

Génération automatique d'ontologies et de graphes de connaissances
via des données des SNS et de mobilité des gens pour la valorisation des POIs peu connus

Landy Rajaonarivo

INRAE, UMR TETIS

Janvier 2025

- 1 Contexte
- 2 Questions de recherche
- 3 Approche proposée
- 4 Expérimentations

Contexte

- De nombreuses villes sont confrontées au "tourisme de masse",
- Cette concentration de personnes peut détériorer l'environnement, perturber les écosystèmes et altérer la qualité de vie des locaux.

Motivation :

- Disperser les visiteurs afin d'atténuer le problème du "tourisme de masse".
- Faire connaître la culture et l'identité des locaux aux touristes, résidents et autorités locales.
- Étudier la mobilité des gens pour déterminer s'il y a un événement local à un endroit donné pendant une période donnée

Challenge :

POIs moins connus sous-entend moins d'informations disponibles

Exemples de POIs moins connus

Japan's Prefectures

Fukui
福井県

Echizen, petite ville connue par : sa **culture artisanale ancienne** de haute qualité de **1500 ans** (fabrication couteau, papier washi, porterie), son **trésor de la nature** (sources thermales, montagnes, mer)



RYUSEN Factory, Echizen, Fukui
Fabrication artisanale de couteau



Endroit **moins connu** mais **représentatif** de sa ville qui est **riche en culture et en nature**



Washi park

L. Rajaonarivo (INRAE)



Echizen Ono Castle

Séminaire

4 / 22

Caractéristique de POIs moins connu

POI : attraction touristique, site culturel, architectural ou récréatif, restauration, hébergement, etc.

POI moins connu : présentant des caractéristiques qui intéressent les touristes, mais rarement mentionné sur les plateformes bien connus tels que Foursquare, TripAdvisor, Yelp, Booking, Google My Business, etc.

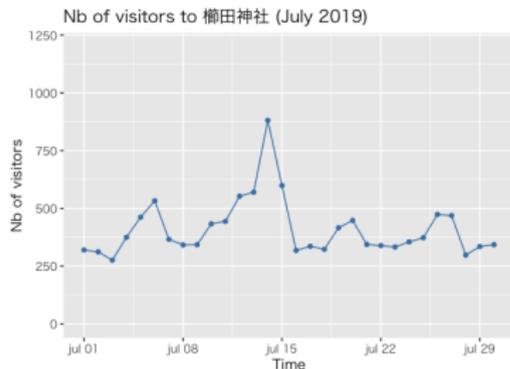
Exemple de POIs moins connus

POI	G		Yp	FQ	JI	Chat GPT	Approche	
	GR	M,P,R	M,P,R	M,P,R	M,P,R	C,D,E	# TW	Estimated categories
Ryūsen hamono (龍泉刀物), Fukui	219K	(Y,14,12)	(Y,0,0)	(Y,0,0)	(Y,9,0)	(Y, N,N)	349	Retail
Kametani yōgyō (亀谷窯業), Shimane	8K	(Y,16,10)	(Y,0,0)	N	(Y,7,1)	N	41	Retail Dining
Noguchi senpo (野口染舗), Hokkaido	16K	(Y,10,7)	N	N	(Y,15,2)	N	63	Retail

GR : recherche sur Google, M: Maps, P: Photo, R: Review, C: Catégorie, D: Détail, E: Explication, Y : Yes, N : No

(en 2023)

Étude de mobilité des gens



Visiteurs du temple Kushida, Fukuoka, Japon
Provenant de NICT Japon

L'étude de mobilité permet de détecter s'il y a un évènement mais **connaître sa nature et ses détails est un challenge.**

Pour connaître le détail :

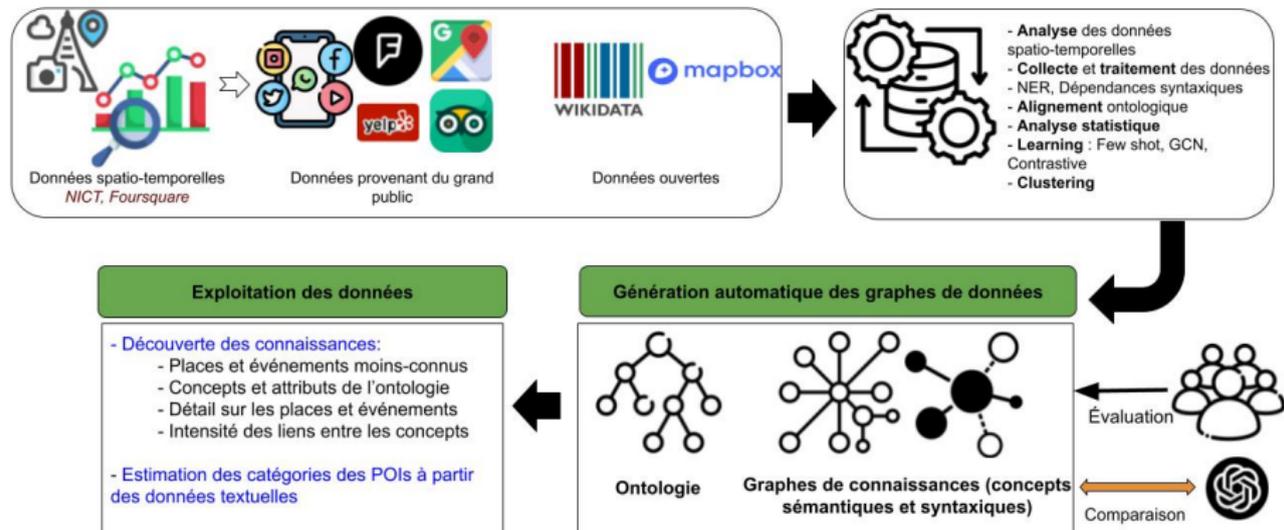
- Qu'est ce que les gens font là-bas ?
- Qui y vont et quand ?
- Qu'est ce qu'on peut visiter autour ?
- Comment y accéder ?



Yamakasa event, Fukuoka (en Juillet)

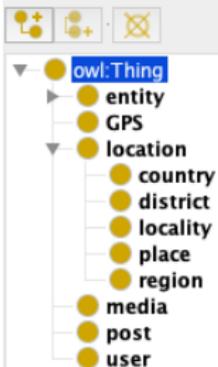
- ① Quelles **sources** permettent de récupérer des informations sur les POIs moins connus ? → **Données textuelles** : réseaux sociaux, articles publiés; **Données spatio-temporelles**
- ② Comment **normaliser les informations** venant de différentes sources afin qu'elles soient re-utilisables après ? → **ontologie et graphe de connaissance** (sémantique et syntaxique)
- ③ Quelles **technologies utiliser pour encoder et récupérer** les informations pertinentes et les liens entre elles ? → **GCN, Contrastive L., NLP, Few-shot L.**

Approche proposée

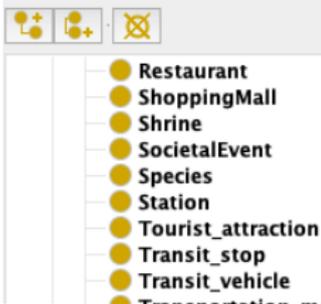


Extrait d'une ontologie générée

Class hierarchy: owl:Thing



Class hierarchy: entity



Active ontology x Entities x Individuals by class x DL Quer

Annotation properties Datatypes Individuals
Classes Object properties Data properties

Object property hierarchy: owl:topObjectPr



Active ontology x Entities x Individuals by class x DL Quer

Annotation properties Datatypes Individuals
Classes Object properties Data properties

Data property hierarchy: owl:topDataProper



Exemple d'instances des ontologies générées

The image displays two screenshots of an ontology editor interface. The top screenshot shows the 'Canal_City_Theater' instance. The left panel lists direct instances: Canal_City_Theater and キャナルシティ. The main panel shows the class hierarchy for 'Tourist_attraction' with subclasses like Shrine, Species, Station, etc. The right panel shows property assertions for 'Canal_City_Theater', including 'locatedIn Japan', 'locatedInAdministrative Hakata', and 'hasPart Canal_City_Theater'.

The bottom screenshot shows the 'Mount Fuji' instance. The left panel lists direct instances: Fukuoka_Tower, Izu Islands, Mount Fuji, Niagara Falls, Tsunoshima, and 白糸ハイランドウェイ. The main panel shows the class hierarchy for 'Tourist_attraction' with subclasses like Shrine, Species, Station, etc. The right panel shows property assertions for 'Mount Fuji', including 'hasGPS lat35_3625-long138_7305555556' and 'hasAdminLocation 'Shizuoka Prefecture''. It also lists data property assertions for aliases like 'Mt. Fuji' and 'Fujiyama'.

Instance d'attraction touristique au Japon

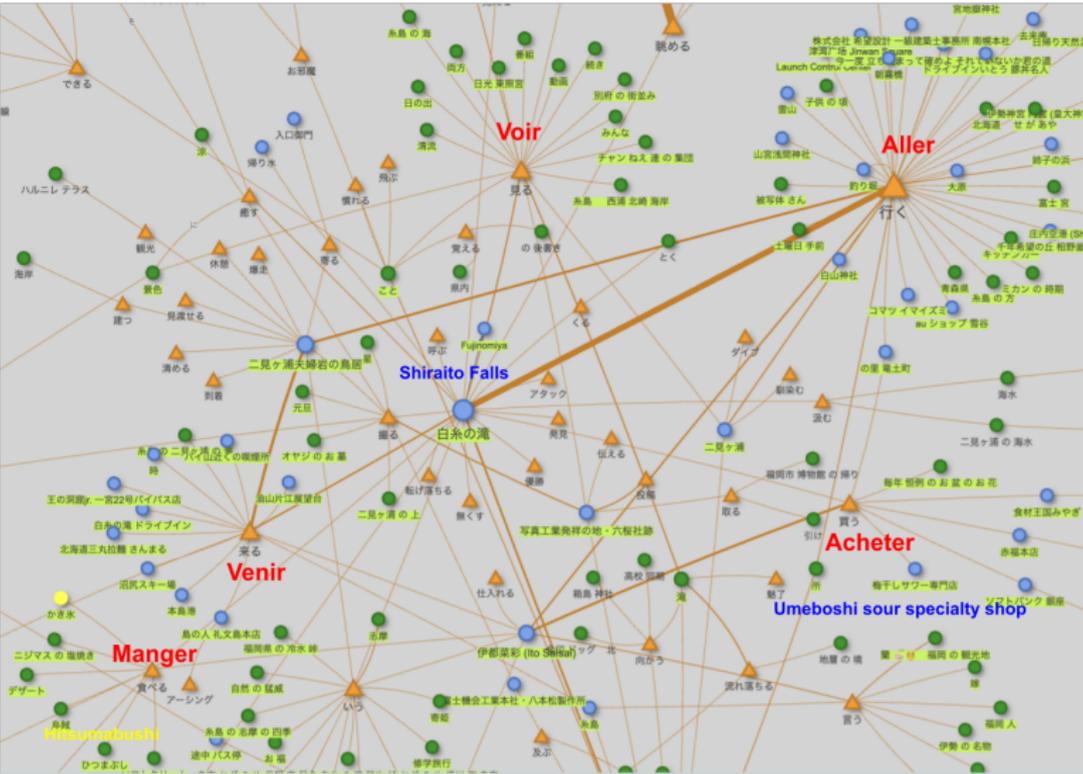
Exemple d'instances des ontologies générées

The screenshot displays a web-based ontology editor interface. At the top, there are tabs for 'Active ontology', 'Entities', 'Individuals by class', and 'DL Query'. The main window is divided into several panels:

- Class hierarchy: Place**: A tree view on the left showing a hierarchy of classes. 'Place' is selected and highlighted in blue. Other classes include Municipality, Organisation, Person, Plant-organ, etc.
- Annotations: hendaye**: A panel showing the annotations for the instance 'hendaye'. It lists 'rdfs:label' with the value 'hendaye' and a type of 'xsd:string'.
- Description: hendaye**: A panel showing the types of the instance. It lists 'Border-town', 'City', 'Commune-of-france', 'Place', and 'Poi'. 'Place' is the most specific type shown.
- Property assertions: hendaye**: A panel showing the property assertions for the instance. It lists several properties with their values and types, such as 'hasAdminLocation' with the value 'pyrénées-atlantiques', 'freetext' with the value 'hendaye france', and 'latitude' with the value '-1.77444'.
- Direct instances: hendaye**: A panel at the bottom left showing a list of direct instances for the class 'Place'. The instance 'hendaye' is highlighted.

Instance de place en France (tweets sur le pays Basque)

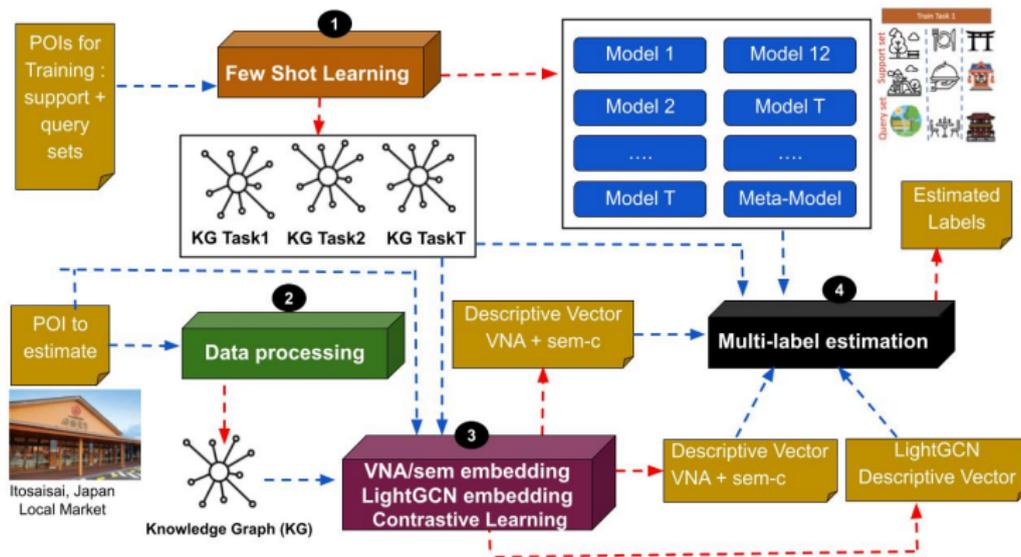
Extrait d'un graphe de connaissance syntaxique



Représentation des données & Apprentissage

Représentation des données :

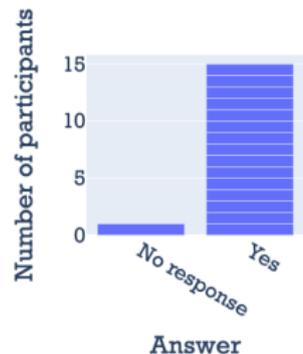
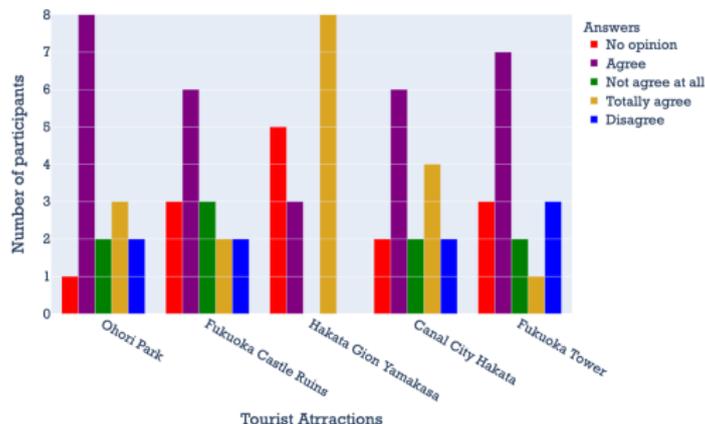
- via des **concepts sémantiques** : activité (sport, loisir, etc.), nourriture, type des lieux (restauration, transport, etc.)
- via des **concepts syntaxiques** : verbe, nom, adjectif



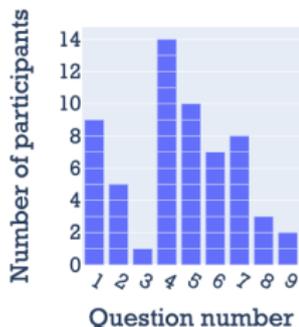
Approche de représentation et de labellisation des POIs

Expérimentation : enquête sur l'ontologie générée

Answers about Tourist Attractions



Question sur la découverte



- 1: Tourist Attractions
- 2: Stations
- 3: Japanese foods
- 4: Religious buildings
- 5: Dates related to Yamakasa event

- 6: Photos taken around KS at the Yamakasa event
- 7: UNESCO's recognition of Yamakasa event
- 8: Shopping malls
- 9: Japanese festivals

Expérimentation : comparaison et ablation

Approches de représentation des données

Evaluation					
	Model	A	P	R	F1
1	BERT (large-uncased + base-japanese)	0.238	0.039	0.166	0.064
2	BERT (bert-large-uncased)	0.023	0.137	0.416	0.172
3	SKLEARN multi-label classification	0.769	0.305	0.305	0.305
4	App + (cosine+sim) + (verbs)	0.771	0.417	0.292	0.333
5	App + (cosine+sim) + (nouns)	0.739	0.087	0.043	0.058
6	App + (cosine+sim) + (adj)	0.753	0.448	0.362	0.391
7	App + sim + (VNA)	0.753	0.425	0.534	0.448
8	App+ cosine + (VNA)	0.643	0.312	0.393	0.32
9	App + (cosine+sim) + (VNA)	0.782	0.534	0.603	0.54
10	App + VNA + semantic	0.9	0.73	0.8	0.77

Exemple des réponses à des questions

Exploitation des données des graphes de connaissances

Que **manger** à **Fukuoka** ?

- Manger du karashi mentaiko de Fukuoka (219)
- Manger motsunabe à Rakutenchi (42)
- Manger du ramen à Fukuoka (25)
- Manger des fleurs de colza en printemps (24)



karisha mentaiko

Que **voir** à **Fukuoka** en printemps?

- voir les fleurs de cerisier à Mizuki Ruins (51), voir les fleurs de cerisier à Ohori Park (38)

D'**où** viennent les gens qui publient des info sur le **temple Kushida**?

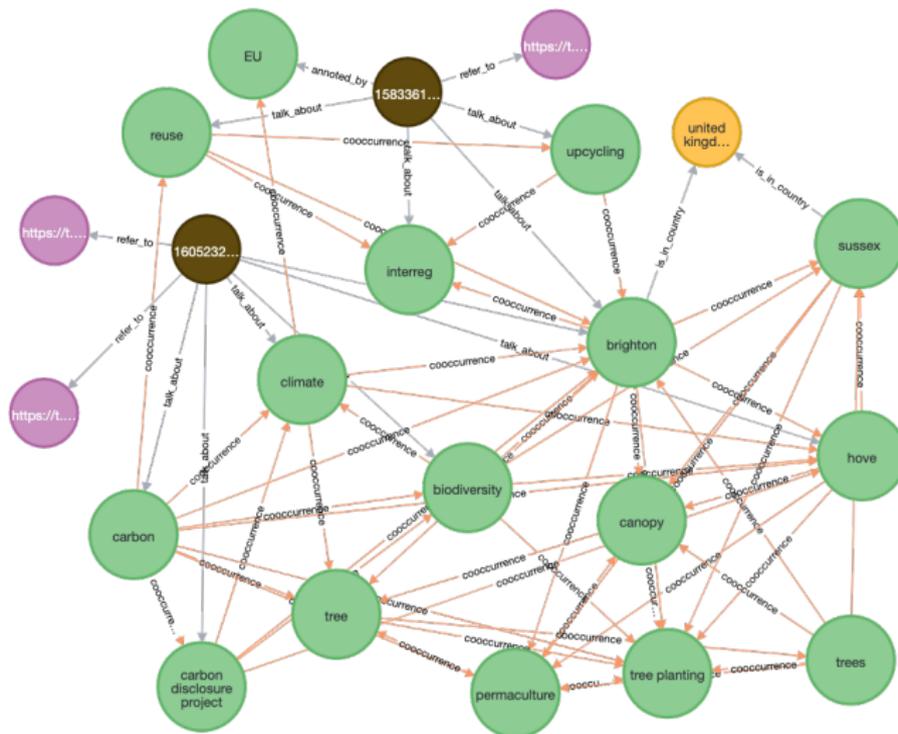
- Miyazaki, Fukuoka, Hiroshima, Toronto, etc.

Comment **se rendre** au marché **Ito Sai Sai**?

- en bus, en voiture

Extrait de graphe de connaissance

Collection des tweets au niveau mondial sur les initiatives de l'Économie Circulaire entre 2013 et 2023



- Approche :
 - Génération automatique d'ontologies et de graphe de connaissance
 - Représentation et découverte de connaissance (peu-connue)
 - Couplage des études sémantiques et syntaxiques
 - Utilisation des apprentissages automatiques

- Limites et Perspectives:
 - Génération d'ontologie : ambiguïté dans certains cas (NER)
 - Estimation de labels : labels influencés par le contexte (*avantage* : informations complémentaires sur les POIs, *inconvénient* : faux labels détectés sur certains POIs)
 - Collecte de données : la langue considérée influence la collecte