

# Évaluation sémantique d'informations symboliques : la cotation

Adrien Revault d'Allonnes

Lundi 11 juillet 2011

Office National d'Étude et de Recherche Aérospatiales



Laboratoire d'Informatique de Paris 6



# Motivations

**La cotation** : évaluation de la confiance en une information

Classiquement présentée comme

- ▶ qualité de l'information
- ▶ certitude de l'information
- ▶ classement relatif d'une réponse à une question

# Motivations

**La cotation** : évaluation de la confiance en une information

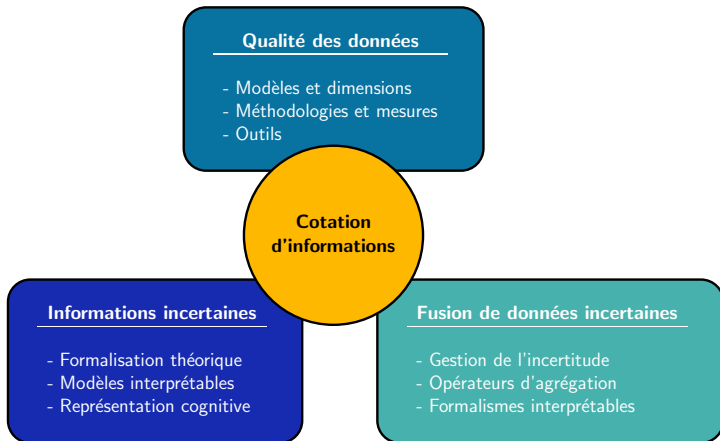
Classiquement présentée comme

- ▶ qualité de l'information
- ▶ certitude de l'information
- ▶ classement relatif d'une réponse à une question

La confiance en une information devrait représenter

- ▶ combien il est raisonnable de la croire
- ▶ la foi qu'on peut lui porter

# Domaines de recherche liés



# Objectif

## Élaboration d'une méthode d'établissement et d'expression de la confiance

qui soit à la fois

- ▶ lisible
- ▶ interprétable
- ▶ adaptable
- ▶ généralisable aux traitements automatiques
  - extraction de connaissances
  - à partir de sources textuelles

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

## Cadre existant

Cotation représentée par bigramme, par exemple A2

Fiabilité	Définition
<b>A</b>	Complètement fiable
<b>B</b>	Habituellement fiable
<b>C</b>	Plutôt fiable
<b>D</b>	Habituellement pas fiable
<b>E</b>	Pas fiable
<b>F</b>	La fiabilité ne peut être estimée

Crédibilité	Définition
<b>1</b>	Confirmée par d'autres sources
<b>2</b>	Probablement vraie
<b>3</b>	Possiblement vraie
<b>4</b>	Douteuse
<b>5</b>	Peu probable
<b>6</b>	La crédibilité ne peut être estimée

- ▶ prise en compte de la source
- ▶ granularité de la confiance
- ▶ représentation de l'inestimable



## Cadre existant

- ▶ manque de lisibilité de l'expression
  - $A2 < B1$  ?
- ▶ notions mesurées confuses
  - source fiable, sûre, experte ?
  - information vraie, probable ou confirmée ?
- ▶ incohérences dans la mesure
  - échelle fiabilité graduelle et positive
  - échelle crédibilité graduelle et positive et négative
  - combinaison peu lisible

## Approche proposée

Formaliser le processus de construction de la confiance

Cote :

expression de la confiance

Cotation :

processus d'établissement

## Approche proposée

Formaliser le processus de construction de la confiance

Cote :

expression de la confiance

- ▶ dimensions
- ▶ expression
- ▶ formalisme

Cotation :

processus d'établissement

- ▶ combinaison
  - intelligibilité
  - souplesse

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

## Sélection des dimensions

Comment établir la confiance en une information ?

*Général*

*Contextuel*

*Source*

Qui la fournit ?

Qu'en sait-il ?

*Contenu*

Est-elle concevable ?

Est-elle rapportée par ailleurs ?

## Sélection des dimensions

Comment établir la confiance en une information ?

	<i>Général</i>	<i>Contextuel</i>
<i>Source</i>	Qui la fournit ? Fiabilité	Qu'en sait-il ? Compétence
<i>Contenu</i>	Est-elle concevable ? Plausibilité	Est-elle rapportée par ailleurs ? Crédibilité

# Participation de la source à la cote

Fiabilité  
Plausibilité

Compétence  
Crédibilité

Fiabilité	Définition
1	Complètement fiable
2	Habituellement fiable
3	Moyennement fiable
4	Habituellement pas fiable
5	Pas fiable
6	La fiabilité ne peut être appréciée

- ▶ échelle discrète
- ▶ étiquettes linguistiques
- ▶ impossibilité de mesurer

## Fiabilité

- ▶ confiance faite à la source
- ▶ ne concerne que la source
  - influence égale pour toute information

# Participation de la source à la cote

Fiabilité  
Plausibilité

Compétence  
Crédibilité

Fiabilité	Définition
1	Complètement fiable
2	Habituellement fiable
3	Moyennement fiable
4	Habituellement pas fiable
5	Pas fiable
6	La fiabilité ne peut être appréciée

Compétence	Définition
1	Expert
2	Compétent
3	Partiellement compétent
4	Insuffisamment compétent
5	Incompétent
6	La compétence ne peut être appréciée

## Compétence

- ▶ connaissance de la source du domaine de l'information
- ▶ dépend du contenu
  - influence relative à l'information



# Participation du contenu à la cote

Fiabilité  
Plausibilité

Compétence  
Crédibilité

Plausibilité	Définition
1	Certaine
2	Réaliste
3	Possible
4	Peu possible
5	Impossible
6	La plausibilité ne peut être appréciée

## Plausibilité

- ▶ compatibilité à la connaissance sur le monde
- ▶ concerne le contenu
  - le modèle du monde est stable pour toute information

## Participation de l'information à la cote

Fiabilité  
PlausibilitéCompétence  
Crédibilité

Plausibilité	Définition
1	Certaine
2	Réaliste
3	Possible
4	Peu possible
5	Impossible
6	La plausibilité ne peut être appréciée

Crédibilité	Définition
1	Totalement confirmée
2	Partiellement confirmée
3	Insuffisamment confirmée
4	Partiellement contredite
5	Totalement contredite
6	La crédibilité ne peut être appréciée

## Crédibilité

- ▶ niveau de recoupement avec autres informations cotées
- ▶ concerne le contenu
  - double contextualisation : recoupement et cote de l'information comparée

# Participation de l'information à la cote

Fiabilité  
Plausibilité

Compétence  
Crédibilité

- ▶ échelle double
  - confirmation si  $\geq 3$
  - infirmation sinon
- ▶ mesure répétée

Crédibilité	Définition
1	Totalement confirmée
2	Partiellement confirmée
3	Insuffisamment confirmée
4	Partiellement contredite
5	Totalement contredite
6	La crédibilité ne peut être appréciée

## Crédibilité

- ▶ niveau de recoupement avec autres informations cotées
- ▶ concerne le contenu
  - double contextualisation : recoupement et cote de l'information comparée

## Expression de la cote

Confiance	Définition
1	Extrêmement probable
2	Probable
3	Envisageable
4	Douteuse
5	Improbable
6	La confiance ne peut être appréciée

### Cote finale

- ▶ préservation de la granularité
- ▶ préservation de l'inestimable
- ⇒ accroissement de la lisibilité

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

# Principe général du processus de cotation

## Projection de l'activation des dimensions sur la cote

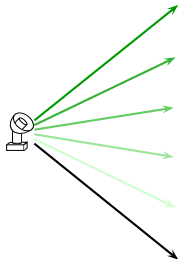
- ▶ procédé séquentiel
- ▶ ordre des dimensions

- |                                      |                 |
|--------------------------------------|-----------------|
| 1. Qui la fournit ?                  | 1. Fiabilité    |
| 2. Qu'en sait-il ?                   | 2. Compétence   |
| 3. Est-elle concevable ?             | 3. Plausibilité |
| 4. Est-elle rapportée par ailleurs ? | 4. Crédibilité  |



# Initialisation

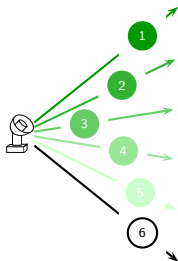
Fiabilité



Premier a priori : la confiance faite à la source

# Initialisation

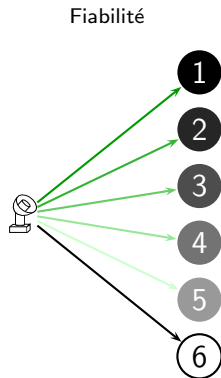
Fiabilité



Premier a priori : la confiance faite à la source

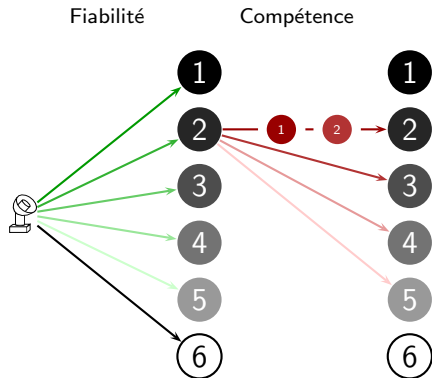


## Initialisation : projection de la fiabilité



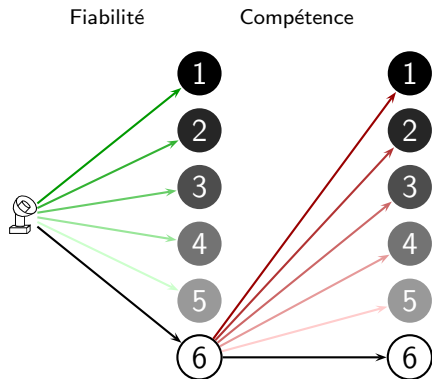
Premier a priori : la confiance faite à la source

## Projection de la compétence



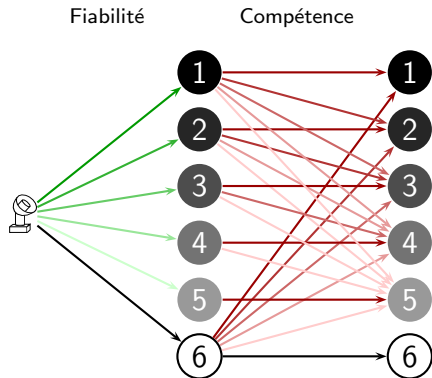
Influence à la baisse

## Projection de la compétence

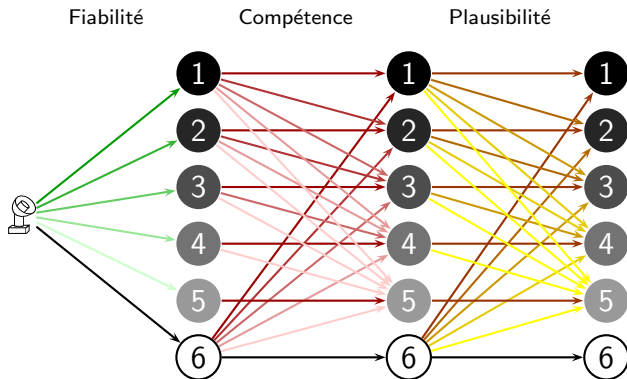


Fiabilité inconnue : initialisation reportée

## Projection de la compétence

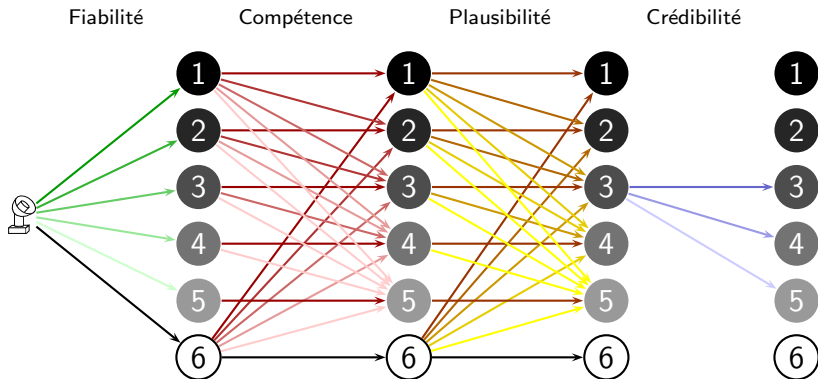


## Projection de la plausibilité



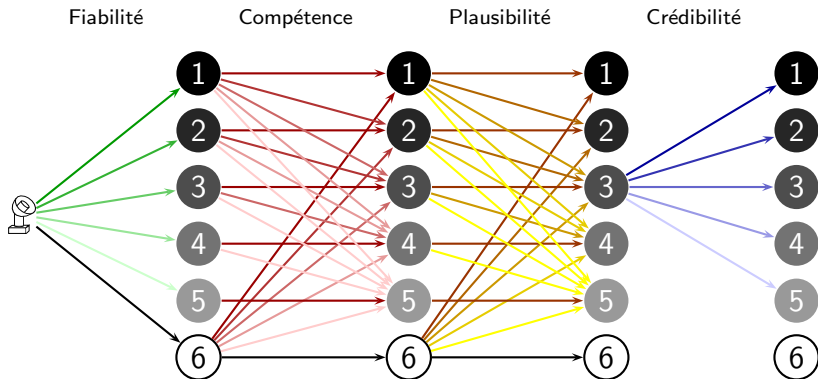
Influence à la baisse

## Projection de la crédibilité



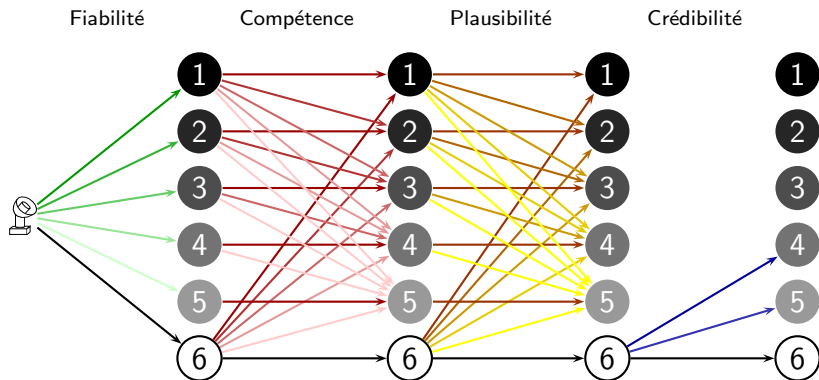
**Infirmation** : influence à la baisse

## Projection de la crédibilité



**Confirmation** : influence à la hausse

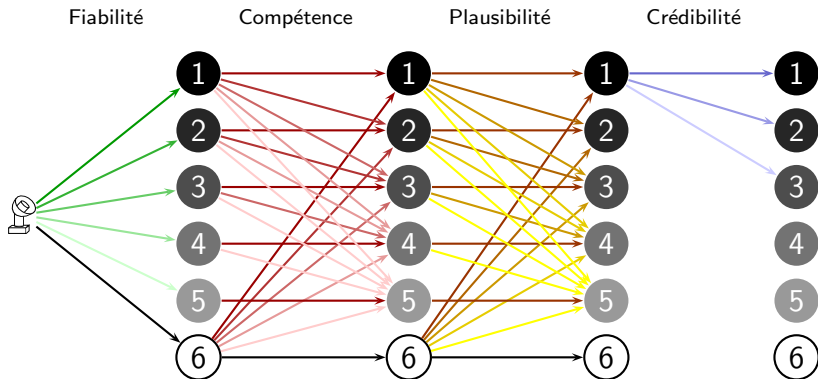
# Projection de la crédibilité



Cote toujours inestimable : initialisation reportée

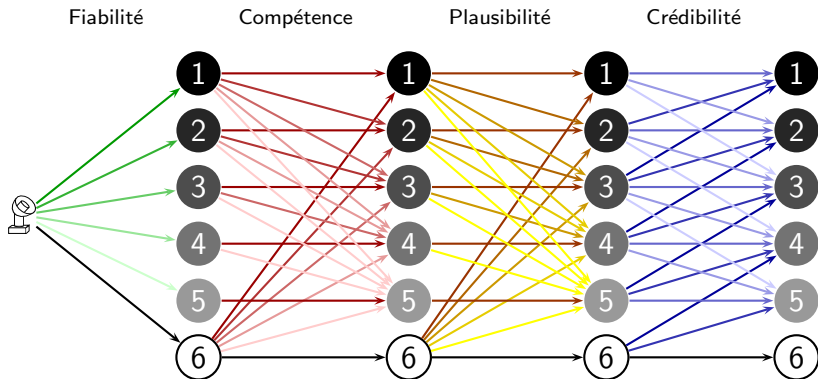


## Projection de la crédibilité



Niveaux d'activation limités

# Architecture de la cotation



## Caractéristiques du modèle

- ▶ séquentialité
- ▶ ordre
- ▶ lisibilité
- ▶ interprétabilité

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

## Modélisation d'attitudes différentes

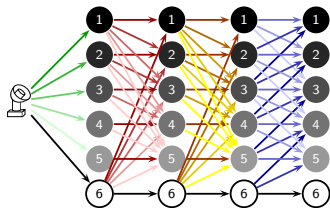
### Posture de crédulité

- ▶ attente de l'évaluateur sur le comportement du modèle
- ▶ choix a priori d'une posture
  - ⇒ représente sa sensibilité aux dimensions
  - ⇒ représente sa facilité de conviction

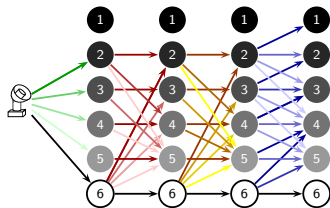
### Stratégie d'utilisation de la chaîne de cotation

- ▶ jeu d'intensités d'influence des dimensions
- ▶ pour chaque dimension
  - ⇒ résultat de la projection de son activation sur la cote
  - ⇒ intensités potentiellement différentes selon la cote

## Modélisation d'attitudes différentes

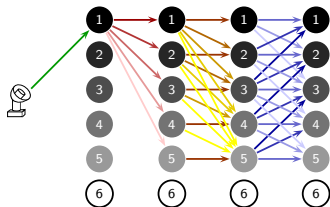


posture standard

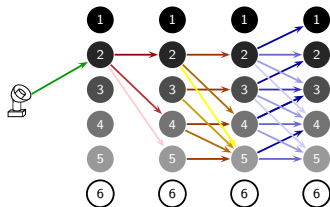


posture méfiante

## Modélisation d'attitudes différentes



posture standard

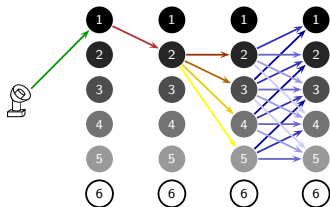


posture méfiante

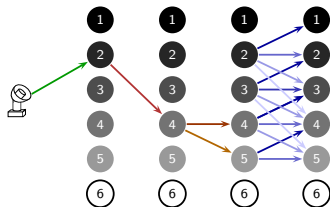
► source sûre



## Modélisation d'attitudes différentes



posture standard

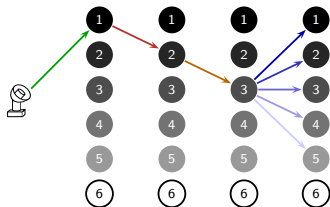


posture méfiante

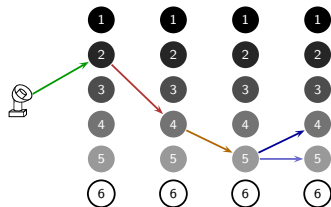
- ▶ source sûre
- ▶ information légèrement hors de son champ d'expertise



# Modélisation d'attitudes différentes



posture standard



posture méfiante

- ▶ source sûre 1
- ▶ information légèrement hors de son champ d'expertise 2
- ▶ intuitivement guère convaincante 2

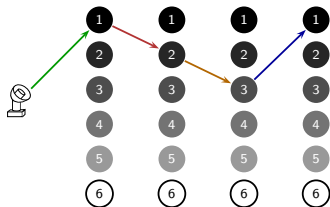
1

2

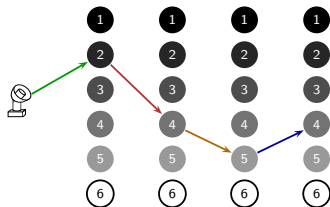
2



# Modélisation d'attitudes différentes



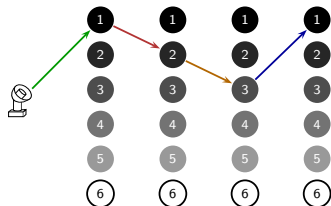
posture standard



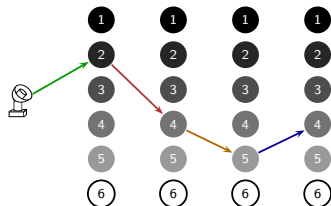
posture méfiante

- ▶ source sûre 1
- ▶ information légèrement hors de son champ d'expertise 2
- ▶ intuitivement guère convaincante 2
- ▶ absolument corroborée 1

# Modélisation d'attitudes différentes



posture standard



posture méfiante

- ▶ source sûre 1
  - ▶ information légèrement hors de son champ d'expertise 2
  - ▶ intuitivement guère convaincante 2
  - ▶ absolument corroborée 1
- ⇒ cote finale

standard

1

méfiante

4

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

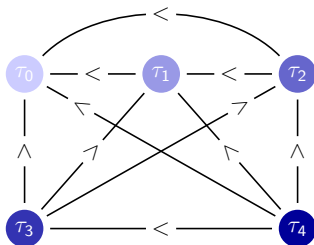
Extraction et cotation de réseau social

## Conclusions et perspectives

## Cadre formel retenu : la logique multivalente

- ▶ granularité de la vérité :  $\mathcal{L}_M = \{\tau_0, \dots, \tau_{M-1}\}$ 
  - ▷ logique à cinq valeurs :  $\mathcal{L}_5$
- ▶ comparabilité des degrés :  $\alpha < \beta \Leftrightarrow \tau_\alpha < \tau_\beta$
- ▶ étiquettes sémantiques

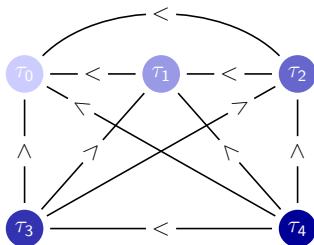
Degré	Étiquette sémantique
$\tau_4$	Vrai
$\tau_3$	Plutôt vrai
$\tau_2$	Moyennement vrai
$\tau_1$	Plutôt faux
$\tau_0$	Faux



## Cadre formel retenu : la logique multivalente

- ▶ granularité de la vérité :  $\mathcal{L}_M = \{\tau_0, \dots, \tau_{M-1}\}$ 
  - ▷ logique à cinq valeurs :  $\mathcal{L}_5$
- ▶ comparabilité des degrés :  $\alpha < \beta \Leftrightarrow \tau_\alpha < \tau_\beta$
- ▶ étiquettes sémantiques
- ▶ inconvénient : incertain  $\Leftrightarrow$  inconnu  $\Leftrightarrow \tau_{\frac{M-1}{2}}$

Degré	Étiquette sémantique
$\tau_4$	Vrai
$\tau_3$	Plutôt vrai
$\tau_2$	Moyennement vrai
$\tau_1$	Plutôt faux
$\tau_0$	Faux



## Proposition de modélisation de l'inconnu

Introduction d'un degré supplémentaire  $\tau_?$  :

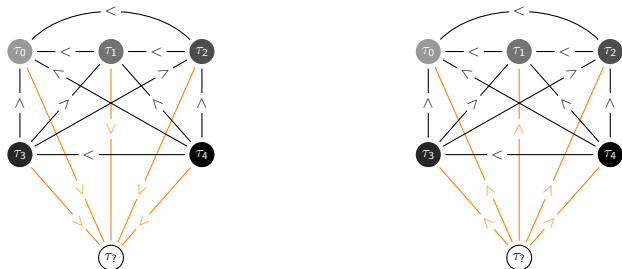
$$\text{incertain} : \tau_{\frac{M-1}{2}} \neq \tau_? : \text{inconnu}$$

- ▶ hors échelle
  - $\tau_? \notin \mathcal{L}_M$
- ▶ élément neutre pour tout opérateur logique
  - $\forall \tau_\alpha \in \mathcal{L}_M, \tau_? \odot \tau_\alpha = \tau_\alpha$
- ▶ comportement involutif
  - $\neg \tau_? = \tau_?$
  - $\tau_? \odot \tau_? = \tau_?$

⇒ Cadre étendu  $\mathcal{L}_M^e = \mathcal{L}_M \cup \{\tau_?\}$

## Propriétés de $\tau_?$

- ▶ perte de la comparabilité dans  $\mathcal{L}_M^e$ 
  - ▷ mais comparaison de  $\tau_?$  à  $\tau_\alpha$  inutile



- ▶ gain d'expressivité : comparabilité renforcée dans  $\mathcal{L}_M$
- ▶ maintien de la cohérence

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

**Combinaisons à la baisse : définition d'opérateurs**

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

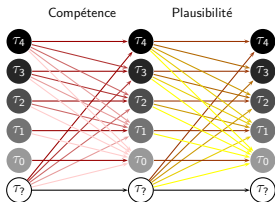
La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives



# Formalisation de la compétence et de la plausibilité



$$F : \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$$

$$(\tau_\alpha, \tau_\beta) \mapsto F(\tau_\alpha, \tau_\beta)$$

- ▷  $\tau_\alpha$  : cote courante
- ▷  $\tau_\beta$  : activation
- ▷  $F(\tau_\alpha, \tau_\beta)$  : cote mise à jour

## ► tendance à la baisse

- $F(\tau_\alpha, \tau_\beta) \leq \tau_\alpha$

## ► croissant en ses deux arguments

- Si  $\tau_\alpha \leq \tau_{\alpha'}$ , alors  $F(\tau_\alpha, \tau_\beta) \leq F(\tau_{\alpha'}, \tau_\beta)$

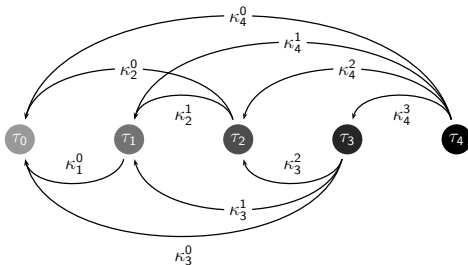
- Si  $\tau_\beta \leq \tau_{\beta'}$ , alors  $F(\tau_\alpha, \tau_\beta) \leq F(\tau_\alpha, \tau_{\beta'})$

## ► $\tau?$ élément neutre

- $F(\tau?, \tau_\alpha) = F(\tau_\alpha, \tau?) = \tau_\alpha$

## Opérateur à la baisse proposé

$$F(\tau_\alpha, \tau_\beta) = \min\{\tau_\gamma \in \mathcal{L}_M \mid \tau_\beta \leq \kappa_\alpha^\gamma\}$$



- ▶ paramétrée par les  $\kappa_\alpha^\beta$ 
  - conditionnant les transitions entre degrés de  $\mathcal{L}_M$
  - soumis à des contraintes d'ordre

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

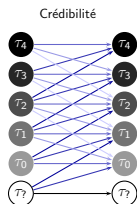
## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

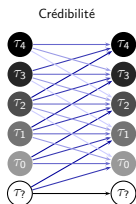
## Conclusions et perspectives

## Formalisation de la crédibilité



- ▶ comportement double
  - à la baisse si infirmation
  - à la hausse si confirmation
- ▶ confrontations d'informations cotées
  - intégration de leur cote
  - mesure répétée : 2 stratégies d'intégration
    - ▶ à la volée : crédibilité immédiate
    - ▶ avec mémoire : crédibilité cumulée

# Crédibilité immédiate



$$G : \mathcal{L}_M^e \times \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$$

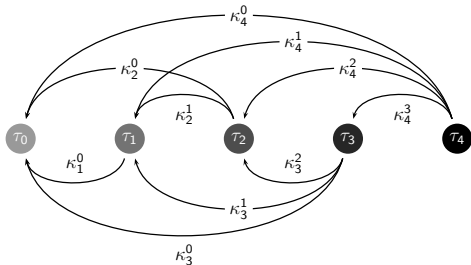
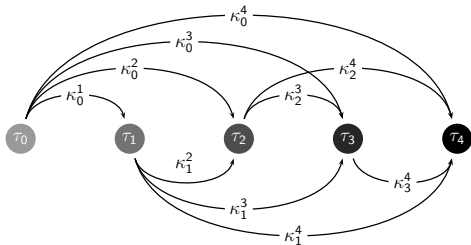
$$(\tau_\alpha, \tau_\beta, \tau_\gamma) \mapsto G(\tau_\alpha, \tau_\beta, \tau_\gamma)$$

- ▶  $\tau_\alpha$  : cote courante
  - ▶  $\tau_\beta$  : activation
  - ▶  $\tau_\gamma$  : cote information comparée
  - ▶  $G(\tau_\alpha, \tau_\beta, \tau_\gamma)$  : cote mise à jour
- } crédibilité pondérée

Définie grâce à deux opérateurs :

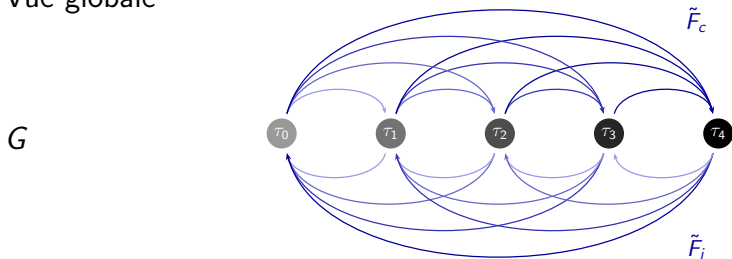
- ▶ à la baisse pour les infirmations :  $\tau_\beta < \tau_{\frac{M-1}{2}}$ 
  - $\tilde{F}_i : \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$
- ▶ à la hausse pour les confirmations :  $\tau_\beta \geq \tau_{\frac{M-1}{2}}$ 
  - $\tilde{F}_c : \mathcal{L}_M^e \times \mathcal{L}_M^e \longrightarrow \mathcal{L}_M^e$

# Opérateur de crédibilité immédiate

 $\tilde{F}_i$ 

 $\tilde{F}_c$ 


# Opérateur de crédibilité immédiate

Vue globale

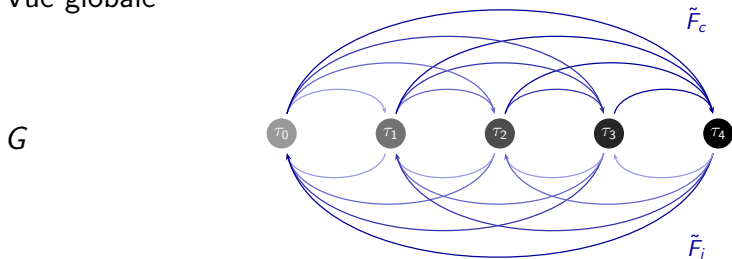


Caractéristiques :

- ▶ pas nécessairement symétrique
- ▶ pas nécessairement associatif

# Opérateur de crédibilité immédiate

Vue globale



Caractéristiques :

- ▶ pas nécessairement symétrique
  - ▶ pas nécessairement associatif
  - ▶  $G$  ignore les successions de recoupements mal cotés
- ⇒ opérateur de crédibilité cumulée



## Crédibilité cumulée

Prise en compte de la **dynamique de la conviction**

- ▶ Ajout d'un effet mémoire à  $G$ 
  - souvenir des confirmations :  $g_+$
  - souvenir des infirmations :  $g_-$

La cote évolue dès que  $g_+$  ou  $g_-$  dépasse un seuil de transition

## Crédibilité cumulée

Prise en compte de la **dynamique de la conviction**

- ▶ Ajout d'un effet mémoire à  $G$ 
  - souvenir des confirmations :  $g_+$
  - souvenir des infirmations :  $g_-$

La cote évolue dès que  $g_+$  ou  $g_-$  dépasse un seuil de transition

Mise à jour de  $g_+$  et  $g_-$

- ▶ réinitialisation à  $\tau?$
- ▶ conservation des tendances
  - résiduelles dans le sens de la dernière évolution
  - acquises dans le sens opposé

## Crédibilité cumulée

Prise en compte de la **dynamique de la conviction**

- ▶ Ajout d'un effet mémoire à  $G$ 
  - souvenir des confirmations :  $g_+$
  - souvenir des infirmations :  $g_-$

La cote évolue dès que  $g_+$  ou  $g_-$  dépasse un seuil de transition

Mise à jour de  $g_+$  et  $g_-$

- ▶ réinitialisation à  $\tau?$
- ▶ conservation des tendances
  - résiduelles dans le sens de la dernière évolution
  - acquises dans le sens opposé

⇒ **Proposition de l'algorithme  $G^*$**

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

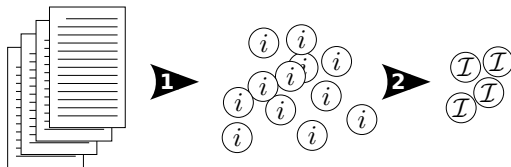
La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

## Cadre applicatif considéré

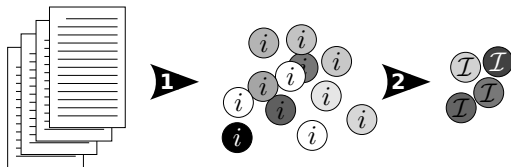
### Extraction de connaissances



- ▶ extraction automatique à partir de données textuelles
- ▶ extraction d'informations atomiques
- ▶ fusion des informations atomiques  
⇒ création de connaissances

## Cadre applicatif considéré

Extraction *et cotation* de connaissances

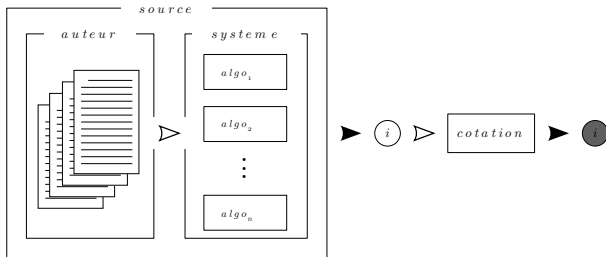


- ▶ extraction automatique à partir de données textuelles
- ▶ extraction *et cotation* d'informations atomiques
- ▶ fusion des informations atomiques  
⇒ création *et cotation* de connaissances

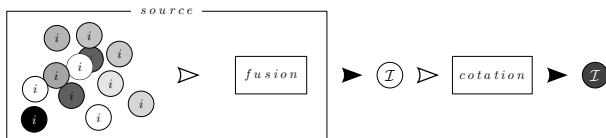
⇒ Double chaîne de cotation

## Double chaîne de cotation

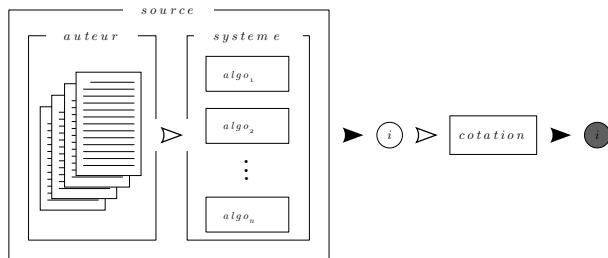
### ► informations atomiques



### ► connaissances



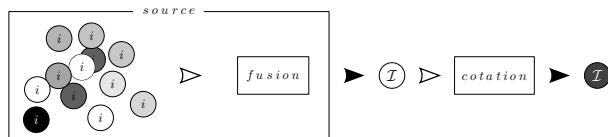
# Transposition à l'extraction de connaissances



- ▶ informations atomiques
  - source = (auteur, système d'extraction)
    - ▶ fiabilité = celle de l'auteur
    - ▶ compétence = performance de l'algorithme
  - contenu
    - ▶ plausibilité indéfinie
    - ▶ crédibilité = degré de recoupement



# Transposition à l'extraction de connaissances



- ▶ connaissances
  - source = (informations atomiques, système de fusion)
    - ▶ fiabilité indéfinie
    - ▶ compétence = performance des règles de fusion
  - contenu
    - ▶ plausibilité = conformité à l'état du monde
    - ▶ crédibilité indéfinie : pas de recoupements

# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

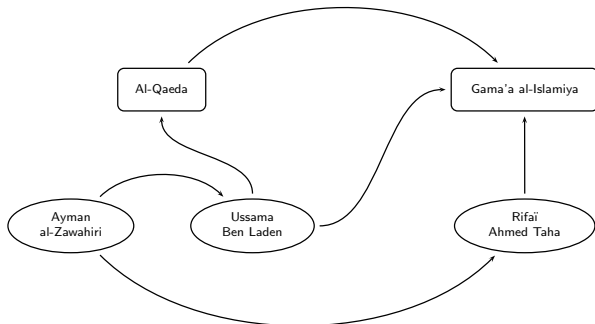
## Conclusions et perspectives

## Mise en œuvre

- ▶ corpus de textes étudiés
  - ‘Guerre secrète contre Al-Qaeda’ dirigé par Éric Denécé
  - version avant relecture de l’éditeur
  - problèmes de fautes de frappe, incohérences de notations
  
- ▶ récolte des besoins utilisateurs
  - extraire des connaissances présentes dans l’ouvrage
  - retrouver les relations décrites entre personnes et organisations

⇒ reconstitution et cotation d’un réseau social

## Extraction de réseau social



► acteurs

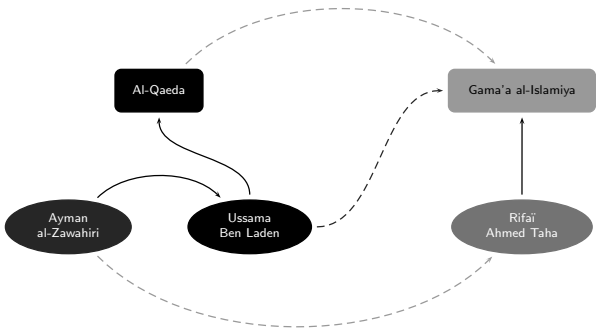
- personnes
- organisations

► attributs

► relations

- personne-personne
- personne-organisation
- organisation-organisation

# Extraction et cotation de réseau social



▶ acteurs

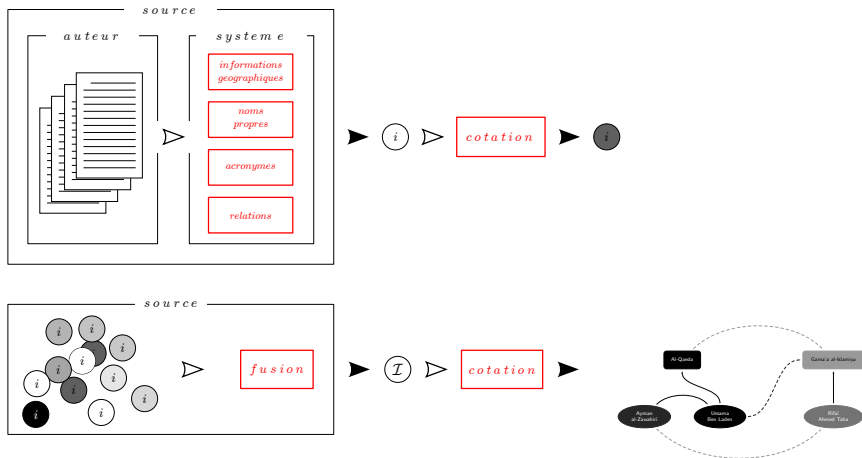
- personnes
- organisations

▶ attributs

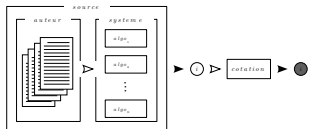
▶ relations

- personne-personne
- personne-organisation
- organisation-organisation

# Extraction et cotation de réseau social



## Extraction d'informations atomiques



### ► 4 modules d'extraction

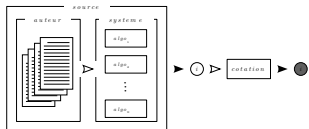
- informations géographiques
- noms propres
- résolution d'acronymes
- relations

⇒ nationalité

⇒ nom, prénom

⇒ organisations

# Extraction d'informations atomiques

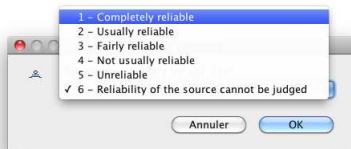
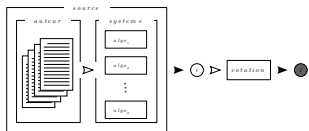


- ▶ 4 modules d'extraction
  - informations géographiques
  - noms propres
  - résolution d'acronymes
  - **relations**

Patron	Type 1	Type 2
sous-groupe de	'Org'	'Organisation'
chef de	'Person'	'Organisation'
membre de	'Person'	'Organisation'
bras droit de	'Person'	'Person'
alias	'Person'	'Person'

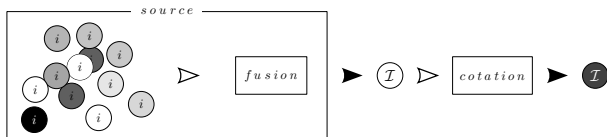


# Cotation d'informations atomiques



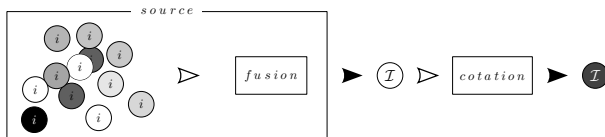
- ▶ fiabilité de la source
  - interface utilisateur
- ▶ compétence des modules d'extraction
  - informations géographiques  $\mathcal{T}_4$
  - noms propres  $\mathcal{T}_3$
  - résolution d'acronymes  $\mathcal{T}_2$
  - relations  $\mathcal{T}_4$
- ▶ crédibilité de l'information
  - mesure de recoupement des chaînes de caractères

## Extraction de connaissances enrichies



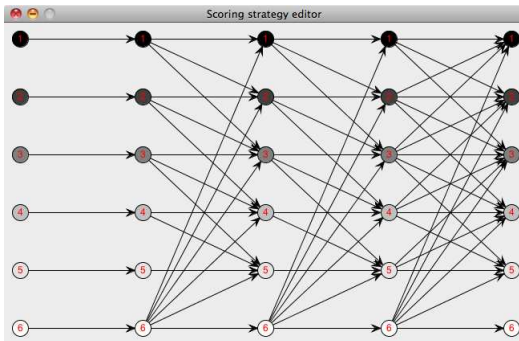
- ▶ fusion d'informations atomiques traitant d'un même sujet
  - descriptions plus informatives
  - personnes : nom, prénom, nationalité, alias
  - organisations : acronyme, nom complet, nationalité
  
- ▶ heuristiques de fusion par des règles
  - identification de toutes les dénominations d'un acteur
  - construction de la connaissance enrichie par rattachement des informations

## Cotation de connaissances enrichies



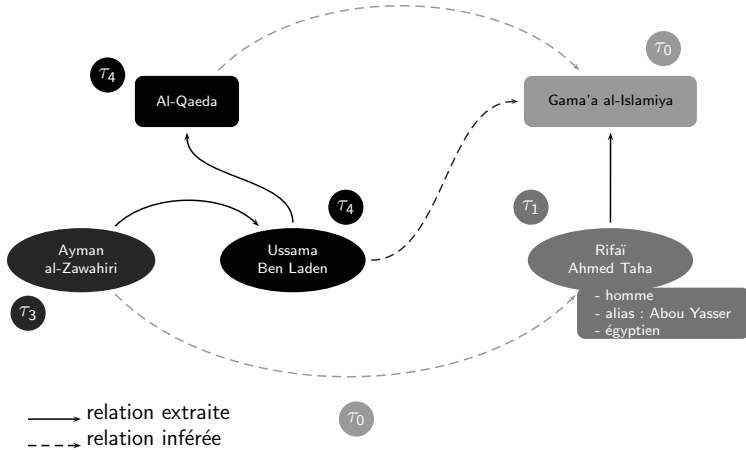
- ▶ compétence du module de fusion :  $\mathcal{T}_3$
- ▶ plausibilité des connaissances
  - richesse des descriptions et des liens
  - compatibilité avec les règles de cohérence

## Sélection de la posture de crédulité



- ▶ éditeur de stratégies
  - spécification du jeu d'intensités
  - 3 stratégies prédéfinies

## Résultat illustratif



# Organisation de la présentation

## Évaluation de la confiance

Dimensions : composantes de la cote

Combinaison : architecture de la cotation

Personnalisation : adaptabilité du modèle

## Formalisation de la confiance

Cadre théorique : extension de la logique multivalente

Combinaisons à la baisse : définition d'opérateurs

Combinaisons de recoupement : opérateurs et stratégies

## Application de la confiance

La cotation dans l'extraction de connaissances

Extraction et cotation de réseau social

## Conclusions et perspectives

# Conclusions

## Proposition d'un modèle d'expression et d'acquisition de la confiance

- ▶ possédant les propriétés de :
  - lisibilité
  - facilité d'interprétation
  - possibilité de justification
  - adaptabilité
  
- ▶ trois niveaux d'étude
  - conceptuel
  - formel
  - pratique

# Conclusions

## Analyse conceptuelle

- ▶ distinction entre cote et cotation ⇒ lisibilité
- ▶ choix de représentation ⇒ lisibilité
  - échelles discrètes munies d'étiquettes linguistiques
- ▶ identification des dimensions y participant
  - source/contenu vs. générale/contextuelle
  - indépendance et non-redondance ⇒ interprétabilité
- ▶ architecture de combinaison ⇒ justification
  - projection de l'activation des dimensions sur la cote
  - postures de crédulité ⇒ adaptabilité



# Conclusions

## Formalisation théorique

- ▶ extension de la logique multivalente
  - introduction de  $\tau?$  ⇒ interprétabilité
  - étude des propriétés de  $\tau?$
  
- ▶ proposition d'opérateurs de combinaison sémantiques
  - 2 tendances : à la baisse ou variable
  - 2 effets : immédiat ou cumulé
  - paramétrisation ⇒ souplesse
  
- ▶ modélisation des propositions conceptuelles

# Conclusions

## Mise en pratique

- ▶ cotation de connaissances extraites à partir de textes
  - formulation d'une double chaîne de cotation
  - transposition des dimensions
- ▶ étude de l'extraction et la cotation de réseau social
  - multimodalité de la cotation : acteurs et relations
- ▶ réalisation d'un démonstrateur
  - modules d'extraction d'informations
  - implémentation du modèle de cotation
- ▶ application à des données réelles

## Perspectives pratiques

### La cotation dans les réseaux

- ▶ présentation de la cotation
  - visualisation graphique de la confiance
- ▶ exploitation de la cotation
  - utilisation de la confiance dans l'étude de graphes

### Validation de l'implémentation

- ▶ utilisation d'une échelle à six niveaux  $\mathcal{L}_5^e$ 
  - compromis expressivité vs. complexité
- ▶ utilisation d'une échelle unique
  - évaluation de la crédibilité dans  $\mathcal{L}_9^e$
- ▶ validation des postures de crédulité

# Perspectives conceptuelles

## Enrichissement des dimensions

- ▶ fiabilité de la source
  - mise à jour dynamique
  - définition dans le cas de discours rapporté
  
- ▶ plausibilité de l'information
  - raisonnement ontologique pour inférer la compatibilité
  
- ▶ crédibilité de l'information
  - comparaison du contenu informationnel
  - indépendance des sources
  - partition amicale des sources

## Perspectives théoriques

### Psychologie du procédé

- ▶ caractérisation cognitive
  - postures de crédulité
  - dynamique de la conviction
- ▶ théorisation sémantique

### Extensions formelles

- ▶ généralisation des opérateurs
  - placement dans la logique multivalente
  - famille d'opérateurs paramétriques
- ▶ extension de l'incertitude multivalente
  - dans le cadre du raisonnement
  - à d'autres logiques non-classiques

# Évaluation sémantique d'informations symboliques : la cotation

Adrien Revault d'Allonnes

Lundi 11 juillet 2011

Office National d'Étude et de Recherche Aérospatiales



Laboratoire d'Informatique de Paris 6



## Publications

- ▶ Incertain et inconnu, deux facettes de la cotation, **Adrien Revault d'Allonnes**, Herman Akdag, Bernadette Bouchon-Meunier, COTA @ IC'2010 : Cotation des Informations : Théorie et Applications, atelier des 21èmes Journées francophones d'Ingénierie des Connaissances, Alès, juin 2010, pp. 99-103
- ▶ For a Data-Driven Interpretation of Rules wrt GMP Conclusions in Abductive Problems, **Adrien Revault d'Allonnes**, Herman Akdag, Bernadette Bouchon-Meunier, Journal of Uncertain Systems, vol 3, nb 4, pp. 280-297, World Academic Press, novembre 2009
- ▶ Critères d'évaluation contextuelle pour le traitement automatique, **Adrien Revault d'Allonnes**, Jérôme Besombes, QDC'09 : Qualité des Données et des Connaissances, Atelier des 9èmes Journées Francophones 'Extraction et Gestion des Connaissances', Strasbourg, janvier 2009, pp. A6 13-20
- ▶ An Extension of STANAG2022 for Information Scoring, Jérôme Besombes, **Adrien Revault d'Allonnes**, Fusion'08 : 11th International Conference on Information Fusion, Cologne - Allemagne, juillet 2008, pp. 1635-1641
- ▶ Trust-moderated information-likelihood. A multi-valued logics approach, **Adrien Revault d'Allonnes**, Herman Akdag, Olivier Poirel, CiE 2007 : Computability in Europe - Computation and Logic in the Real World, Sienna - Italie, juin 2007, pp. 1-6
- ▶ Selecting Implications in Fuzzy Abductive Problems, **Adrien Revault d'Allonnes**, Herman Akdag, Bernadette Bouchon-Meunier, FOCI 2007 : IEEE Symposium on Foundations of Computational Intelligence, Honolulu - Hawaï, avril 2007, pp. 597-602
- ▶ Vers une classification de problèmes abductifs en fonction d'observations possibles, **Adrien Revault d'Allonnes**, Herman Akdag, Bernadette Bouchon-Meunier, LFA 2006 : Rencontres francophones sur la Logique Floue et ses Applications, Toulouse, octobre 2006, Cépaduès, pp. 121-128