

Gestion de points de vue sur des données RDF

Ludivine Duroyon
avec François Goasdoué et Ioana Manolescu

13 novembre 2017

Ce travail a été effectué dans le cadre du projet ANR
ContentCheck « Modèles, algorithmes et architectures pour le
fact-checking et le journalisme de données », 2016-2019

Partenaires du projet

Inria Saclay, IRISA, LIMSI, LIRIS, et Le Monde (*Les Décodeurs*)

Gestion d'informations pour le journalisme

Quels types d'informations ?

- **Des faits** *"il existe un complot contre M. Fillon"* qui peuvent nécessiter de compléments :
- ▶ **Du contexte/des connaissances du domaine** *"M. Fillon est un politicien, candidat aux élections présidentielles"*
- ▶ **Une dimension temporelle** *"Le complot a duré le temps des élections"*
- ▶ **Des points de vue** *"M. Larcher est persuadé qu'il y a bien un complot, M. Hollande affirme qu'il n'y a pas de complot"*

Gestion d'informations pour le journalisme

Quelles questions sont posées ?

- Qui a été l'assistant parlementaire de M. Fillon en 2012 ?
- Qui pense qu'il y a eu un complot contre M. Fillon durant les élections ?
- Qui est au courant du point de vue de M. Hollande ?

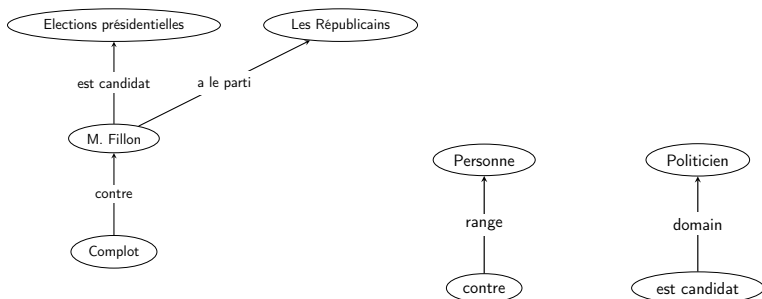
Contributions

- ① **Modèle de données (W3M)** expressif et flexible pour les faits, des agents, les points de vue des agents, du temps et des connaissances.
- ② **Inférence** (formalisation des conséquences logiques d'une instance du modèle), notamment par la propagation des connaissances entre les agents ; **approximation finie** de l'ensemble de conséquences (qui peut être infini)
- ③ **Langage de requêtes (W3Q)** pour interroger des instances du modèle, basé sur la sémantique temporelle et des points de vue.
- ④ **Représentation finie** du modèle (même infini) et approche pour évaluer des requêtes.

Sommaire

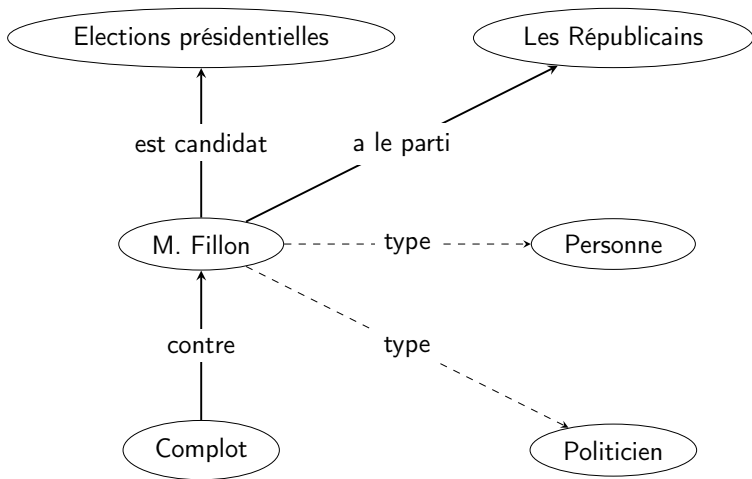
- 1 Introduction
- 2 RDF
- 3 W3M
- 4 W3Q
- 5 Structure de Kripke
- 6 Conclusion

Graphes RDF et règles RDF Schema



Nom	Règle de dérivation
rdfs2	$(b_p, domain, b_o), (b_{s_1}, b_p, b_{o_1}) \rightarrow (b_{s_1}, type, b_o)$
rdfs3	$(b_p, range, b_o), (b_{s_1}, b_p, b_{o_1}) \rightarrow (b_{o_1}, type, b_o)$

Assertions RDF explicites et implicites

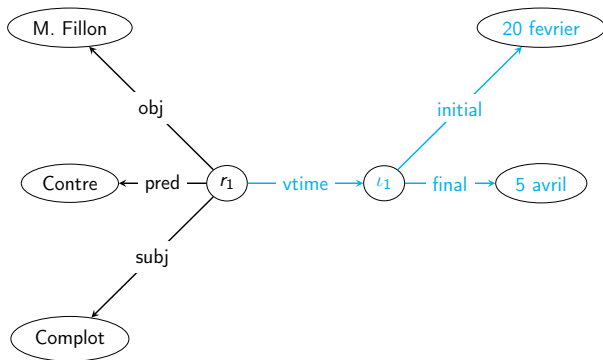


Modèle de données *Who said What When (W3M)*

- ▶ Basé sur RDF et RDFS
- ▶ Temps des faits : **temps de validité**
- ▶ Points de vue des agents sur les faits par des **croyances**, *publiques* ou *privées*, *positives* ou *négatives*.
- ▶ **Temps des croyances**
- ▶ **Règles d'inférence** de nouvelles croyances + positionnement temporel

Exprimer le temps

- Tous les faits et toutes les croyances ont un intervalle de temps associé.
- Le temps est attaché aux triples par *réification*
- Exemple : « Il existe un complot contre M. Fillon du 20 février au 5 avril »



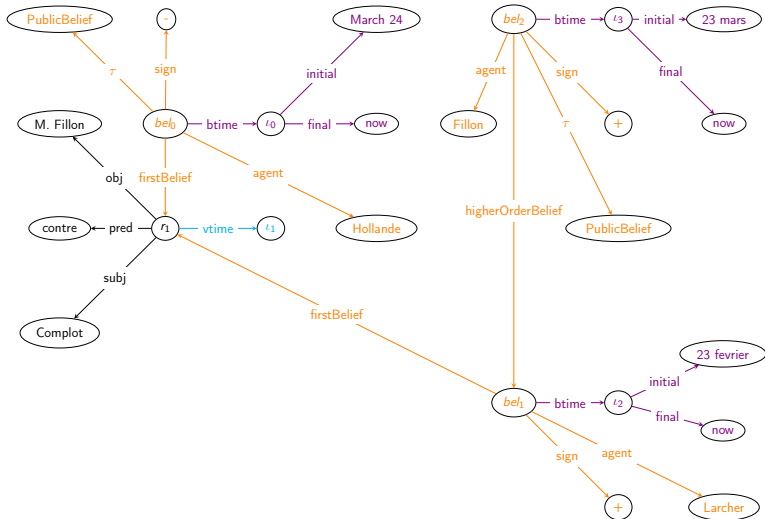
Exprimer les points de vue

Le point de vue des agents est représenté par ses **croyances**.

Une croyance est définie par :

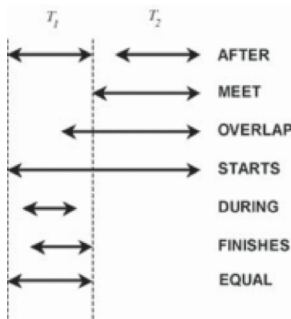
- Un **agent**
- Un **signe**
- Un **type : croyance privée ou publique**
- Un lien vers son sujet ; si c'est un fait, c'est une croyance de premier ordre ; si c'est une croyance c'est une croyance d'ordre supérieur.

Instance de W3M



Règles sémantiques temporelles

Inspirées des prédicats d'Allen, bien connues pour caractériser les relations temporelles



Exemple de règle

Si l_1 est avant l_2 et l_3 commence quand l_2 finit, alors l_1 est avant l_3 .

Règles sémantiques des points de vue

Modélisation de l'espace public :

- Tous les agents sont au courant de toutes les croyances publiques ;
- Tout le monde sait que tout le monde est au courant.

Règles d'inférence de croyances :

Nom	Règle de dérivation
viewpoint1	$(b_1, firstBelief, r), (b_1, \tau, PublicBelief),$ $(a_1, agent, b_1), (a_2, \tau, Agent),$ $(b_1, btime, \iota_1), (\iota_1, initial, i_1) \rightarrow$ $(b_2, agent, a_2), (b_2, higherOrderBelief, b_1),$ $(b_2, \tau, PublicBelief)(b_2, sign, +),$ $(b_2, btime, \iota_2), (\iota_2, initial, i_1), (\iota_1, final, now)$

Déduction de croyances

A partir de notre instance W3M, on peut déduire que :

- M. Fillon croit que M. Hollande croit qu'il n'y a pas de complot $F H \bar{C}$
- M. Hollande croit M. Fillon croit que M. Hollande croit qu'il n'y a pas de complot $H F H \bar{C}$
- M. Fillon croit que M. Hollande M. Fillon croit que M. Hollande croit qu'il n'y a pas de complot $F H F H \bar{C}$
- ...

Saturation d'une instance de W3M

Definition (Saturation)

La *saturation* \mathcal{I}^{W3M} d'une instance W3M \mathcal{I} , est \mathcal{I}^{RUTUV} , où R désigne les règles RDF, T les règles temporelles et V les règles des points de vue.

\mathcal{I}^{W3M} peut être infinie à cause des règles V .

Definition (Saturation bornée)

Une *approximation finie* de \mathcal{I}^{W3M} est $\mathcal{I}^{(RUTUV)_k}$ (chaînes d'inférence de longueur maximale k)

Proposition (Saturation stratifiée)

La *saturation stratifiée* $((\mathcal{I}^R)^{V_k})^T)^R$ est finie, et :

$$\mathcal{I}^{(RUTUV)_k} \subseteq (((\mathcal{I}^R)^{V_k})^T)^R \subseteq \mathcal{I}^{W3M}$$

La saturation bornée et celle stratifiée sont des graphes RDF.

Langage de requêtes W3Q

Le langage est fondé sur les *Basic Graph Patterns* au coeur de SPARQL :

- Contre qui y a-t-il un complot ?

$q(x) \leftarrow (\textit{complot}, \textit{contre}, x)$
 $\{M.Fillon\}$

- Qui est candidat aux présidentielles, et pour quel parti ?

$q(x, y) \leftarrow (x, \textit{est candidat}, \textit{elections presidentielles}),$
 $(x, \textit{a le parti}, y)$

$\{(M.Fillon, LesRepublicains)\}$

Langage de requête W3Q

agent *PRIVATE/PUBLIC*^{+/-} AT (y,z) BGP AT (y₁,z₁)

- Qui est visé par un complot et quand ?

$Q_1(x, y, z) \leftarrow (\textit{complot}, \textit{contre}, x) \textit{AT} (y, z)$

- Qui croit en privé qu'il y a un complot contre M. Fillon et depuis quand ?

$Q_2(x, y) \leftarrow x \textit{PRIVATE}^+ \textit{AT} (y, \textit{now})$
(*complot, contre, Fillon*)

- Qui croit publiquement que M. Larcher croit en privé qu'il y a un complot contre Fillon ?

$Q_3(x, y, z) \leftarrow x \textit{PUBLIC}^+ \textit{M. Larcher PRIVATE}^+$
(*complot, contre, M. Fillon*)

De W3Q à SPARQL

- W3Q :

$Q_1(x, y, z) \leftarrow (\text{complot}, \text{contre}, x) \text{ AT } (y, z)$

SPARQL :

$q_1(x) \leftarrow (v, \text{subj}, \text{complot}), (v, \text{pred}, \text{contre}), (v, \text{obj}, x),$
 $(v, \text{vtime}, \iota), (\iota, \text{initial}, y), (\iota, \text{final}, z)$

- W3Q :

$Q_2(x, y) \leftarrow x \text{ PRIVATE}^+ \text{ AT } (y, \text{now})$

$(\text{complot}, \text{contre}, \text{Fillon})$

SPARQL :

$q_2(x) \leftarrow (v, \text{agent}, x), (v, \text{type}, \text{PrivateBelief}), (v, \text{sign}, +)$
 $(v, \text{fistbelief}, w), (v, \text{btime}, \iota), (\iota, \text{initial}, y),$
 $(\iota, \text{final}, \text{now}), (w, \text{subj}, \text{complot}), (w, \text{pred}, \text{contre}),$
 $(w, \text{obj}, \text{Fillon}), (\text{complot}, \text{contre}, \text{Fillon})$

De W3Q à SPARQL

- W3Q :

$Q_3(x) \leftarrow x \text{ PUBLIC}^+ \text{ M. Larcher PRIVATE}^+$
(*complot, contre, M. Fillon*)

SPARQL :

$q_3(x) \leftarrow (v, \text{agent}, x), (v, \text{type}, \text{PublicBelief}), (v, \text{sign}, +),$
 $(v, \text{higherOrderBelief}, w), (w, \text{agent}, \text{Larcher}),$
 $(w, \text{type}, \text{PrivateBelief}), (w, \text{sign}, +),$
 $(w, \text{fistbelief}, w_2), (w_2, \text{subj}, \text{complot}),$
 $(w_2, \text{pred}, \text{contre}), (w_2, \text{obj}, \text{Fillon}),$
(*complot, contre, Fillon*)

Evaluation des requêtes

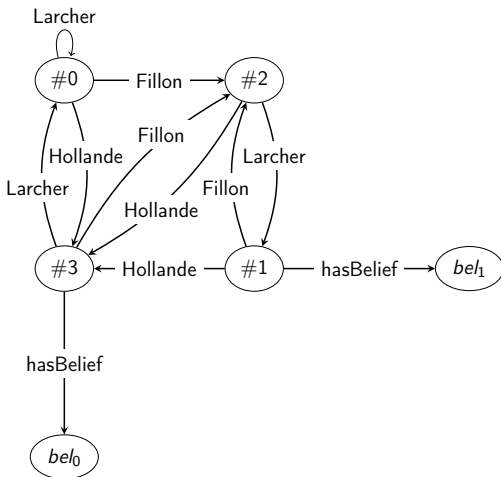
Avec les concepts introduits jusqu'ici, on peut :

- Stocker la stratification bornée ou saturée comme un graphe RDF, dans un SGBD RDF
- Traduire une requête W3Q dans SPARQL
- Les résultats ainsi obtenus sont limités par la borne k sur la longueur des chaînes d'inférence.

Une structure finie pour un modèle infini

- La structure de Kripke est introduite dans la logique épistémique (la logique des connaissances), on en utilise un fragment.
- L'intérêt de cette structure est de représenter de façon finie un graphe même infini de croyance.
- Cette structure se représente aussi en RDF ; une traduction différente des requêtes est nécessaire.

Instance de Kripke utile à notre modèle



Qui croit publiquement que M. Larcher croit en privé qu'il y a un complot contre Fillon ?

W3Q :

$Q_3(x) \leftarrow x \text{ PUBLIC}^+ \text{ M. Larcher PRIVATE}^+$

(*complot, contre, M. Fillon*)

SPARQL sur W3M :

$q_3(x) \leftarrow (v, \text{agent}, x), (v, \text{type}, \text{PublicBelief}), (v, \text{sign}, +),$

(*v, higherOrderBelief, w*), (*w, agent, Larcher*),

(*w, type, PrivateBelief*), (*w, sign, +*), (*v, fistbelief, w₂*),

(*w₂, subj, complot*), (*w₂, pred, contre*), (*w₂, obj, Fillon*), (*complot, contre, F*)

SPARQL sur W3M et la structure de Kripke :

$q_3(x) \leftarrow$

(*#0, x, ?#n1*), (*?#n1, Larcher, ?#n2*), (*#n2, hasBelief, ?bel2*),

(*?bel2, type, PrivateBelief*), (*?bel2, sign, +*), (*?bel2, fistbelief, w₂*),

(*w₂, subj, complot*), (*w₂, pred, contre*), (*w₂, obj, Fillon*), (*complot, contre, F*)

La structure de Kripke représente l'instance saturée

Theorem

Etant donné une instance de W3M \mathcal{I} et K sa structure de Kripke correspondante, et une requête de W3Q Q à évaluer sur \mathcal{I} et q la BGPQ correspondante :

① $\mathcal{I} \models_{RUTUV} q$ holds iff $\mathcal{I} \models_{RUV} q_T$ holds

\Rightarrow Une méthode concrète pour évaluer Q !

Conjecture

① $\mathcal{I} \models_{RUV} q_T$ holds iff $\mathcal{I} \cup K \models_R q_{T,K}$ holds

Contributions

- Un modèle de données W3M fondé sur RDF permettant de représenter de façon uniforme des faits (assertions RDF) et connaissances du domaine sur les faits (assertions RDFS), le **temps à l'aide d'intervalles temporel** et leurs relations (au sens d'Allen), et **les points de vue d'agents** à l'aide d'une logique épistémique.
- Un langage de requête W3Q de haut niveau permettant d'interroger une instance W3Q, se compilant en SPARQL (langage de requête pour RDF)
- Des méthodes pour stocker une (approximation d') instance W3M dans un SGBD RDF et de l'interroger via la compilation SPARQL d'une requête W3Q

Objectifs et perspectives

A court terme :

- Prouver la conjecture
- Réaliser une implémentation

A moyen terme :

- Enrichir le modèle W3M et le langage W3Q
- Optimisation de stockage et de requêtes
- Maintenance de la saturation et de la structure de Kripke
- Raisonnement pour le fact-checking

- [GBKS09] Wolfgang Gatterbauer, Magdalena Balazinska, Nodira Khoussainova, and Dan Suciu. Believe it or not : Adding belief annotations to databases. *PVLDB*, 2(1) :1–12, 2009.
- [GHV05] Claudio Gutiérrez, Carlos A. Hurtado, and Alejandro A. Vaisman. Temporal RDF. In *The Semantic Web : Research and Applications, Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29 - June 1, 2005, Proceedings*, pages 93–107, 2005.
- [HV06] Carlos A. Hurtado and Alejandro A. Vaisman. Reasoning with temporal constraints in RDF. In *Principles and Practice of Semantic Web Reasoning, 4th International Workshop, PPSWR 2006, Budva, Montenegro, June 10-11, 2006, Revised Selected Papers*, pages 164–178, 2006.
- [JC15] Wajee Jiratanachit and Suphamit Chittayasothorn. A belief-based bitemporal database model. In Ngoc Thanh Nguyen, Bogdan Trawinski, and Raymond Kosala, editors, *Intelligent Information and Database Systems - 7th Asian Conference, ACIIDS 2015, Bali, Indonesia, March 23-25, 2015, Proceedings, Part I*, volume 9011 of *Lecture Notes in Computer Science*, pages 170–178. Springer, 2015.