



**HAL**  
open science

## Services augmentés pour le tourisme intelligent et l'analyse des pratiques

Maxime Masson

► **To cite this version:**

Maxime Masson. Services augmentés pour le tourisme intelligent et l'analyse des pratiques. Forum Jeunes Chercheuses Jeunes Chercheurs (JCJC) - INFORSID 2022, INFORSID, May 2022, Dijon, France. hal-04326824

**HAL Id: hal-04326824**

**<https://hal.science/hal-04326824>**

Submitted on 6 Dec 2023

**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.

---

# Services augmentés pour le tourisme intelligent et l'analyse des pratiques

**Maxime Masson**

*Laboratoire LIUPPA, Collège STEE, Université de Pau et des Pays de l'Adour  
Avenue de l'Université, 64000 Pau, France  
maxime.masson@univ-pau.fr*

---

*MOTS-CLES : Réseaux Sociaux, Tourisme, Traitement Automatique du Langage (TAL), Analyses Multidimensionnelles, Proxémique, Extraction d'Informations.*

*KEYWORDS : Social Media, Tourism, Natural Language Processing (NLP), Multidimensional Analysis, Proxemics, Information Extraction.*

*ENCADREMENT : Christian Sallaberry, Rodrigo Agerri, Marie-Noelle Bessagnet, Philippe Roose, Annig Le Parc Lacayrelle.*

---

## 1. Introduction

Les dernières décennies ont vu une croissance significative des sources de données disponibles couvrant de nombreux sujets ainsi que l'essor du contenu généré par les utilisateurs. Le tourisme, en particulier, est un domaine dans lequel l'obtention et l'étude d'ensembles de données thématiques massifs et précis sont essentiels pour mieux comprendre certains phénomènes et tendances.

Ces jeux de données peuvent être utilisées par les professionnels du tourisme à différentes échelles. Tout d'abord, pour aider au processus de prise de décision des acteurs du tourisme en leur permettant de faire les choix les plus éclairés possibles. Cela se fait en explorant les données pour arriver à mieux comprendre la pratique et les besoins des visiteurs. Ce type d'analyse est aussi particulièrement utile pour les entreprises, notamment celles spécialisées dans le marketing touristique (comme les *organisations de gestion de destination*), pour lesquelles il est crucial de bien cerner les désirs et les attentes des touristes (Belias *et al.*, 2021). D'autre part, les systèmes de recommandation touristique sont devenus un enjeu important à la fois pour les touristes et les acteurs locaux. Ils permettent, en examinant la pratique touristique et les lieux visités par un grand nombre de personnes, de suggérer les activités ou itinéraires touristiques les plus adaptés possible au profil du visiteur.

Différentes ressources peuvent être utilisées pour accéder à des données touristiques. Historiquement, il s'agit principalement de bases de données, avec : (1) les bases de données commerciales, telles que celles des agences de voyages en

ligne (OTA, *Online Travel Agencies*) et (2) les bases de données publiques, par exemple celles produites par les agences gouvernementales ou celles fonctionnant sur un principe de *crowdsourcing*. Ces dernières font partie de ce que l'on appelle le « contenu généré par les utilisateurs ». Cette catégorie de données s'est considérablement développée, via les réseaux sociaux généralistes (*Twitter*, *Facebook*) ou les sites d'avis (*TripAdvisor*). L'Internet des objets est également un nouveau vecteur de données clef pour le tourisme intelligent (ex : surveillance et prédiction des flux de visiteurs). Enfin, les systèmes de traçabilité des touristes sont de plus en plus utilisés pour suivre avec précision les itinéraires empruntés.

Notre projet s'inscrit dans la continuité de ce phénomène. Nous souhaitons concevoir et développer une plateforme visant à collecter, traiter, analyser puis valoriser des données relatives à la pratique du tourisme, aux flux de visiteurs et à la fréquentation des points d'intérêt dans la région du Pays Basque, un territoire hautement touristique qui s'étend entre la France et l'Espagne. Il convient de déterminer si la construction de trajectoires touristiques multidimensionnelles (spatiale, temporelle, thématique) individuelles dans un premier temps puis agrégées corrélée à la mise en place de services d'analyse et l'introduction de la notion de proxémique à ces dernières peut contribuer à des analyses pertinentes, à fois pour les touristes (*recommandation*) et les professionnels (*aide à la décision*). Nous évoquerons les motivations derrière notre travail puis les actions réalisées et futures.

## 2. Motivations

Afin d'illustrer les grandes phases de mon travail de thèse, nous prenons appui sur le concept de *WaterWheel* (Bucher *et al.*, 2021), une roue montrant le cycle de vie des données dans le système d'information que nous allons mettre en place.

La 1<sup>ère</sup> phase (Fig. 1, *Collecte*) consiste en la recherche de ressources liées au tourisme et à la collecte des données associées. A cette fin, il faut définir précisément les jeux de données nécessaires en fonction des besoins (*tourisme, zone et temporalité ciblés*) ; mettre en place une méthodologie de collecte de données sur les réseaux sociaux générique, multilingue et indépendante du domaine. Cette méthodologie doit également s'assurer de l'évaluation du jeu de données résultant. Au cours de la 2<sup>ème</sup> phase (Fig. 1, *Traitement*), nous souhaitons mettre en place une batterie de services dédiés à l'enrichissement, l'édition, et l'agrégation des données touristiques précédemment collectées. Le but est de les transformer et les croiser pour augmenter leur valeur ajoutée (*extraction de concepts, classification, détection de sentiments, etc.*). Les données seront ensuite enrichies afin d'être en mesure de définir la pratique touristique de la manière la plus complète possible. Une fois les données nécessaires en main (*collecte*) et dans la forme souhaitée (*traitement*), nous allons concevoir des services d'analyse (Fig. 1, *Analyse*) basés sur les dimensions : spatiale (*ensemble des lieux visités*), temporelle (*période et durée du séjour*) et thématique (*sémantique de la pratique touristique : type d'activités, condition du séjour, profil touristique, coût*). L'une des originalités de notre travail est que nous souhaitons définir des trajectoires touristiques multidimensionnelles à partir des données collectées et y intégrer la notion de *proxémique* (Hall, 1966). La proxémique est la science qui étudie l'organisation de

l'espace et l'effet des distances dans les relations interpersonnelles. Cette dernière est généralement appliquée au monde physique (*mouvement dans l'espace*) mais l'un des défis de mon travail est de l'adapter au monde virtuel au service du tourisme intelligent. Nous cherchons à concevoir un modèle proxémique permettant d'offrir une perspective nouvelle d'analyse des pratiques touristiques et qui soit applicable aux réseaux sociaux, un lieu d'expression fréquent pour les touristes. De plus, nous souhaitons que la dimension thématique prenne le pas sur le côté strictement géographique. L'espace pourrait s'apparenter à l'espace thématique du tourisme sur le réseau social (*types de tourisme, d'activité, typologies de visiteur*).

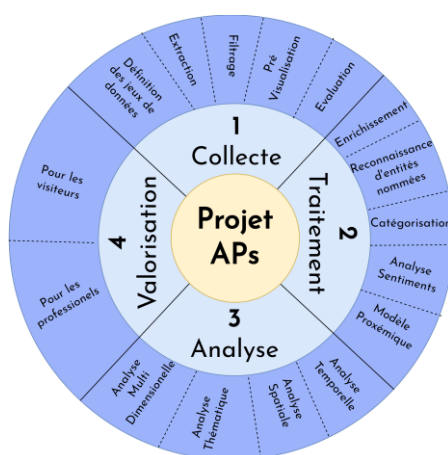


Figure 1 : Cycle de vie de l'information touristique dans le cadre de notre projet. Représentation inspirée de la WaterWheel (Bucher et al., 2021).

La 4<sup>ème</sup> phase (Fig. 1, *Valorisation*) s'intéresse à la valorisation des données précédemment collectées et traitées en visant deux cibles : les **professionnels** (i) et les **visiteurs** (ii). L'expérience des visiteurs pourrait être présentée dans des tableaux de bord multimodaux corrélant l'offre touristique et les pratiques afin d'orienter la prise de décision des aménageurs. Il serait ainsi possible de mettre en lumière les points chauds (*zones d'activité touristique intense*) passés, actuels et futurs, les couloirs de passage, ou les pratiques touristiques récurrentes. Nous souhaitons aussi construire un système de recommandation touristique calibré pour cette région si particulière qu'est le Pays Basque en prenant appui sur l'analyse des trajectoires touristiques multidimensionnelles précédemment citées afin d'être en mesure de suggérer les itinéraires les plus adaptés en fonction du profil du visiteur.

### 3. Actions réalisées

Rappelons que la 1<sup>ère</sup> phase de notre travail de recherche est la mise en place d'une méthodologie générique pour la collecte de données touristiques. Nous avons choisi les réseaux sociaux comme principale ressource pour plusieurs raisons : (i)

leur facilité d'accès, (ii) pas de mise en place de longues campagnes de collecte, (iii) pas de nécessité d'acheter de coûteux systèmes de suivi des touristes ou même des données commerciales à des sociétés spécialisées. De plus, ces données sont massives et très diverses, c'est-à-dire que de multiples aspects du tourisme sont couverts. L'extraction sélective à partir de cette masse de données est un défi complexe et a été l'objet du début de notre travail.

A notre connaissance, il n'existe pas d'outil ou de méthode entièrement générique et applicable à n'importe quel domaine pour extraire des données thématiques très ciblées à partir de n'importe quel réseau social, les méthodes actuelles étant fortement corrélées au domaine (Scholz et Jezni, 2020). Nous avons donc conçu une nouvelle méthode générique, indépendante du domaine et itérative destinée à faciliter ce processus de collecte et l'avons expérimentée sur le thème du tourisme avec le réseau social *Twitter*. Cette méthode s'architecture autour des 3 dimensions au cœur de nos futures analyses : spatiale (*étendue spatiale des données*), temporelle (*période temporelle concernée*) et thématique (*sujet d'étude représenté sous la forme d'un vocabulaire plus ou moins complexe*). Afin de s'assurer de la qualité des données collectées, un processus d'évaluation et de prévisualisation des données a été mis en place. Les expérimentations menées nous permettent de construire des jeux de données thématiques exhaustifs avec un niveau de bruit minimal et tenant compte des contraintes inhérentes aux réseaux sociaux.

#### 4. Actions futures

En suivant les phases décrites dans notre cycle de vie, nous allons démarrer la phase 2, avec l'enrichissement puis l'extraction de concepts fins et d'entités nommées de ces données brutes via des méthodes de Traitement Automatique du Langage. En parallèle, nous travaillerons à l'intégration du concept de proxémique dans nos travaux, en l'adaptant aux données issues du cyber espace, et avec un focus sur les interactions dans les réseaux sociaux ayant trait au tourisme tout en proposant des méthodes d'analyse et d'extraction novatrices. Le but final restant toujours de concevoir un système de recommandation pour les touristes qui soit adapté à la région et différents outils d'aide à la décision pour les professionnels.

#### Bibliographie

- Belias A., Malik A., S. Rossidis, I. Mantas (2021). Use of Big Data in Tourism: Current Trends and Directions for Future Research. *Academic Journal of Interdisciplinary Studies*, vol. 10, n° 5, p. 357.
- Bucher B., Hein C., Raines D., Gouet Brunet V. (2021) Towards Culture-Aware Smart and Sustainable Cities: Integrating Historical Sources in Spatial Information Infrastructures. *ISPRS International Journal of Geo-Information*, vol. 10, n° 9, 588.
- Scholz J., Jezni J. (2020) Evaluating Geo-Tagged Twitter Data to Analyze Tourist Flows in Styria, Austria. *ISPRS International Journal of Geo-Information*, vol. 9, n° 11, 681.
- Hall E. T. (1966). *The hidden dimension*, Anchor, vol. 609.