



Cahier des charges

Projet d'Ouverture à la Recherche

Découverte de relations insolites sur le patrimoine

Réalisé par : Hawa Fofana, p2212201 (Master 1 Informatique, UCBL1)
Grace Tatidoung Tsoualla, p2212516 (Master 1 Informatique, UCBL1)

Encadrées par : Fabien Duchateau (Maître de conférences, LIRIS)
Franck Favetta (Maître de conférences, LIRIS)

Table des matières

I) Contexte	2
II) Présentation du projet	3
1) Objectifs	3
2) Travail à faire	3
III) Contraintes	4
IV) Livrables	4
V) Planning prévisionnel	4

I) Contexte

Ce projet s'inscrit dans le cadre d'un constat simple : La nécessité d'explorer en matière d'extraction d'informations, de nouveaux types de relations, notamment celles dites insolites. En effet, de nombreux outils d'extraction et bases de connaissances se concentrent encore aujourd'hui sur les types de relations fréquents (né-e à, capitale de, auteur-e de, etc.)¹. Pour ce faire, ces systèmes se basent sur une liste prédéfinie de types de relations. Il est cependant possible de travailler sur des systèmes plus ouverts nécessitant un minimum de support (au moins K exemples de relations pour un type à partir des documents textuels existants) afin de détecter ces nouveaux types de relations. Il faut noter que les documents textuels sont difficilement exploitables directement par les machines².

Parmi ces nouveaux types, on distingue la catégorie des relations insolites qui sont probablement les plus difficiles à détecter indépendamment de l'un ou l'autre des systèmes précédents. Par exemple, la basilique de Fourvière a inspiré l'architecture de l'église Notre Dame des Victoires de San Francisco, ou le quai de la pêcherie à Lyon a servi successivement de halle aux poissons, port de pêche puis sentier de promenade et marché aux livres. Tout au long de ce projet, il sera donc question pour nous d'étudier davantage ces relations insolites avec pour domaine d'application le patrimoine local de Lyon.

¹ C. Niklaus, M. Cetto, A. Freitas, and S. Handschuh. A survey on open information extraction. Computational Linguistics, 2018. Lien CL2018.

² X. Han, T. Gao, Y. Lin, H. Peng, Y. Yang, C. Xiao, Z. Liu, P. Li, M. Sun, and J. Zhou. More data, more relations, more context and more openness: A review and outlook for relation extraction. JCNLP, 2020. Lien JCNLP2020.

II) Présentation du projet

1) Objectifs

L'objectif de ce projet sera :

- D'établir une liste de types de relation (plus ou moins) insolites avec des exemples à partir de diverses sources de données sur le patrimoine local (la ville de Lyon en particulier et la région Auvergne-Rhône-Alpes en général), les classifier et construire un jeu de données expertisé.
- Tester et comparer ensuite plusieurs outils d'extraction d'informations sur le jeu de données produit.
- Souligner enfin les limites des outils actuels dans un rapport, les défis à relever et proposer si possible quelques pistes pour améliorer les résultats d'extraction de relations insolites.

2) Travail à faire

Le travail est découpé en 4 grandes tâches :

Tâche 1 : Extraction manuelle de relations insolites

- 1.1 Rapide état de l'art sur l'extraction d'informations (articles en bibliographie)
- 1.2 Étude et analyse de diverses sources de données sur le patrimoine ([patrimoine de Lyon](#), [patrimoine Auvergne Rhône Alpes](#), [grand Lyon](#), [Wikipédia](#)...)
- 1.3 Extraction de plusieurs types de relations insolites (une dizaine)
 - 1.3.1. Identifier les éléments (paragraphe, phrases, images) intéressants des sources de données
 - 1.3.2. Identifier les types de relation insolite éventuellement contenu dans les éléments repérés
- 1.4 Relever si possible à partir des mêmes sources de données une dizaine d'exemples pour chaque type

Tâche 2 : Modélisation de relations insolites

- 2.1. Étudier la structure des types de relations insolites identifiés
- 2.2. Proposer une représentation, compréhensible par la machine
- 2.3. Créer le jeu de données (un fichier) en fonction de la représentation adoptée
- 2.4. Structurer et organiser nos relations dans le jeu de données

Tâche 3 : Tester la pertinence et la cohérence du jeu de données

- 3.1. Tester au moins deux outils d'extraction d'informations sur ce dernier
- 3.2. Comparer les différents résultats obtenus

Tâche 4 : Identifier les perspectives de relations insolites

- 4.1. Souligner les limites des outils d'extraction en général
- 4.2. Souligner les limites des outils d'extraction appliqués aux relations insolites en particulier
- 4.3. Souligner les défis à relever
- 4.4. Proposer des pistes d'amélioration

III) Contraintes

- Outils d'extraction : Ollie, ReVerb, Stanford OpenIE
- Langage : Python
- Organisation : deux réunions par mois.

IV) Livrables

- Cahier de charge
- Un jeu de données de relations insolites
- Résultats des tests des outils d'extraction sur le jeu de données
- Rapport final sur les limites de ces outils et les pistes d'amélioration
- Poster et vidéo de vulgarisation

V) Planning prévisionnel

