

# Analyse de trajectoires sémantiques dans un environnement fermé pour la compréhension des mouvements des visiteurs dans un musée (A-SITM)

## Encadrant.e.s

- Claudia MARINICA (co-encadrante 30%) [Claudia.Marinica@univ-nantes.fr](mailto:Claudia.Marinica@univ-nantes.fr)
- Fabien PICAROUGNE (co-encadrant 30%) [Fabien.Picarougne@univ-nantes.fr](mailto:Fabien.Picarougne@univ-nantes.fr)
- Fabrice GUILLET (directeur de thèse 40%) [Fabrice.Guillet@univ-nantes.fr](mailto:Fabrice.Guillet@univ-nantes.fr)

## Les partenaires

Ce sujet de thèse se place dans le cadre d'une collaboration internationale entre l'équipe de recherche DUKe du laboratoire LS2N, Nantes, France, et le Laboratoire de recherche en muséologie du Département des archives, bibliothéconomie et muséologie de l'Université Ionienne de Corfou, Grèce, en collaboration avec des musées de Corfou, Grèce, dont le Musée d'art asiatique, le Musée Mon Repos et la Galerie d'art. Les données des visiteurs déjà collectées et à collecter via des services de localisation ou les systèmes des institutions susmentionnées comprennent des ensembles de données à analyser.

Participants du LS2N - DUKe : *Claudia Marinica, Fabien Picarougne, Fabrice Guillet.*

Participant du Laboratoire de recherche en muséologie, Département des archives, bibliothéconomie et muséologie, Université Ionienne de Corfou, Grèce : *Georgios Papaioannou, Mat Damigos, Eleutherios Kalogeros.*

Participants des musées de Corfou : *personnel à nommer une fois le projet démarré.*

**Résumé.** De nombreux musées utilisent des services de localisation (LBS) pour suivre les mouvements des visiteurs (par exemple le Louvre, le British Museum, etc.) afin de mieux comprendre les visiteurs. En plus des informations de localisation, un système LBS peut fournir des informations contextuelles ou sémantiques sur les visiteurs. De plus, les musées peuvent fournir des données supplémentaires concernant la description topologique de leur espace intérieur. Ainsi, l'objectif principal de cette thèse est de développer de nouvelles techniques d'analyse du mouvement des visiteurs à travers des approches d'analyse de trajectoires, où les informations de positionnement dans les trajectoires sont enrichies avec des informations topologiques et sémantiques. À cette fin, nous proposons dans cette thèse de répondre à trois défis : (1) fournir une norme pour les trajectoires enrichies basée sur le Modèle de Trajectoires Sémantiques en Milieu Fermé (SITM) que nous avons proposé précédemment et la norme IndoorGML, (2) développer une analyse descriptive pour extraire des modèles à partir des trajectoires formalisées sous le modèle SITM, et (3) développer des algorithmes de classification de séries temporelles de trajectoires pour fournir des prédictions de mouvement.

## Description du sujet

### **1. Contexte et objectifs globaux**

Depuis des décennies, les musées étudient leurs visiteurs pour comprendre pourquoi ils vont dans les musées, ce qu'ils y font, comment ils apprennent et quels peuvent être leur engagement et leur satisfaction. En effet, outre leur rôle de médiation culturelle, les musées souhaitent enrichir l'expérience des visiteurs [1]. Plus précisément, une meilleure compréhension de leurs visiteurs pourrait permettre aux musées, dans un premier temps, de personnaliser leurs expériences, de promouvoir l'accessibilité et/ou de proposer des visites de musées dynamiques (comme présenté dans nos travaux précédents dans [2]). Deuxièmement, étudier les visiteurs pourrait encourager les musées à prendre des décisions de gestion telles que l'amélioration des voies d'évacuation [2].

De nos jours, les services basés sur la localisation (LBS) permettent de capturer assez facilement des informations sur les mouvements de personnes ou d'objets. De nombreux musées

les utilisent pour suivre les déplacements des visiteurs et leurs habitudes de visite : par exemple le Metropolitan Museum of Arts de New York, la Smithsonian Institution de Washington DC, le British Museum de Londres, le Louvre et le Rijksmuseum d'Amsterdam. Au Louvre, par exemple, des audioguides spéciaux (par exemple la Nintendo 3DS) et des applications pour smartphone (par exemple Louvre Ma Visite) sont utilisés. À mesure que cette technologie devient plus sophistiquée, nous pouvons nous attendre à voir encore plus de musées l'utiliser pour améliorer l'expérience des visiteurs. De plus, ces dispositifs peuvent fournir, avec les données de localisation, certaines informations contextuelles ou sémantiques sur le visiteur (profil du visiteur, activité du visiteur dans l'application lors de la visite, etc.).

L'objectif principal de ce sujet de thèse est de développer de nouvelles techniques d'analyse du mouvement des visiteurs, en construisant pour chaque visiteur sa *trajectoire* à l'intérieur du musée. L'originalité de ce travail vient d'*enrichir les trajectoires des visiteurs* avec (1) *des contraintes d'espace intérieur* restreignant les mouvements du visiteur (ex. : position des portes, couloirs, etc.), et (2) *des informations contextuelles et/ou sémantiques* liées au musée ou au visiteur.

## 2. Défis scientifiques

Le sujet de thèse aborde trois défis scientifiques qui seront développés ci-après.

### Défi 1. Vers une norme pour les trajectoires sémantiques en milieu fermé basées sur le modèle SITM.

La collecte des données est prévue pendant la thèse, ce qui nous permet de modifier le protocole de collecte selon les besoins afin d'obtenir des données qualitatives. Pour un musée, trois types d'informations doivent être collectées : les informations topologiques sur le bâtiment du musée (couloir, porte, ascenseur, etc.), les informations de localisation des visiteurs et les informations contextuelles/sémantiques liées aux déplacements des visiteurs (par exemple l'activité sur l'application, intention de visite, etc.) et les points d'intérêt (par exemple nature et sémantique associés aux œuvres du musée).

Ce défi bénéficiera des travaux que nous avons déjà réalisés dans [3] où nous avons proposé un nouveau modèle de trajectoires spatio-temporelles en milieu fermé enrichies d'annotations sémantiques, appelé *Semantic Indoor Trajectory Model (SITM)*. Le modèle de trajectoire SITM est composé de :

- Un *modèle d'espace intérieur*, basé sur le modèle d'espace multicouche de IndoorGML<sup>1</sup>, sous la forme d'un arbre où à la verticale nous avons une relation *contient* et à l'horizontale nous avons une relation *accessible*.
- Un *formalisme pour une trajectoire sémantique* composée de traces :  $T_{IDmo, t_{start}, t_{end}} = (\text{trace}_{IDmo, t_{start}, t_{end}}, A_{traj})$ , où  $\text{trace}_{IDmo, t_{start}, t_{end}} = (e_k, v_k, t_{startk}, t_{endk}, A_k) \quad k \in [1, n]$
- Un ensemble d'*informations sémantiques* enrichissant les trajectoires et les traces, représentées par des ensembles  $A$ .

Ici, le défi est d'exprimer les données collectées sous le formalisme proposé en fournissant, si nécessaire, une série d'adaptations. Par exemple, dans nos travaux précédents, nous avons limité les relations entre deux nœuds à *contient* et *accessible*. La question ici est de savoir s'il est adapté et possible d'utiliser d'autres relations entre nœuds déjà intégrés dans IndoorGML. De plus, nous pouvons constater que chaque cas d'utilisation doit adapter le modèle SITM existant. Le sujet de thèse présenté vise à proposer le modèle SITM comme *standard pour la représentation des trajectoires sémantiques en milieu fermé*, basé sur IndoorGML.

### Défi 2. Analyse descriptive des trajectoires sémantiques en milieu fermé (SITM)

Depuis des décennies, les musées souhaitent *comprendre* leurs visiteurs. Ainsi, dans ce deuxième défi, nous visons à développer un ensemble d'*analyses descriptives* sur des données de trajectoire représentées à travers le modèle SITM présenté dans [3]. Ainsi, nous nous concentrons ici sur les techniques d'exploration de données de trajectoires qui ont rencontré un intérêt important au cours de la dernière décennie [5], et l'objectif est d'*extraire des motifs de trajectoire*

<sup>1</sup><http://www.indoorgml.net/>

décrivant le mouvement des visiteurs. Les motifs permettraient de comprendre parmi quelle(s) catégorie(s) se situent les visiteurs du musée selon la forme de leur visite : fourmi, papillon, sauterelle ou poisson comme décrit dans [4], ou de découvrir de nouvelles catégories.

Les études précédentes s'intéressaient à l'exploration de motifs séquentiels annotés temporellement [6], à l'extraction de motifs à partir de mouvements humains enrichis en sémantique [7] ou à l'intégration de la topologie des milieux fermés dans le processus de fouille de motifs [8]. À notre connaissance, aucune des approches de fouille de motifs existantes ne permet d'extraire des motifs à partir de données de trajectoire enrichies des spécificités de l'espace intérieur et d'informations sémantiques concernant le visiteur et la visite.

### **Défi 3. Approches prédictives des trajectoires sémantiques en milieu fermé (SITM)**

Alors que le défi précédent vise à aider les musées à améliorer l'expérience des visiteurs, ce troisième défi vise à encourager les musées à prendre *des décisions de gestion* (telles que l'amélioration des itinéraires d'évacuation) en fournissant *des prévisions de mouvements*.

À cette fin, nous proposons de formaliser les trajectoires SITM sous forme *de séries temporelles de trajectoires* et de travailler au développement *d'algorithmes de classification de séries temporelles de trajectoires*. Notre choix est motivé par l'existence de séries temporelles multidimensionnelles permettant de représenter la sémantique et la topologie au sein du SITM. Ces dernières années, les méthodes d'apprentissage profond ont rejoint la classification des séries temporelles (TSC) [9] et l'algorithme InceptionTime [10] surpasse les modèles non profonds pour la TSC.

Pour la classification des séries temporelles de trajectoires SITM, la difficulté est de trouver comment intégrer les informations topologiques et les informations sémantiques dans les algorithmes d'apprentissage profond. Une solution que nous envisageons est l'utilisation de réseaux de neurones convolutifs (CNN) utilisés pour le traitement des données spatiales et/ou temporelles. Une deuxième solution consiste dans le mécanisme d'attention (transformers) qui est utilisé pour le traitement séquentiel des données et permet au modèle d'ajuster dynamiquement l'importance de certains éléments de la séquence d'entrée à la sortie.

**Votre profil.** Nous espérons accueillir un candidat remplissant les conditions suivantes :

- Réalisation d'un excellent master ou diplôme en informatique
- Solides compétences et expérience en programmation
- Des connaissances de base dans les domaines suivants sont très appréciées : fouille de données, apprentissage profond, classification de séries temporelles
- Capacité à développer des méthodes et des concepts
- Volonté de contribuer à des projets interdisciplinaires
- Compétences organisationnelles et analytiques
- Capacité à travailler en équipe, compétences en résolution de problèmes et pensée créative
- Excellentes compétences en communication orale et écrite en anglais

**Conditions de la thèse.** La thèse se déroulera au sein de l'équipe de recherche DUKe du laboratoire LS2N, Nantes, France. La plupart des membres de l'équipe DUKe travaillent à Polytech Nantes (Ecole d'Ingénieurs de Nantes Université), le doctorant y disposera donc d'un bureau. Nous sommes une équipe de recherche dynamique composée de 20 chercheurs permanents et d'une dizaine de doctorants et postdoctorants.

Compte tenu du contexte de collaboration internationale du sujet de thèse, le doctorant devra peut-être effectuer des séjours de recherche au Laboratoire de Recherche en Muséologie de l'Université Ionienne de Corfou, en Grèce et également visiter certains des musées fournissant les données.

**Comment candidater.** Les candidats intéressés peuvent joindre à leur candidature les documents suivants :

- Curriculum vitae

- Lettre de motivation spécifique au poste proposé
- Résumé du mémoire de master
- Diplômes des dernières années
- Au moins deux lettres de recommandation
- Liste des publications (le cas échéant)

Les candidats intéressés doivent envoyer leur candidature à :  
Claudia Marinica ([Claudia.Marinica@univ-nantes.fr](mailto:Claudia.Marinica@univ-nantes.fr))

### Jalons.

- T+0 - T+2 : Prise en main du sujet et état de l'art autour du domaine d'application : analyse des mouvements des visiteurs dans les musées
- T+2 - T+8 : Etat de l'art sur le sujet de modélisation des trajectoires enrichies (en se basant sur un état de l'art existant)
- Etat de l'art sur le sujet d'extraction de motifs fréquents à partir de trajectoires
  - Etat de l'art sur le sujet de classification de séries temporelles multivariées avec des méthodes d'apprentissage profond, et de classification des trajectoires avec des techniques de séries temporelles
- T+8 - T+12 : Collecte et nettoyage de données de mouvement de visiteurs dans les musées partenaires. Modélisation en utilisant le formalisme SITM
- Proposition d'un standard pour les trajectoires enrichies en milieu fermé
- T+12 - T+18 : Proposition de nouvelles approches d'analyse descriptive des trajectoires SITM en se basant sur les techniques de fouille de trajectoire
- T+18 - T+20 : Evaluation des approches proposées
- T+20 - T+21 : Publication des résultats
- T+21 - T+27 : Proposition de nouveaux algorithmes pour la classification de trajectoires modélisées sous la forme de séries temporelles. Les nouveaux algorithmes se baseront sur des techniques d'apprentissage profond.
- T+27 - T+29 : Évaluation des algorithmes proposés.
- T+29 - T+30 : Publication des résultats
- T+30 - T+36 : Rédaction du manuscrit

### Références

- [1] John H. Falk and Lynn D. Dierking. The Museum Experience Revisited. *Routledge*, 1st edition, 2013.
- [2] Alexandros Kontarinis, **Claudia Marinica**, Dan Vodislav, Karine Zeitouni, Anne Krebs. Towards a Better Understanding of Museum Visitors' Behavior through Indoor Trajectory Analysis. *Seventh International Conference on Digital Presentation and Preservation of Cultural and Scientific Heritage (DiPP2017)*, Sep 2017, Burgas, Bulgaria. pp.19-30
- [3] Kontarinis, A., Zeitouni, K., **Marinica, C.** et al. Towards a semantic indoor trajectory model: application to museum visits. *Geoinformatica* 25, 311–352 (2021).
- [4] Lanir, J., Kuflik, T., Sheidin, J. et al. Visualizing museum visitors' behavior: Where do they go and what do they do there?. *Pers Ubiquit Comput* 21, 313–326 (2017).
- [5] Mazimpaka, J.D.; Timpf, S. Trajectory data mining: A review of methods and applications. *J. Spat. Inf. Sci.* 2016, 2016, 61–99
- [6] Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. 2007. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '07)*. Association for Computing Machinery, New York, NY, USA, 330–339.
- [7] Renso, C., Baglioni, M., de Macedo, J.A.F. et al. How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowl Inf Syst* 37, 331–362 (2013).
- [8] L. Radaelli, D. Sabonis, H. Lu and C. S. Jensen, Identifying Typical Movements among Indoor Objects - Concepts and Empirical Study, 2013 *IEEE 14th International Conference on Mobile Data Management*, Milan, Italy, 2013, pp. 197-206.

- [9] Ismail Fawaz, H., Forestier, G., Weber, J. et al. Deep learning for time series classification: a review. *Data Min Knowl Disc* 33, 917–963 (2019).
- [10] Ismail Fawaz, Hassan, et al. "InceptionTime: Finding alexnet for time series classification." *Data Mining and Knowledge Discovery* 34.6 (2020): 1936-1962.