”. In: *International Symposium on Web and Wireless Geographical Information Systems*. Springer, 2022, pp. 1–17. DOI: [10.1007/978-3-031-06245-2_1](https://doi.org/10.1007/978-3-031-06245-2_1).
- [69] Hassan Nouredine, Cyril Ray, and Christophe Claramunt. “Multiple Views of Semantic Trajectories in Indoor and Outdoor Spaces”. In: *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. Association for Computing Machinery, 2021, pp. 99–102. DOI: [10.1145/3474717.3483915](https://doi.org/10.1145/3474717.3483915).
- [70] Hassan Nouredine, Cyril Ray, and Christophe Claramunt. “Semantic trajectory modelling in indoor and outdoor spaces”. In: *2020 21st IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2020, pp. 131–136. DOI: [10.1109/MDM48529.2020.00035](https://doi.org/10.1109/MDM48529.2020.00035).
- [71] Emmanouil Ntoulas, Elias Alevizos, Alexander Artikis, and Athanasios Koumparos. “Online trajectory analysis with scalable event recognition”. In: *EDBT/ICDT Workshops*. 2021.
- [72] *OGC Moving Features*. <https://www.ogc.org/standards/movingfeatures>.
- [73] *Open Geospatial Consortium (OGC)*. <https://www.opengeospatial.org/standards/>.
- [74] *OpenStreetMap*. <https://openstreetmap.org>.
- [75] *Oracle*. <https://www.oracle.com/>.
- [76] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady Andrienko, Natalia Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, Jose Macedo, Nikos Pelekis, et al. “Semantic trajectories modeling and analysis”. In: *ACM Computing Surveys (CSUR)* 45.4 (2013), pp. 1–32. DOI: [10.1145/2501654.2501656](https://doi.org/10.1145/2501654.2501656).
- [77] Kostas Patroumpas, Elias Alevizos, Alexander Artikis, Marios Vodas, Nikos Pelekis, and Yannis Theodoridis. “Online event recognition from moving vessel trajectories”. In: *GeoInformatica* 21.2 (2017), pp. 389–427. DOI: [10.1007/s10707-016-0266-x](https://doi.org/10.1007/s10707-016-0266-x).
- [78] Kostas Patroumpas, Nikos Pelekis, and Yannis Theodoridis. “On-the-fly mobility event detection over aircraft trajectories”. In: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2018, pp. 259–268. DOI: [10.1145/3274895.3274970](https://doi.org/10.1145/3274895.3274970).
- [79] Nikos Pelekis and Yannis Theodoridis. *Mobility data management and exploration*. Springer, 2014. DOI: [10.1007/978-1-4939-0392-4](https://doi.org/10.1007/978-1-4939-0392-4).

- [80] Nikos Pelekis, Yannis Theodoridis, and Davy Janssens. “On the management and analysis of our lifesteps”. In: *ACM SIGKDD Explorations Newsletter* 15.1 (2014), pp. 23–32. DOI: [10.1145/2594473.2594478](https://doi.org/10.1145/2594473.2594478).
- [81] Nikos Pelekis, Yannis Theodoridis, Spyros Vosinakis, and Themis Panayiotopoulos. “Hermes—a framework for location-based data management”. In: *International Conference on Extending Database Technology*. Springer, 2006, pp. 1130–1134. DOI: [10.1007/11687238_75](https://doi.org/10.1007/11687238_75).
- [82] Manolis Pitsikalis, Alexander Artikis, Richard Dreo, Cyril Ray, Elena Camossi, and Anne-Laure Jousselme. “Composite event recognition for maritime monitoring”. In: *Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems*. 2019, pp. 163–174. DOI: [10.1145/3328905.3329762](https://doi.org/10.1145/3328905.3329762).
- [83] *Polluscope*. <http://polluscope.uvsq.fr>.
- [84] *PostgreSQL*. <https://www.postgresql.org/>.
- [85] *Prometheus*. <https://prometheus.io/>.
- [86] Chiara Renso, Stefano Spaccapietra, and Esteban Zimányi. *Mobility data*. Cambridge University Press, 2013.
- [87] Milos B Roganovic and Dragan H Stojanovic. “Online detection of patterns in semantic trajectory data streams”. In: *2013 11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS)*. Vol. 2. IEEE, 2013, pp. 575–578. DOI: [10.1109/TELSIKS.2013.6704444](https://doi.org/10.1109/TELSIKS.2013.6704444).
- [88] Livia Ruback, Marco Antonio Casanova, Alessandra Raffaetà, Chiara Renso, and Vania Vidal. “Enriching mobility data with linked open data”. In: *Proceedings of the 20th International Database Engineering & Applications Symposium*. 2016, pp. 173–182. DOI: [10.1145/2938503.2938550](https://doi.org/10.1145/2938503.2938550).
- [89] Loic Salmon and Cyril Ray. “Design principles of a stream-based framework for mobility analysis”. In: *GeoInformatica* 21.2 (2017), pp. 237–261. DOI: [10.1007/s10707-016-0256-z](https://doi.org/10.1007/s10707-016-0256-z).
- [90] Korbinian Schmid, Teodor Tomic, Felix Ruess, Heiko Hirschmüller, and Michael Suppa. “Stereo vision based indoor/outdoor navigation for flying robots”. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 3955–3962. DOI: [10.1109/IROS.2013.6696922](https://doi.org/10.1109/IROS.2013.6696922).
- [91] Scarlet Schwiderski-Grosche and Ken Moody. “The SpaTeC composite event language for spatio-temporal reasoning in mobile systems”. In: *Proceedings of the 7th ACM international conference on Distributed event-based systems*. 2009. DOI: [10.1145/1619258.1619273](https://doi.org/10.1145/1619258.1619273).
- [92] Yuval Shahar. “A Framework for Knowledge-Based Temporal Abstraction”. In: *Artif. Intell.* 90.1–2 (Feb. 1997), 79–133. ISSN: 0004-3702. DOI: [10.1016/S0004-3702\(96\)00025-2](https://doi.org/10.1016/S0004-3702(96)00025-2).
- [93] *Siddhi CEP*. <https://github.com/siddhi-io/siddhi>.
- [94] Stefano Spaccapietra, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, and Christelle Vangenot. “A conceptual view on trajectories”. In: *Data & knowledge engineering* 65.1 (2008), pp. 126–146. DOI: [10.1016/j.datak.2007.10.008](https://doi.org/10.1016/j.datak.2007.10.008).
- [95] *Spark Streaming*. <https://spark.apache.org/streaming/>.

- [96] Dragan Stojanovic, Billur Barshan, Apostolos Papadopoulos, Nico Van de Weghe, and Christophe Claramunt. “Positioning Methods and Technologies in Mobile and Pervasive Computing”. In: *Encyclopedia of Information Science and Technology, Third Edition*. IGI Global, 2015, pp. 5713–5722. DOI: [10.4018/978-1-4666-5888-2.ch564](https://doi.org/10.4018/978-1-4666-5888-2.ch564).
- [97] Christian Strobl. “PostGIS”. In: (2008). DOI: [10.1007/978-0-387-35973-1_1012](https://doi.org/10.1007/978-0-387-35973-1_1012).
- [98] Rajesh Tamilmani and Emmanuel Stefanakis. “Modelling and Analysis of Semantically Enriched Simplified Trajectories Using Graph Databases”. In: *Advances in Cartography and GIScience of the ICA 1* (2019), NA–NA. DOI: [10.5194/ica-adv-1-20-2019](https://doi.org/10.5194/ica-adv-1-20-2019).
- [99] Hosna Tashakkori, Abbas Rajabifard, and Mohsen Kalantari. “A new 3D indoor/outdoor spatial model for indoor emergency response facilitation”. In: *Building and Environment* 89 (2015), pp. 170–182. DOI: [10.1016/j.buildenv.2015.02.036](https://doi.org/10.1016/j.buildenv.2015.02.036).
- [100] Tee-Ann Teo and Kuan-Hsun Cho. “BIM-oriented indoor network model for indoor and outdoor combined route planning”. In: *Advanced Engineering Informatics* 30.3 (2016), pp. 268–282. DOI: [10.1016/j.aei.2016.04.007](https://doi.org/10.1016/j.aei.2016.04.007).
- [101] Fernando Terroso-Saenz, Mercedes Valdés-Vela, Eric den Breejen, Patrick Hanckmann, Rob Dekker, and Antonio F Skarmeta-Gomez. “CEP-traj: An event-based solution to process trajectory data”. In: *Information Systems* 52 (2015), pp. 34–54. DOI: [10.1016/j.is.2015.03.005](https://doi.org/10.1016/j.is.2015.03.005).
- [102] *TimescaleDB*. <https://www.timescale.com/>.
- [103] Ankit Toshniwal, Siddarth Taneja, Amit Shukla, Karthik Ramasamy, Jignesh M Patel, Sanjeev Kulkarni, Jason Jackson, Krishna Gade, Maosong Fu, Jake Donham, et al. “Storm@ twitter”. In: *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 2014, pp. 147–156. DOI: [10.1145/2588555.2595641](https://doi.org/10.1145/2588555.2595641).
- [104] Andreas Tritsarolis, George-Stylios Theodoropoulos, and Yannis Theodoridis. “Online discovery of co-movement patterns in mobility data”. In: *International Journal of Geographical Information Science* 35.4 (2021), pp. 819–845. DOI: [10.1080/13658816.2020.1834562](https://doi.org/10.1080/13658816.2020.1834562).
- [105] Eftimis Tsilionis, Nikolaos Koutroumanis, Panagiotis Nikitopoulos, Christos Doukeridis, and Alexander Artikis. “Online event recognition from moving vehicles: Application paper”. In: *Theory and Practice of Logic Programming* 19.5-6 (2019), pp. 841–856. DOI: [10.1017/S147106841900022X](https://doi.org/10.1017/S147106841900022X).
- [106] Fabio Valdés, Maria Luisa Damiani, and Ralf Hartmut Güting. “Symbolic trajectories in secondo: Pattern matching and rewriting”. In: *International Conference on Database Systems for Advanced Applications*. Springer. 2013, pp. 450–453. DOI: [10.1007/978-3-642-37450-0_34](https://doi.org/10.1007/978-3-642-37450-0_34).
- [107] Ann Vanclooster and Philippe De Maeyer. “Combining Indoor and Outdoor Navigation: The Current Approach of Route Planners”. In: *Advances in Location-Based Services: 8th International Symposium on Location-Based Services, Vienna 2011*. Ed. by Georg Gartner and Felix Ortog. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 283–303. DOI: [10.1007/978-3-642-24198-7_19](https://doi.org/10.1007/978-3-642-24198-7_19).

- [108] Ann Vanclooster, Nico Van de Weghe, and Philippe De Maeyer. “Integrating indoor and outdoor spaces for pedestrian navigation guidance: A review”. In: *Transactions in GIS* 20.4 (2016), pp. 491–525. DOI: [10.1111/tgis.12178](https://doi.org/10.1111/tgis.12178).
- [109] Jarbas Nunes Vidal-Filho, Valéria Cesário Times, Jugurta Lisboa-Filho, and Chiara Renso. “Towards the Semantic Enrichment of Trajectories Using Spatial Data Infrastructures”. In: *ISPRS International Journal of Geo-Information* 10.12 (2021), p. 825. DOI: [10.3390/ijgi10120825](https://doi.org/10.3390/ijgi10120825).
- [110] Marcos R Vieira, Petko Bakalov, and Vassilis J Tsotras. “On-line discovery of flock patterns in spatio-temporal data”. In: *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*. 2009, pp. 286–295. DOI: [10.1145/1653771.1653812](https://doi.org/10.1145/1653771.1653812).
- [111] Sheng Wang, Zhifeng Bao, J Shane Culpepper, and Gao Cong. “A survey on trajectory data management, analytics, and learning”. In: *arXiv preprint arXiv:2003.11547* (2020). DOI: [10.1145/3440207](https://doi.org/10.1145/3440207).
- [112] Weiping Wang, Qiang Chang, Qun Li, Zesen Shi, and Wei Chen. “Indoor-outdoor detection using a smart phone sensor”. In: *Sensors* 16.10 (2016), p. 1563. DOI: [10.3390/s16101563](https://doi.org/10.3390/s16101563).
- [113] Michael Worboys. “Modeling indoor space”. In: *Proceedings of the 3rd ACM SIGSPATIAL international workshop on indoor spatial awareness*. 2011, pp. 1–6. DOI: [10.1145/2077357.2077358](https://doi.org/10.1145/2077357.2077358).
- [114] Eugene Wu, Yanlei Diao, and Shariq Rizvi. “High-performance complex event processing over streams”. In: *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. 2006, pp. 407–418. DOI: [10.1145/1142473.1142520](https://doi.org/10.1145/1142473.1142520).
- [115] Jianqiu Xu and Ralf Hartmut Güting. “A generic data model for moving objects”. In: *GeoInformatica* 17.1 (2013), pp. 125–172. DOI: [10.1007/s10707-012-0158-7](https://doi.org/10.1007/s10707-012-0158-7).
- [116] Jinjin Yan, Abdoulaye A Diakit , and Sisi Zlatanova. “A generic space definition framework to support seamless indoor/outdoor navigation systems”. In: *Transactions in GIS* 23.6 (2019), pp. 1273–1295. DOI: [10.1111/tgis.12574](https://doi.org/10.1111/tgis.12574).
- [117] Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer. “SeMiTri: a framework for semantic annotation of heterogeneous trajectories”. In: *Proceedings of the 14th international conference on extending database technology*. 2011, pp. 259–270. DOI: [10.1145/1951365.1951398](https://doi.org/10.1145/1951365.1951398).
- [118] Liping Yang and Michael Worboys. “A navigation ontology for outdoor-indoor space: (work-in-progress)”. In: *Proceedings of the 3rd ACM SIGSPATIAL international workshop on indoor spatial awareness*. 2011, pp. 31–34. DOI: [10.1145/2077357.2077364](https://doi.org/10.1145/2077357.2077364).
- [119] Liping Yang and Michael Worboys. “Similarities and differences between outdoor and indoor space from the perspective of navigation”. In: *Poster presented at COSIT* (2011).
- [120] Haopeng Zhang, Yanlei Diao, and Neil Immerman. “On complexity and optimization of expensive queries in complex event processing”. In: *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 2014, pp. 217–228. DOI: [10.1145/2588555.2593671](https://doi.org/10.1145/2588555.2593671).

-
- [121] Hengcai Zhang, F Lu, S Zhen, and Enchao Wu. “Unified navigation graph model of indoor space and outdoor space”. In: *Proceedings of the 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sapporo, Japan. 2017, pp. 18–21.
- [122] Pengfei Zhou, Yuanqing Zheng, Zhenjiang Li, Mo Li, and Guobin Shen. “Iodetector: A generic service for indoor outdoor detection”. In: *Proceedings of the 10th acm conference on embedded network sensor systems*. 2012, pp. 113–126. DOI: [10.1145/2659466](https://doi.org/10.1145/2659466).
- [123] Esteban Zimányi, Mahmoud Sakr, and Arthur Lesuisse. “MobilityDB: A mobility database based on PostgreSQL and PostGIS”. In: *ACM Transactions on Database Systems (TODS)* 45.4 (2020), pp. 1–42. DOI: [10.1145/3406534](https://doi.org/10.1145/3406534).
- [124] Sisi Zlatanova, Jinjin Yan, Yijing Wang, Abdoulaye Diakit , Umit Isikdag, George Sithole, and Jack Barton. “Spaces in spatial science and urban applications—state of the art review”. In: *ISPRS International Journal of Geo-Information* 9.1 (2020), p. 58. DOI: [10.3390/ijgi9010058](https://doi.org/10.3390/ijgi9010058).

Titre : Un modèle hiérarchique pour les trajectoires sémantiques et l'extraction d'événements dans les espaces intérieurs et extérieurs

Mot clés : Gestion des données de mobilité, Trajectoires sémantiques, événements complexes, modélisation d'espace

Résumé : L'intérêt pour l'exploitation des informations provenant de collecte participative a récemment émergé car elle peut apporter de nombreux avantages précieux pour de nombreux domaines d'application. C'est notamment le cas des données de crowd-sourcing mobiles en temps réel, souvent disponibles dans des environnements "indoor" et "outdoor". Ces données offrent de nombreuses possibilités d'analyse de la mobilité humaine, notamment lorsqu'elles sont associées à des informations contextuelles multidimensionnelles. La prise en compte de la diversité des espaces multi-environnements et des lieux de mobilité soulève plusieurs défis de recherche en modélisation, gestion et traitement des données. Associée à de multiples informations contextuelles, l'analyse de la mobilité "indoor" et "outdoor" souligne la nécessité d'abstractions de données appropriées et flexibles au niveau de la modélisation pour représenter les données spatiales, temporelles et sémantiques qui apparaissent dans un environnement de ville intelligente. Si les approches récentes ont souvent abordé cette question en utilisant le modèle commun des "stops and moves", celui-ci ne couvre pas complètement les informations contextuelles multidimensionnelles qui apparaissent en temps réel sur les humains naviguant dans les espaces "indoor" et "outdoor". Cela renforce également le besoin de systèmes permettant de traiter les données spatio-temporelles en temps réel lors de la recherche d'événements de mobilité complexe. Malgré leur capacité à représenter des événements spatio-temporels, ces systèmes nécessitent des langages de manipulation des données bien définis et flexibles pour

prendre en charge les mécanismes d'abstraction et de composition permettant d'analyser les mobilités urbaines. Cette thèse a pour objectif de fournir les constructions nécessaires à l'analyse des informations de collecte participative mobile dans les espaces "indoor" et "outdoor" tout en considérant un ensemble non limité d'informations contextuelles qui peuvent être associées afin de mieux comprendre les données de mobilité urbaine en batch et en temps réel. Nous présentons un modèle de données spatiales "indoor" et "outdoor" représenté sous la forme d'un graphe multicouche et annoté avec des données de trajectoire provenant de collecte participative. La nouveauté de cette approche réside dans le fait qu'elle fournit un modèle spatial "indoor" et "outdoor" hiérarchique, homogène et flexible, qui peut être associé à la volée à des données de trajectoire provenant d'un milieu urbain. Notre approche de modélisation définit des trajectoires sémantiques génériques et flexibles prenant en compte de multiples sémantiques de données collaboratives à différentes granularités et où la segmentation des trajectoires repose sur des valeurs sémantiques évolutives. Cette thèse développe un cadre de modélisation des événements complexes appliqué aux trajectoires sémantiques humaines "indoor" et "outdoor" basé sur un langage formel qui établit les opérations requises pour la composition des événements. Nous avons implémenté des pipelines de données pour examiner l'efficacité de notre approche. L'ensemble de l'approche est expérimenté et appliqué à des données participatives issues d'une étude de cas réelle afin de montrer sa pertinence, sa scalabilité et ses performances.

Title: A hierarchical model for semantic trajectories and event extraction in indoor and outdoor spaces

Keywords: Mobility data management, semantic trajectories, complex event recognition, indoor and outdoor modelling

Abstract: The interest in exploiting crowd-sourced information has recently emerged as it can bring many valuable benefits for many application domains. This is particularly the case for real-time mobile crowd-sourcing data often available in indoor and outdoor environments. Such data offers many opportunities for analysing human mobility, especially when associated with multi-dimensional contextual information. Considering the diversity of multi-environment spaces and where mobility occurs, raises several data modelling, management and processing research challenges. When associated with multiple contextual information, indoor and outdoor mobility analysis stresses the need for appropriate and flexible data abstractions at the modelling level to represent the spatial, temporal and semantic data that arise in a smart city environment. While recent approaches often considered this issue using the common stops and moves model, this does not completely cover the multi-dimensional contextual information that arises in real-time on humans navigating through indoor and outdoor spaces. It also increases the need for computing systems and data architectures to process spatio-temporal data in a timely manner when searching for complex mobility events of interest. Despite the ability to represent spatio-temporal events, such systems require well-defined and flexible data manipulation languages to support abstraction and composition mechanisms

for analysing urban mobilities.

This thesis aims to provide the necessary constructs for analysing mobile crowd-sensed information that arises in indoor and outdoor spaces. In order to better understand urban mobility data in batch and real-time, we consider a broad range of contextual information that can be associated with mobility data. We introduce an indoor and outdoor spatial data model represented as a multi-layered graph and constructed with crowd-sourced trajectory data. The novelty of the approach lies in the fact that it provides a homogeneous and flexible hierarchical indoor and outdoor spatial model that can be associated with crowd-sensed trajectory data on the fly. Our modelling approach defines generic and flexible semantic trajectories considering multiple collaborative data semantics at different granularities and where trajectory segmentation relies on evolving semantic values. This thesis develops a modelling framework for complex events applied to our indoor and outdoor semantic trajectory model based on a formal language that establishes the required operations for the composition of the events. We have implemented data pipelines to examine our approach's efficiency. The whole approach is experimented and applied to participatory data from a real case study to show its suitability, scalability and performance.