

Projet de thèse LIA **Université d'Avignon et des Pays de Vaucluse**

Unité de recherche concernée : Laboratoire Informatique d'Avignon (LIA EA 4128), Université d'Avignon et des Pays de Vaucluse.

Directeur de thèse : Juan-Manuel Torres, Maître de Conférences en Informatique (27 section) , Habilité à Diriger des Recherches, membre du Laboratoire d'Informatique d'Avignon (LIA, EA 4128), Université d'Avignon et des Pays de Vaucluse. Professeur Associé à l'École Polytechnique de Montréal, Québec (Canada).

Encadrant : Fen Zhou, Maître de Conférences en Informatique (27 section), membre du Laboratoire d'Informatique d'Avignon (LIA), Université d'Avignon et des Pays de Vaucluse ;

Encadrant : Didier Josselin, Directeur de Recherche CNRS (ESPACE/LIA).

Sujet de thèse: « Un musée à ciel ouvert : système de recommandation intelligent pour les visites touristiques »

Résumé/objectif : Ce projet de thèse permettra d'atteindre un objectif innovant : développer un système de recommandation (SdR) multilingue qui génère des parcours de visites personnalisés en fonction des contraintes (temps, lieu, budget) et des intentions (désir, goût, envie) des visiteurs, qu'ils soient seuls ou en groupe. Il s'agit d'une transposition du problème du parcours de musées mais à ciel ouvert ; avec des enjeux de granularité particulières. Les contraintes et les intentions seront énoncées en langue naturelle par les touristes visitant une ville. Nous voulons également évaluer l'impact de visites proposées par le SdR pour satisfaire au mieux les besoins des touristes qui l'utilisent. Ces recherches seront menées en croisant les expertises du LIA en Optimisation combinatoire et en Traitement Automatique de Langues (TAL).

Objectif scientifique

Dans des villes touristiques la progression permanente du nombre de touristes montre le goût persistant du public pour une pratique tangible et partagée de la culture et des patrimoines historiques (dans le cas d'Avignon, par exemple, on a le Festival d'Avignon, le Palais des Papes, le Pont d'Avignon, les musées, les champs de lavande, etc). Le tourisme telle que nous l'entendons ici est une expérience participative qui s'appréhende à travers des interactions entre le touriste, ses partenaires, une ville, l'environnement, les monuments, les documents associés et les dispositifs numériques. Cet engouement pour le tourisme pose des problèmes critiques tant aux offices de tourisme qu'aux visiteurs potentiels : la gestion et l'accès aux monuments, le choix des parcours de visite, la liste des monuments à visiter selon les intentions, les goûts, les connaissances préalables ou les partenaires de visite. Par ailleurs, on sait que les sites Internet d'une ville et les réseaux sociaux sont de plus en plus utilisés pour préparer une visite touristique, visite qui s'effectue à plusieurs dans 62% des cas (1). Grâce à une téléphone intelligente, un système recommandant (SdR) pourrait être installé comme une application qui permettrait de localiser les touristes et de leur proposer un ou plusieurs parcours personnalisés et dynamiques. L'objectif radicalement innovant du projet est le suivant : formaliser le fonctionnement et les impacts d'un SdR des parcours touristiques en prenant appui sur plusieurs domaines d'activité en Informatique: le traitement automatique des langues (TAL) et la recherche opérationnelle (RO). Ce SdR devrait permettre, de proposer des parcours de touristiques personnalisés et adaptés aux intentions d'un groupe de touristes (dyade ou plus). Il devra également tenir compte de l'ensemble des contraintes et motivations fournies par chaque touriste potentiel. Enfin grâce aux potentialités de procédures de *reporting*, *des blogs et des réseaux sociaux*, il est envisagé d'évaluer les expériences de chaque touriste, ainsi que de dresser des connaissances des pratiques de visite. Ces expériences sont liées au vécu subjectif de la visite des lieux et des œuvres, et au

parcours dans l'espace géographique de ces visiteurs. Ces trajectoires individuelles constituent autant de graphes localisés qui dépendent des profils et des pratiques des touristes, de leur représentation de l'espace (via des « cartes mentales » (7)), et de la façon dont l'urbanisme les influence (concept de « space syntax » proposé en par HULLIER (8) et influence de la structure des réseaux, par exemple (9)). Le projet de thèse se justifie par l'expression de ces représentations par le langage (partie TAL) et par la trajectoire suivie par le touriste (partie recherche opérationnelle et optimisation combinatoire des points d'intérêt). De plus, on présuppose une certaine rationalité et l'existence de profils typiques qui vont permettre au SDR de proposer des parcours adaptés aux différents types de visiteurs, en tenant compte de la multiplicité des chemins alternatifs possibles (10) selon la combinaison des préférences individuelles.

Verrous scientifiques

Cette thèse pose des verrous scientifiques pluridisciplinaires importants :

i/ Avant la visite : il faut une sélection automatique de textes permettant au touriste de préparer la visite dans la langue ou dans un bouquet de langues de son choix et selon des critères qui lui sont propres. Le problème scientifique concerne entre autres, le résumé automatique personnalisé multi-documents et multilingue (français, espagnol, anglais, chinois) (2).

ii/ Arrivé dans la ville : Le GPS dans un téléphone intelligent permettra aux touristes de se repérer dans les différents monuments, ce qui peut fournir des statistiques sur les disponibilités des monuments dans la ville. En fonction du trajet effectué en début de visite, et les disponibilités des monuments, il faudra conseiller au touriste une suite des monuments tenant compte du temps qu'il souhaite y consacrer et tenant compte des contraintes de confort tel que l'accessibilité aux handicapés, restaurant et hôtel. La problématique scientifique abordée concerne les méthodes d'optimisation de routage et l'allocation de ressource dans les réseaux de communication (5,6).

iii/ Après la visite : dans l'application de la téléphone intelligent, on devra permettre au touriste d'exprimer son avis sur ce qu'il a visité, et aussi et surtout consulter celui des autres. La problématique scientifique concernera la fouille de textes et plus précisément, la détection automatique d'opinion. Ce pavé peut bien entendu être utilisé avant la visite sur le versant de la consultation. Il a des questions ouvertes concernant la constitution d'un corpus d'apprentissage selon les préférences des touristes qui seront également étudiées. À ce niveau, des inférences statistiques permettront de dégager des profils types en termes de motivations et de pratiques spatiales des visites. Des parcours optimaux seront proposés et pourront être testés auprès du public. Un point original est qu'il ne s'agira pas obligatoirement de faire des visites qui minimisent le temps de visite, mais plutôt qui maximisent la satisfaction (certaines oeuvres amenant probablement davantage de temps de contemplation que d'autres). Les successions de mouvement, de pause et de repos fixeront probablement une fonction d'objectif assez originale en la matière.

Cadre expérimental et originalité du sujet

Nous utiliserons de outils de Traitement Automatique de Langues (TALN) (2,3,4), afin pouvoir constituer et traiter des corpus issus des questionnaires, des sites des villes, des sites des agences de voyages et sur internet. Ces algorithmes ont déjà obtenu de bons résultats dans des tâches telles que la classification automatique, le résumé automatique, la détection et la classification d'opinions (4). Ces outils restent assez indépendants de la langue et de la thématique, donc ils sont facilement adaptables aux besoins des expériences. Nous utiliserons également des outils de programmation linéaire en nombres entiers pour résoudre les problèmes d'optimisation de parcours de tourisme de manière optimale et les algorithmes gloutons pour proposer des itinéraires de bonne qualité en ligne. Ces méthodes sont déjà utilisées dans les réseaux optiques pour le routage de lumière (6) et dans les réseaux de contenu pour la diffusion de vidéos (5),

ainsi qu'en analyse spatiale des réseaux (11). Des représentations des cartes des déplacements individuels (géométrie des mobilités) et mentales (graphe de perception des œuvres et monuments et de leurs interactions) seront associées pour décrire les profils des touristes et aider à la conception/représentation des solutions optimales.

Évaluation

L'évaluation des résultats se fera du point de vue quantitatif et qualitatif. En ce qui concerne l'évaluation quantitative, des modèles de programmation linéaire en nombres entiers et des algorithmes heuristiques pour optimiser les parcours des visites seront étudiés afin de fournir un *webservice* pour le SdR. D'ailleurs, des statistiques et une analyse fine au niveau des algorithmes TALN et des résultats en termes de Recherche d'information, seront effectués: précision, rappel, F-score. Le terrain de recherche, la méthodologie et les références scientifiques formeront les appuis de l'évaluation qualitative. En amont des analyses, mais tout au long de la recherche elle-même, l'évaluation tiendra compte des allers-retours entre les références scientifiques et les résultats des analyses. Une attention sera portée aux publications du doctorant qui, pendant la durée du contrat doctoral, sera encouragé à publier dans des revues scientifiques nationales et internationales.

- Mots clés : Programmation linéaire en nombres entiers, Algorithme heuristique, Planification des parcours de visites dans une ville, Traitement Automatique de Langues, Apprentissage automatique, Évaluation.

Références

- (1) Eidelman, Jacqueline. et Jonchery, Anne. 2012. *À l'écoute des visiteurs édition* Paris, La documentation française.
- (2) Juan-Manuel Torres-Moreno. *Résumé automatique de documents : Une approche statistique*. Hermès Lavoisier, 2011, ISBN 978-2-7462-3212-9
- (3) Jean-Valère Cossu, Benjamin Bigot, Ludovic Bonnefoy, Mohamed Morchid, Xavier Bost, Grégory Senay, Richard Dufour, Vincent Bouvie, Juan-Manuel Torres-Moreno and Marc El-Bèze, *LIA @ RepLab 2013*, CLEF 2013 Evaluation Labs and Workshop, Online Reputation Management – RepLab.
- (4) Juan-Manuel Torres-Moreno, Marc El-Bèze, Frédéric Béchet, Nathalie Camelin, *Fusion probabiliste appliquée à la détection et classification d'opinions*, DEFT'09, Paris, France, 22 juin 2009, 15p.
- (5) Fen Zhou, Jiayi Liu , Gwendal Simon, Raouf Boutaba. *Joint optimization for the delivery of multiple video channels in Telco-CDN*. In Proceeding of IEEE/ACM CNSM, pp1-5, 2013.
- (6) Fen Zhou et al. *Cost Bounds and Approximation Ratios of Multicast Light-trees in WDM Networks*. IEEE Journal of Optical Communications and Networking, 3(4): 323-334, 2011.
- (7) Gould Peter, White Rodney, *Mental Maps*, 1986, Taylor & Francis.
- (8) Hillier Bill, *Space is the machine : a configurational theory of architecture*, 2007
- (9) Garaix Thierry, Artigues Christian, Feillet Dominique, Josselin Didier, *Vehicle routing problems with alternative paths: an application to on-demand transportation*, *EJOR*, 204, 1 (2010) 62-75.
- (10) Genre-Grandpierre Cyrille, Foltête Jean-Christophe, Josselin Didier, *Impacts of Road Networks on urban mobility*, in Theriault M., Des Rosiers F (ed) *Modeling urban Dynamics, Geographical Information systems series*, ISTE, Wiley, 2011, pp. 103-128.

(11) Gerd Finke (Ed), *Recherche opérationnelle et réseaux, méthodes d'analyse spatiale*, traité IGAT, Hermès Lavoisier, 2002.

Moyens

Type Financement : Contrat doctoral MESR LIA

Date de début de la thèse: *(Septembre 2014)*

Candidature : Candidat identifié

Correspondant/Contact : {juan-manuel.torres, fen.zhou}@univ-avignon.fr

Nom : Torres Prénom : Juan-Manuel Tél : 04 90 84 35 68 Fax : 04 90 84 35 01

Mail : juan-manuel.torres@univ-avignon.fr

Nom : Zhou Prénom : Fen Tél : 04 90 84 35 46 Fax : 04 90 84 35 01

Mail : Fen.Zhou@univ-avignon.fr

Nom : Josselin Prénom : Didier Tél : 04 90 84 35 74 Fax : 04 90 84 35 01

Mail : didier.josselin@univ-avignon.fr