



Doctorale EARIA

Recherche d'information précise par processus hybrides

Romain Beaumont

Organisme d'accueil :

LIMSI-CNRS

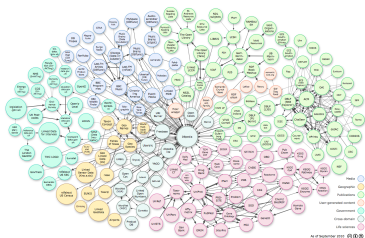
Encadrantes :

Brigitte Grau

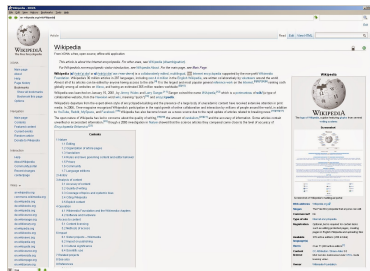
Anne-Laure Ligozat

15 octobre 2014

Question : *Who is the daughter of Bill Clinton married to?*

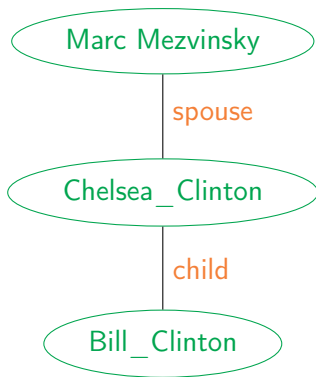


Réponse : *Marc Mezvinsky*



Who is the daughter of Bill Clinton married to ?

Base de connaissance



Texte

Chelsea Clinton, daughter of former President Bill Clinton and former Secretary of State Hillary Clinton... In 2010, Chelsea Clinton and investment banker Marc Mezvinsky were married in an interfaith ceremony in Rhinebeck, New York.

Réponse : Marc Mezvinsky

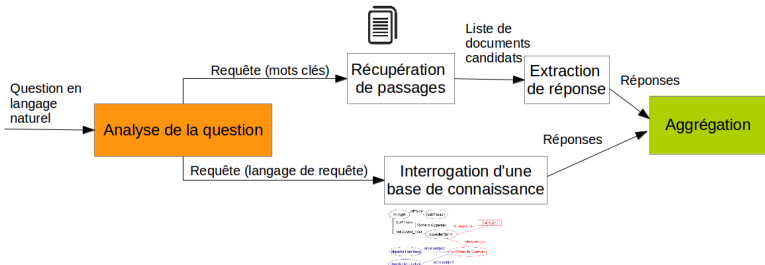
Recherche d'information précise dans des bases de connaissance et des corpus textuels

Base de connaissance

- + structure explicite
- + plus normalisé
- + moins ambigu
- informations manquantes

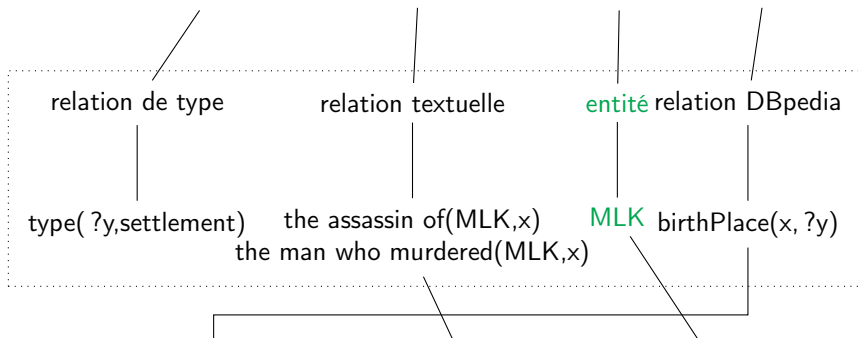
Texte

- pas de structure explicite
- beaucoup de variations
- parfois ambigu
- + nombreuses informations et corpus



Recherche d'information précise dans des bases de connaissance et des corpus textuels

In which city was the assassin of Martin Luther King born ?



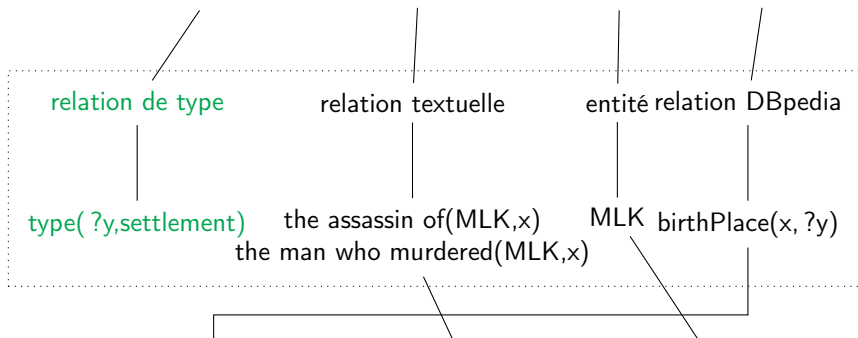
What is the place of birth of the man who murdered Martin Luther King ?

distance lexicale

ambiguïté lexicale

Recherche d'information précise dans des bases de connaissance et des corpus textuels

In which city was the assassin of Martin Luther King born ?



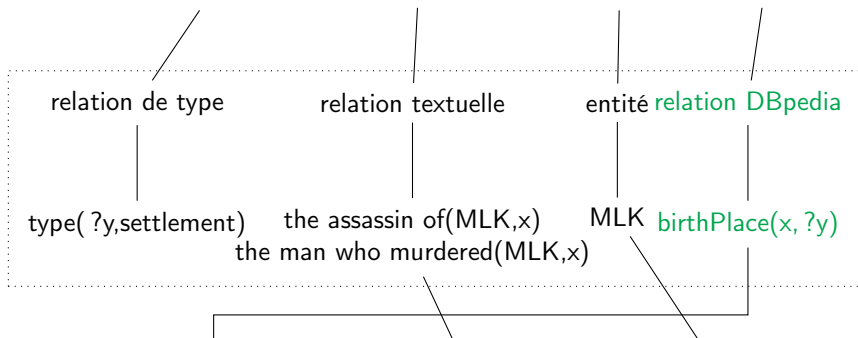
What is the place of birth of the man who murdered Martin Luther King ?

distance lexicale

ambiguïté lexicale

Recherche d'information précise dans des bases de connaissance et des corpus textuels

In which city was the assassin of Martin Luther King born ?



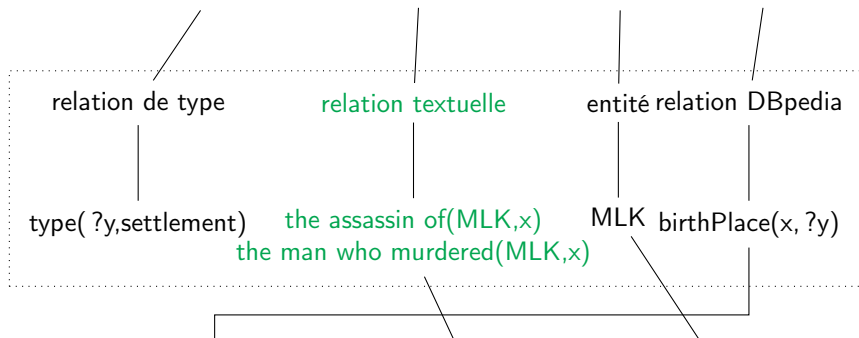
What is the place of birth of the man who murdered Martin Luther King ?

distance lexicale

ambiguïté lexicale

Recherche d'information précise dans des bases de connaissance et des corpus textuels

In which city was the assassin of Martin Luther King born ?



What is the place of birth of the man who murdered Martin Luther King ?

distance lexicale

ambiguïté lexicale

Représentation sémantique de la question

Ensemble de triplets partiellement instanciés (e_1, L_r, e_2)

- e_1 et e_2 des entités et L_r une liste de relations possibles (conservation des ambiguïtés)

Méthode de formation de cette représentation

- Reconnaissance des entités : guidé par DBpedia ou un outil de reconnaissance d'entités nommées
- Reconnaissance des types de relation : gérer les variations linguistiques avec WordNet et des bases de paraphrase
- Former la représentation sémantique : utilisation des dépendances syntaxiques pour lier les entités et relations
- Génération d'une requête : pour l'instant uniquement SPARQL

Thèse

- Modèle d'analyse de la question intégré
 - Reconnaissance conjointe des entités et relations
 - Meilleure prise en compte des ambiguïtés
- Amélioration du traitement des variations linguistiques
- Développer le modèle de recherche hybride

Références



Unger, C., Forascu, C., Lopez, V., Ngomo, A., Cabrio, E., Cimiano, P., and Walter, S. (2014). Question answering over linked data (qald-4). CLEF.



Xu, K., Feng, Y., and Zhao, D.
Xser@ qald-4 : Answering natural language questions via phrasal semantic parsing.



Yahya, M., Berberich, K., Elbassuoni, S., and Weikum, G. (2013). Robust question answering over the web of linked data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1107–1116. ACM.