Reconstitution d’un jardin à partir d’un texte

Robin LEMAÎTRE

POM 2022 - Encadré par M. DUCHATEAU et M. FAVETTA

Contexte et problématique

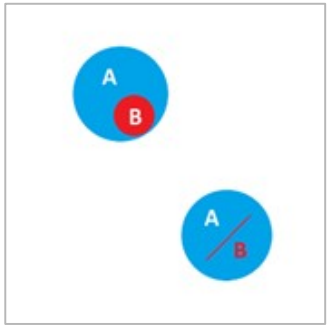
- Des jardins et des parcs qui n’existent plus
  - Jardins suspendus de Babylone
  - Parc de Malmaison

  ⇒ **Comment aider à reconstituer ces jardins disparus ?**

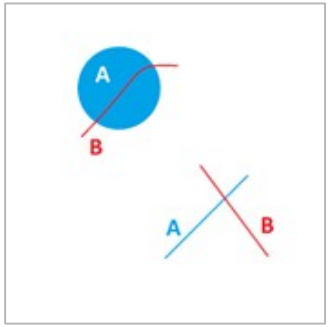
- Exploitation des traces écrites
  - Identifier les informations pertinentes dans un texte
  - Évaluer les différentes interprétations d’un texte

État de l’art

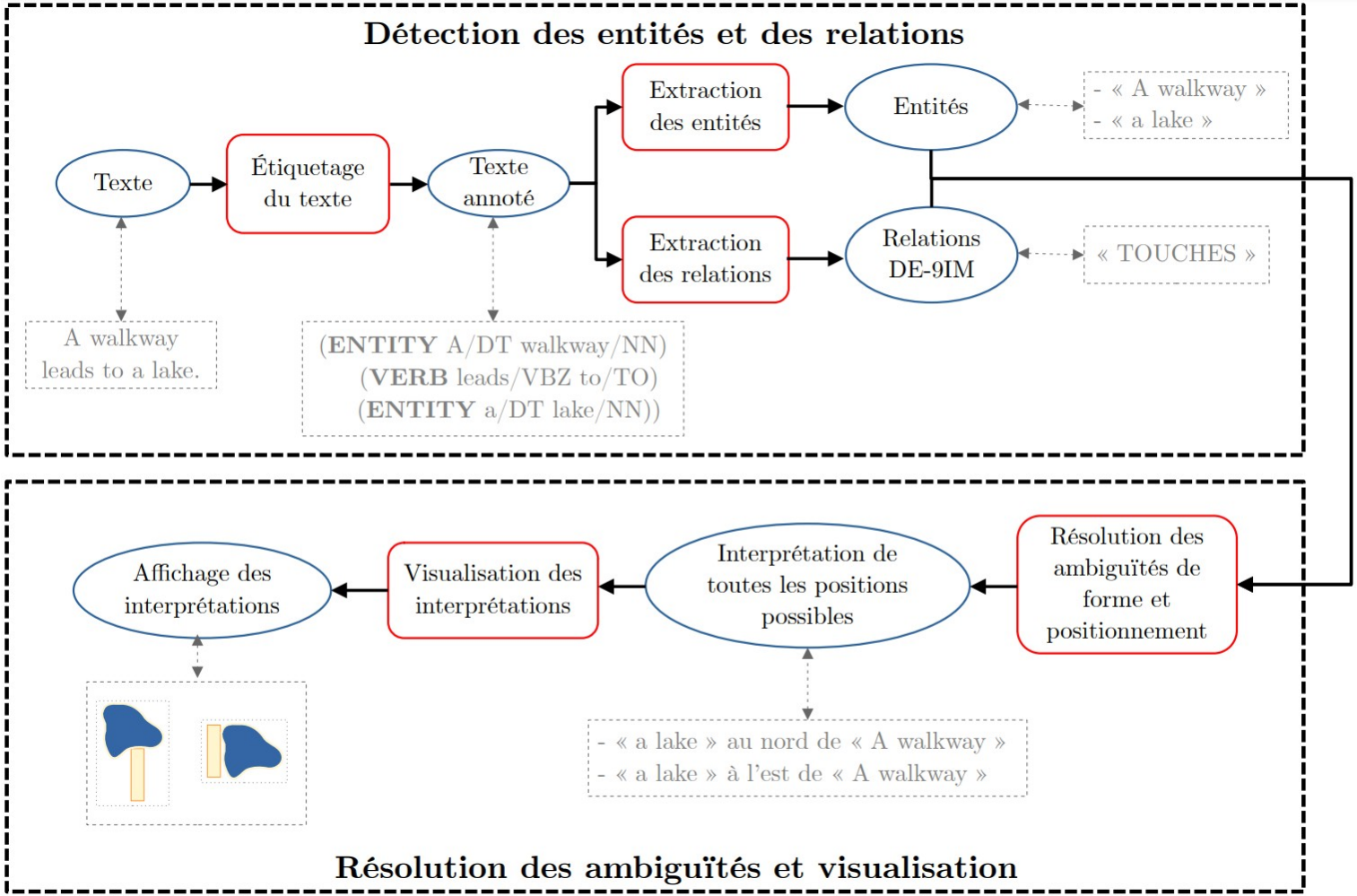
- Traitement Automatique de la Langue Naturelle
  - Étiquetage des mots et des phrases (POS-Tagging & NP-Chunking)
  - Reconnaissance d’entités nommées (NER)
- Modèles spatiaux décrivant des relations
  - Plusieurs types de relations spatiales
  - Relations topologiques : DE-9IM
- Traitement des ambiguïtés visuelles



A CONTAINS B



B CROSSES A


Détection des entités et des relations
Texte
Étiquetage du texte
Texte annoté
Extraction des entités
Extraction des relations
Entités
Relations DE-9IM
- « A walkway »
- « a lake »
« TOUCHES »
A walkway leads to a lake.
(ENTITY A/DT walkway/NN)
(VERB leads/VBZ to/TO)
(ENTITY a/DT lake/NN))
Résolution des ambiguïtés de forme et positionnement
Interprétation de toutes les positions possibles
Visualisation des interprétations
Affichage des interprétations
- « a lake » au nord de « A walkway »
- « a lake » à l'est de « A walkway »
Résolution des ambiguïtés et visualisation

Première partie de notre proposition

Détection d’entités et de relations

I - Les étiquetages

- Étiquetage de chaque mot (*POS-Tagging*)
  - "There is a lake to the north of the park Tête d'Or. A large island emerges from the lake."
  - → There/EX is/VBZ a/DT lake/NN to/TO the/DT north/NN of/IN the/DT park/NN Tête/NNP d/NN '/NN Or/NNP./. A/DT large/JJ island/NN emerges/VBZ from/IN the/DT lake/NN./.

- Étiquetage des groupes grammaticaux (*NP-Chunking*)
  - Définition d'une grammaire (une douzaine de règles)
  - → (SPREL (ENTITY A large island) (VERB emerges from) (ENTITY the lake))

```
VERB:
{<VB.*>+<TO|RP|IN>*}
SPREL:
{<ENTITY><VERB><ENTITY>}
```

2 règles de grammaire

II - Extraction d’entités

- Un groupe nominal = une entité
- Des ambiguïtés grammaticales de liens entre les entités :
  - Articles indéfinis et définis (“Un lac”, “Le lac”)
  - Pronoms personnels (“Il” = “un lac”)
  - Pronoms relatifs : Qui, Que, etc.
  - Cas plus complexes : “Ce dernier”, etc.

- « There is a lake to the north of the park Tête d’Or. A large island emerges from the lake. »

→ a lake, the park Tête d’Or, A large island, the lake

III - Extraction de relations

- Un verbe (de relation) = une relation DE-9IM ou cardinale (Nord, etc.)
  - Exploitation du type de formulation de la phrase
  - Utilisation de WordNet pour les synonymes
  - Représentation sous forme de tuples : < A, relation, B>

→ < a lake, NORTH, the park Tête d’Or >

→ < A large island, WITHIN, the lake >

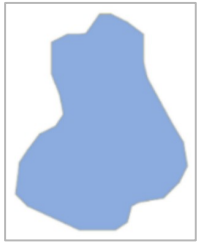
Seconde partie de notre proposition

Résolution des ambiguïtés et visualisation

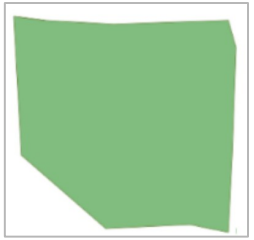
I - Les formes

- Quelle forme adopter pour une entité ?
  - Texte
  - Image raster
  - Image vectorielle (polygone)

- Utilisation d'images vectorielles
  - Formes prédéfinies pour certains types d'entités
  - Formes par défaut



a lake

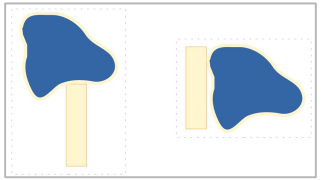


the park Tête-d'Or

a large island

II - Les positionnements

- Comment positionner des entités par rapport aux autres ?
  - Différentes interprétations possibles
  - Manque d'informations



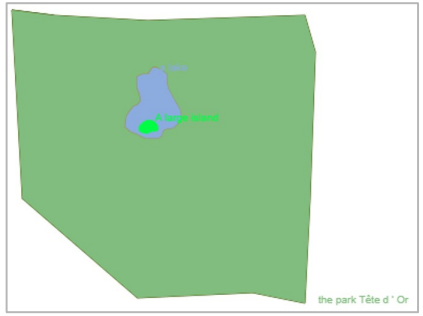
« Un chemin mène au lac »

- Parcours récursif des entités
  - Le parc, entité principale
  - Conservation de toutes les interprétations pour chaque relation

- Interprétation 1
  a. *a lake* → **nord** du *parc Tête d'Or*
  b. *A large island* → **centre** du *lac* → nord du *parc Tête d'Or*

- Interprétation 2
  a. *a lake* → **nord** du *parc Tête d'Or*
  b. *A large island* → **sud** du *lac* → nord du *parc Tête d'Or*

Interprétation 1



Interprétation 2

The text

Write a description here, or select a pre-defined text below.

There is a lake to the north of the park Tête d'Or. A large island emerges from a lake.

Render!

- A walkway leads to a lake, where there is an island.
- A tree-lined path leads to a fountain, which is surrounded by flowers.
- An island emerges from the lake.
- A path leads to the lake. It is surrounded by trees. An island is located in the lake.
- There are flowers to the north of the park.
- There is a lake and a fountain.
- a lake is surrounded by trees.
- There is the lake to the north of the park. A large path leads to the lake. an island emerges from the lake.

Result (text)

The following entities / relations have been detected :

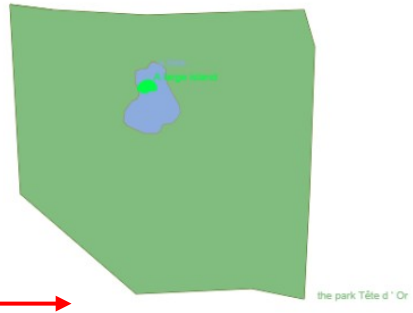
Entities
- a lake (id: 2)
- the park Tête d ' Or (id: 0)
- A large island (id: 6)

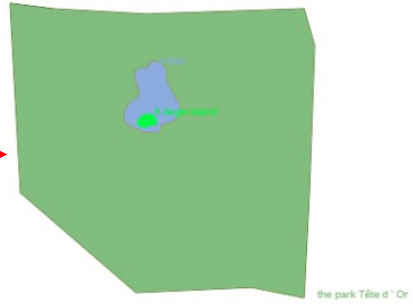
DE-9IM relations
- < a lake (id: 2), the north, the park Tête d ' Or (id: 0)>
- < A large island (id: 6), within, a lake (id: 2)>

Spatial relations ambiguities
- within : A large island can be located in a lake from different location and size (a lake can be partially or nearly-totally covered by A large island, or A large island can be located to the center or South of a lake).

Different interpretations
- Version 0
- Version 1

Result (draw)





within : A large island can be located in a lake from different location and size (a lake can be partially or nearly-totally covered by A large island, or A large island can be located to the center or South of a lake).

Conclusion

- Proposition d’une solution à la problématique de reconstitution des jardins
  - Définition de règles de grammaire
  - Correspondance verbes <-> relation DE-9IM avec WordNet
  - Gestion de nombreuses ambiguïtés (grammaticales, formes, positionnements)

- Perspectives
  - Amélioration de la détection des liens grammaticaux (apprentissage)
  - Traitement des relations métriques (application des solutions existantes en NER)
  - Résolution des superpositions et amélioration de la représentation graphique (interactions, appariement des entités nommées)
  - Confrontation de plusieurs textes sur le même jardin
  - Évaluation de la proposition (comparaison avec de vrais parcs)

- Expérience enrichissante avec cette initiation à la recherche

Fin de la présentation

Merci pour votre attention

Robin Lemaître – POM (Groupe 11) – 16 juin 2022