



**HAL**  
open science

## Critères d'évaluation contextuelle pour le traitement automatique

Adrien Revault d'Allonnes, Jérôme Besombes

► **To cite this version:**

Adrien Revault d'Allonnes, Jérôme Besombes. Critères d'évaluation contextuelle pour le traitement automatique. 9èmes Journées Francophones "Extraction et Gestion des Connaissances", Jan 2009, Strasbourg, France. pp.A6 13-20. hal-00600715

**HAL Id: hal-00600715**

**<https://hal.science/hal-00600715>**

Submitted on 6 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Critères d'évaluation contextuelle pour le traitement automatique

Adrien Revault d'Allonnes\*, Jérôme Besombes\*\*

\*Laboratoire d'Informatique de Paris VI – LIP6  
104 avenue du président Kennedy, 75016 Paris,  
Adrien.Revault-d'Allonnes@lip6.fr,

\*\*Office National d'Études et de Recherche Aérospatiales – ONERA  
ONERA - BP72, 29 avenue de la Division Leclerc, 92322 Châtillon Cx  
Jerome.Besombes@onera.fr

**Résumé.** Les traitements automatiques de gros volumes de données deviennent de plus en plus incontournables dans tous les domaines d'activité. Certains domaines sont, cependant, plus sensibles à la qualité des données produites. Si les mesures statistiques classiques offrent une bonne indication des performances d'algorithmes, elles qualifient rarement les données produites proprement dites. Nos travaux proposent des critères d'évaluation de ces données et les intègrent à un procédé de qualification de la qualité d'une information. Nous montrons comment intégrer ces différents paramètres à un système d'extraction automatique et comment celui-ci va évaluer sa production à la volée. Ce score de qualité servira à assister l'expert chargé de la vérification des données soit en tant qu'index de classement des informations en fonction de leur crédibilité, soit en mettant en avant les informations douteuses à vérifier en priorité. Afin de fournir le score le plus exhaustif et le plus interprétable possible, nous considérerons différents paramètres : la confiance faite à la source de l'information, la capacité de cette source à produire l'information considérée et l'information elle-même lorsque confrontée tant à des connaissances expertes qu'à d'autres informations produites.

## 1 Introduction

Les évolutions récentes des attributions des forces militaires, comme l'évolution de leurs contextes d'utilisation (e.g. opérations de maintien de la paix, guerres asymétriques, lutte contre le terrorisme, . . .), offrent de nombreux défis pour la conception et la mise-en-œuvre de systèmes d'informations dédiés. Les volumes d'informations disponibles, en augmentation continue, rendent le besoin de systèmes automatisés incontournable. Néanmoins, de la sensibilité des informations manipulées, et de leurs conséquences, découle un besoin de manipuler des informations sûres et donc une méfiance envers les systèmes totalement automatisés. Cet inconvénient peut être minimisé en associant automatiquement un score de qualité à chaque information extraite ou produite. On pourra, par la suite, utiliser ce score pour ordonner les informations, voire fixer des seuils d'alertes sous lesquels il sera nécessaire d'effectuer des vérifications, avant toute prise de décision.

Pour fournir le score le plus exhaustif et le plus interprétable possible, celui-ci devra tenir compte de différents paramètres : la confiance faite à la source, la capacité de cette source à produire l'information considérée et l'information elle-même lorsque confrontée tant à des connaissances expertes qu'à d'autres informations produites.

## 2 Évolution de la proposition

L'annexe du STANAG2022 OTAN (1997) a défini pour la première fois la grille de cotation du renseignement pour les forces de l'OTAN. Elle proposait une échelle à deux dimensions et six degrés de vérité (cf. tableau 1), dans laquelle la qualité d'une information était représentée par une combinaison de sa plausibilité et de la fiabilité de sa source. La fiabilité caractérise donc la source, indépendamment de l'information considérée. Toute information fournie par une même source bénéficierait donc de la même fiabilité.

Le second axe de cotation du STANAG2022 évalue la certitude de l'information. Le degré supérieur de cet axe est affecté aux informations confirmées par d'autres sources. Ceci tend à suggérer que la certitude de l'information se mesure comme un degré de confirmation : plus une information est confirmée, plus elle est certaine et, à l'inverse, plus elle est infirmée, moins elle est crédible.

De cette échelle de lecture nous retenons au moins les deux critiques suivantes :

Fiabilité	Définition	Crédibilité	Définition
F	Fiabilité ne peut être estimée	6	Crédibilité ne peut être estimée
E	Pas fiable	5	Peu probable
D	Habituellement pas fiable	4	Douteuse
C	Plutôt fiable	3	Possiblement vraie
B	Habituellement fiable	2	Probablement vraie
A	Complètement fiable	1	Confirmée par d'autres sources

TAB. 1 – STANAG2022 : Fiabilité de la source et crédibilité de l'information

1. L'évaluation sur deux axes, au lieu d'augmenter la lisibilité du score, le rend plus obscur. En effet, quelle information est la plus probable si l'une est cotée B3 et l'autre C2 ? Si l'on ajoute à ceci que la plausibilité découle de la confirmation par des sources *distinctes*, la simple lecture de la fiabilité devient problématique.
2. Comme nous l'avons dit, les critères d'évaluation de la plausibilité d'une information se réduisent à son degré de confirmation / infirmation. D'autres critères seraient sans doute indiqués pour mieux estimer la qualité d'une information. Mathews et al. (2005) présentent un point de vue intéressant sur les 'pédigrées' d'informations. Le contexte de production de deux informations par la même source permettrait, par exemple, de contrebalancer la fiabilité de la source et introduirait donc de la flexibilité dans son impact sur sa production.

Ces remarques peuvent paraître, de prime abord, contradictoires. D'un côté, nous voulons offrir à l'utilisateur un score unique exprimant la qualité d'une information et ainsi la confiance qu'il doit avoir en elle. En particulier, ceci simplifierait les comparaisons entre informations, ce qui serait une nette amélioration de l'existant. D'autre part, nous souhaitons conserver les critères d'évaluation existants et en ajouter des nouveaux ou, du moins, clarifier les cas où les domaines de définition étaient mal définis ou confus.

Pour parer à ces problèmes nous allons :

- définir deux critères additionnels : la compétence de la source et la plausibilité de l'information
- redéfinir les critères existants
- intégrer chaque critère à une chaîne de calcul qui définira un protocole général pour le calcul d'un score unique

### 3 Définition de critères

Nous allons maintenant donner notre point de vue sur les critères existants et introduire les critères additionnels que nous proposons d'intégrer à notre score de qualité.

Pour nous conformer à l'esprit d'origine, nous avons choisi d'exprimer les valeurs de vérité de chaque critère sur une échelle à six niveaux, mais ce choix n'est en rien limitatif.

#### 3.1 Fiabilité

La fiabilité de la source est toujours définie selon la grille initiale (cf. tableau 1). Néanmoins, nous limitons maintenant l'étendue de ce que ce critère juge, cf. Rogova et Nimier (2004). Dans la documentation clairssemée traitant du sujet (l'AJP-2.1 OTAN (2002)) la fiabilité de la source ne se contente pas de représenter la confiance faite à la source, mais également des notions de la compétence de la source à fournir l'information ainsi que la confiance exprimée par cette source pour son information.

Nous définissons la fiabilité comme totalement indépendante de l'information. De cette manière cet 'avis' aura le même impact sur toutes les informations provenant d'une même source. Bien entendu, la fiabilité de la source n'est pas un critère définitif. Le temps et l'expérience pourront la faire varier, comme illustré dans la figure 3. Une source fiable donnant systématiquement des informations fausses perdra de la confiance et, de la même manière, une source inconnue se verra réévaluée à l'aune de sa production. Néanmoins, à un instant donné, toutes les informations produites par une même source bénéficieront du même taux de fiabilité.

Supposons, par exemple, que notre source soit un capteur électronique. Nous prétendons que la fiabilité de la source correspond à l'état du capteur. De cette manière quelle que soit l'information produite, la fiabilité de la source aura la même influence sur le calcul de sa qualité.

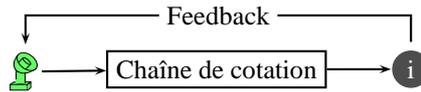


FIG. 1 – Mise à jour de la fiabilité de la source

### 3.2 Compétence

La capacité qu'une source a à fournir une information peut dépendre de l'information elle-même. Il est possible qu'une source donne une information hors de son domaine d'expertise. Une telle information ne doit pas, pour autant, être rejetée pour cette unique raison, mais le destinataire doit en tenir compte lors de la classification de l'information ou lors de l'évaluation de sa qualité.

Si nous revenons à notre exemple, un capteur électronique est étalonné pour fonctionner sur une certaine plage. Ceci n'empêche pas ce capteur de fournir des informations en dehors de cette plage. Dans un tel cas de figure, la qualité devrait intégrer cette estimation de la compétence pour pondérer la confiance.

Une autre interprétation est le contexte d'utilisation. Prenons, par exemple, les conditions météorologiques. Ces conditions vont impacter toutes les observations faites pendant qu'elles s'appliqueront.

Nous définissons donc la compétence de la source comme une fonction de la source en regard de l'information considérée, et mesurée sur l'échelle donnée par le tableau 2.

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

TAB. 2 – Compétence de la source, fonction de l'information considérée

### 3.3 Plausibilité

Supposons maintenant que nous disposons de connaissances extérieures, par exemple stockées dans une ontologie. Nous désignons par connaissances extérieures notre connaissance sur l'état du monde, indépendamment de nos informations extraites. Si une information est produite en désaccord avec ces connaissances extérieures, nous prétendons que cette contradiction est différente d'une contradiction avec des informations construites. Il est vraisemblable qu'un doute en l'information émerge, dans la mesure où notre connaissance du monde est supposée sûre.

Si notre capteur est, par exemple, un radar de vitesse braqué sur une route. En connaissant la route et quelques rudiments des lois de la physique, il est possible de contredire certaines mesures. Ceci est différent de la comparaison de mesures, parce que notre connaissance du monde étant certaine, elle ne sera pas remise en cause par les informations que nous évaluons. En revanche, les informations acquises peuvent – et doivent – être réévaluées.

La plausibilité mesure donc l'adéquation entre l'information et notre point de vue sur l'état du monde. Un exemple de degrés de vérité pour ce critère est présenté dans le tableau 3.

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

TAB. 3 – Plausibilité de l'information : conformité à la connaissance extérieure

### 3.4 Crédibilité

En raison du degré maximal de crédibilité de la grille de cotation existante, nous choisissons de l'évaluer comme un indice de confirmation par les informations acquises. Une fois encore, nous insistons sur la distinction faite entre les comparaisons entre informations acquises et leur compatibilité avec le modèle de raisonnement. Les informations que nous comparons en évaluant la crédibilité sont toutes construites et ont donc toutes un indice de qualité mouvant, par opposition aux connaissances du monde dont la confiance est fixe et maximale. Le tableau 4 présente les degrés de vérité de la crédibilité.

Crédibilité	Définition
6	La crédibilité ne peut être appréciée
5	Contredite par des informations fiables
4	Partiellement contredite
3	Insuffisamment confirmée
2	Partiellement confirmée
1	Confirmée par des informations fiables

TAB. 4 – Degrés de crédibilité : indice de confirmation entre informations extraites

### 3.5 Qualité

En sortie, notre système proposera un degré de qualité associé à l'information considérée. Celui-ci exprime, en un score unique et comparable, la combinaison des critères ci-dessus. Un exemple de degrés de vérité associés à la qualité est présenté dans le tableau 5.

Confiance	Définition
6	La qualité ne peut être appréciée
5	Peu probable
4	Douteuse
3	Possible
2	Probable
1	Extrêmement probable

TAB. 5 – Sortie du système : degrés de qualité associé à une information

## 4 La chaîne de cotation

Maintenant que nous avons défini les critères que nous souhaitons inclure dans notre méthode d'évaluation, nous allons préciser les tendances d'impact de chaque facteur sur l'évolution de la qualité de l'information et présenter notre point de vue sur leur utilisation.

### 4.1 Présentation et origine

Nous représentons, dans la figure 1, le processus de cotation de manière séquentielle. L'ordre d'intégration des différents critères provient d'une interprétation naturelle et intuitive du procédé partant de l'évaluation de la source, suivie d'un critère traitant source *et* information et se concluant par les confrontations de l'information aux autres connaissances. La figure 1 indique également les pentes générales de la qualité, pour chaque critère. Chaque étape consiste en la réévaluation du score précédent – celui de l'étape précédente – lorsque confrontée au nouveau critère. Nous présenterons également l'idée de stratégies d'évaluation permettant de modéliser différentes postures de crédulité de manière semblable à ce que nous présentions dans Revault d'Allonnes et al. (2007). Les notions de stratégies seront illustrées d'exemples, dans la section 5.

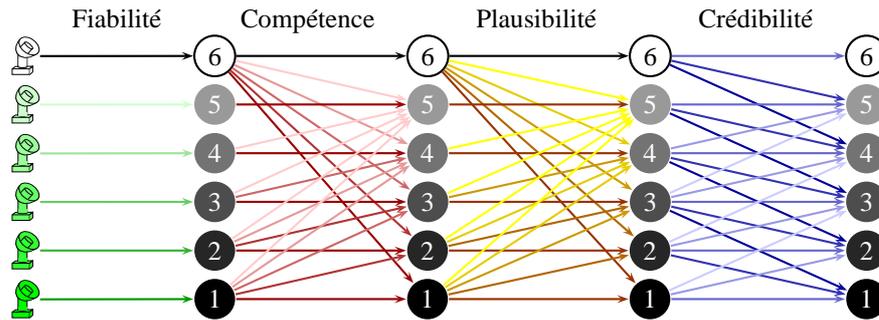


FIG. 2 – La chaîne de cotation

## 4.2 Degrés de vérité et incertitude

Avant d'entrer dans le détail de notre méthode, il faut noter que chaque évaluation de critère s'opère sur une même échelle, i.e. de 1≡'haute' à 5≡'basse', avec 6≡'inqualifiable'. Cependant, chaque évaluation possède sa propre interprétation, comme les tableaux de la section précédente le précisent. Nous allons maintenant construire un degré de qualité d'une information en fonction de ces critères.

Dans chaque grille de lecture des degrés de vérité, le score 'inqualifiable' est une valeur neutre. C'est à dire que la combinaison de n'importe quel score de confiance  $s$  à une évaluation de critère de niveau 6 renverra  $s$ . De même, n'importe quelle évaluation de critère  $e$  combinée avec un degré de confiance de 6 renverra  $e$  ou, dans le cas de la fiabilité de la source, l'interprétation de  $e$  fonction de la stratégie. Ceci explique que la figure 1 montre que n'importe quel score est atteignable depuis une confiance de 6. D'un autre côté, en raison de l'interprétation intuitive de la valeur 'l'indice ne peut être exprimé', aucun retour vers cette valeur n'est possible, une fois que la confiance l'a quittée. Ceci explique pourquoi aucune flèche ne pointe vers une confiance de 6.

Ce score, ou degré de confiance, ressemble en partie à des théories existantes et usitées. En effet, ce que nous proposons est l'évaluation de données incertaines et potentiellement imprécises. Nous proposons par exemple, dans Revault d'Allonnes et al. (2007), une interprétation multi-valuée du problème. Nous pensons donc que ces théories offrent de nombreux outils pour combiner les évaluations de nos critères, voire pour leur évaluation. Comme nous l'avons dit, notre échelle à six niveaux est un héritage direct de l'existant. Chaque critère pourrait, cependant, être défini et calculé de manière continue, puis exprimé à l'utilisateur à l'aide de variables linguistiques (cf. Zadeh (1975)). Le calcul de la confirmation, infirmation ou conflit pourrait être modélisé par la théorie de Dempster-Shafer Shafer (1976, 1990). Notre degré de confiance, lui-même, pourrait provenir d'une sorte de Transferable-Belief Model Smets et al. (1991), et associé à son interprétation pignistique.

## 5 Principes de la chaîne de cotation et stratégies

### 5.1 Fiabilité de la source

Supposons que nous ayons une nouvelle information entrante. Elle n'est donc pas encore évaluée. Puisque cette information est inconnue, nous allons la croire à la hauteur de la confiance faite à sa source. La première étape de notre processus va, ainsi, attribuer à une information inconnue le niveau de confiance (tableau 5, niveaux de gris dans la figure 1) équivalent à la fiabilité de sa source. La figure 1 présente la 'stratégie par défaut' qui associe  $A \Leftrightarrow 1, \dots, F \Leftrightarrow 6$ . À part la nécessité d'avoir une équivalence entre les degrés 'inqualifiables' – i.e.  $F \Leftrightarrow 6$  – toute autre stratégie acceptable – i.e. où  $A \geq B \geq C \geq D \geq E$ , voir tableau 6 pour des exemples – pourrait être utilisée.

### 5.2 Compétence et plausibilité

L'étape d'intégration de la 'compétence' de notre processus a un impact négatif sur le score de qualité. Ceci est dû au fait que l'initialisation de la confiance s'opérant au niveau de la confiance faite à la source, on aura sans doute supposé que ladite source 'sait de quoi elle parle' ou, du moins, que nous lui faisons confiance à hauteur de ceci. Il est donc logique que le seul impact de la compétence soit négatif. Une source de toute confiance fournissant une information hors de son domaine d'expertise nous incitera à prendre l'information avec circonspection.

De manière semblable, l'impact de la plausibilité de l'information aura tendance à faire baisser le score global. Si nous croyons une information à hauteur de sa source ; si sa source est compétente en la matière, le fait que l'information

	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>
A	1	2	3
B	2	2	4
C	3	3	4
D	4	5	5
E	5	5	5
F	6	6	6

TAB. 6 – 3 stratégies différentes pour le calcul de la fiabilité de la source. La stratégie S<sub>1</sub> est la stratégie par défaut de la figure 1

soit possible ne doit pas augmenter notre confiance, puisque nous avons commencé par la supposer au moins possible. En revanche, si elle paraît improbable, nous commencerons à en douter.

Pour déterminer de combien chacune de ces étapes va réduire la confiance, nous nous tournons maintenant vers les stratégies. La figure 1 montre les tendances générales attendues pour chaque niveau de confiance. Une fois encore, elle présente la stratégie par défaut pour laquelle l'impact croît proportionnellement à l'évaluation du critère. Cette stratégie et deux autres sont présentées dans le tableau 7.

Puisque chaque étape de la chaîne de cotation peut être vue comme une fusion du score courant avec l'évaluation d'un nouveau critère, toute méthode probabiliste Nimier (2004); Rogova et Nimier (2004) ou tout autre opérateur de fusion Bloch (1996) peut être envisagé. Par exemple, des opérateurs conjonctifs<sup>1</sup> pourraient modéliser une stratégie méfiante et les opérateurs disjonctifs représenteraient une stratégie confiante, de manière semblable à ce que nous proposons dans Revault d'Allonnes et al. (2007). Le choix des opérateurs, et donc des stratégies, sera laissé à l'utilisateur en fonction de l'application visée.

	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>
1	$c$	$c$	$c$
2	$\min(c + 1, 5)$	$\min(c + 2, 5)$	$\min(c + 1, 5)$
3	$\min(c + 2, 5)$	$\min(c + 3, 5)$	$\min(c + 3, 5)$
4	$\min(c + 3, 5)$	$\min(c + 4, 5)$	$\min(c + 3, 5)$
5	$\min(c + 4, 5)$	$\min(c + 4, 5)$	$\min(c + 5, 5)$
6	$c$	$c$	$c$

TAB. 7 – 3 stratégies différentes pour l'évaluation de la compétence de la source et la plausibilité de l'information, où  $c$  est l'indice de qualité courant. La stratégie S<sub>1</sub> est la stratégie par défaut de la figure 1

### 5.3 Crédibilité

La dernière étape de la chaîne de cotation est le calcul de l'indice de confirmation, hérité du STANAG2022. Puisque nous avons séparé les notions d'évaluation d'information et de comparaison d'informations acquises, cette dernière étape compare des informations possibles bien qu'incertaines et cotées. Ceci est une différence majeure avec l'étape précédente de plausibilité. En effet, ce que nous considérons à l'étape de la plausibilité est notre connaissance du monde. Nous l'avons appelée 'connaissance extérieure' pour la distinguer de la connaissance construite, et cotée. Nous insistons sur le fait que la connaissance extérieure ne sera pas remise en cause lorsque confrontée à nos informations construites. La confiance en sera toujours maximale.

À cette étape de crédibilité, cependant, nous comparons des informations incertaines. Chacune de ces informations a un degré de confiance associé, degré qui pourra être mis à jour lorsque nous comparerons les informations. L'évolution de la confiance en fonction des confirmations ou infirmations est un procédé naturel. Ceci étant, puisque nous considérons à la fois confirmation et infirmation, il coule de source que ce critère peut faire augmenter la confiance aussi bien qu'il peut la faire baisser. Puisque toutes les informations comparées sont cotées, leur score pourra être utilisé pour pondérer leur degré de confirmation ou d'infirmation (cf. Besombes et Cholvy (2007)).

<sup>1</sup>Un opérateur de fusion est dit conjonctif (resp. disjonctif) si et seulement si le résultat de la fusion est inférieur (resp. supérieur) ou égal au minimum (resp. maximum) des valeurs fusionnées.

## 6 Application symbolique : l'extraction d'information

Afin de pouvoir traiter le nombre croissant d'informations textuelles (sources ouvertes, interceptions, ...), des systèmes automatiques à partir de reconnaissance d'entités nommées peuvent assister un opérateur dans sa tâche d'extraction d'information. De tels systèmes peuvent extraire de l'information de textes de manière automatique et peuvent encore classifier et combiner ces informations pour créer de la connaissance plus élaborée. L'inconvénient de ce genre de technique est la perte de contrôle que l'utilisateur a sur le résultat. En effet, le système va extraire de l'information et la fournir à l'utilisateur, par exemple sous forme d'annotations. À ce moment l'utilisateur peut :

- faire totalement confiance au système en supposant qu'une information extraite est nécessairement vraie
- ou, se souvenant qu'aucun système n'est parfait, vérifier toutes les informations lui-même

De toute évidence, la seconde option est éliminée d'office puisqu'elle annule tout avantage de l'utilisation du système automatique. La confiance aveugle peut également être dangereuse, en particulier pour tout utilisateur manipulant des données sensibles.

Nous allons maintenant présenter comment notre chaîne de cotation, définie en section 4, s'intègre naturellement à une évaluation des informations produites par un système automatique.

Donnons-nous, tout d'abord, un rapport textuel, émanant d'une source fiable ainsi qu'une date automatiquement extraite de ce rapport (cf. Makkonen et Ahonen-Myka (2003)). La fiabilité de la source entraîne une confiance *a priori* que nous évaluerons à 1. Ce premier score est l'initialisation de notre chaîne de cotation.

Étant donné le texte et divers types d'information (une date, un sujet, un nom d'organisation, ..., voir National Institute of Standards and Technology (2007)), un algorithme d'extraction spécifique sera utilisé. Chaque algorithme pourra être assorti d'une évaluation (rappel, précision, f-mesure, ..., voir LDC (2001); NIST (2007) pour une liste exhaustive) constituant la compétence de l'algorithme à produire une information.

Supposons que l'algorithme ait une précision de 90% convertie en une compétence de niveau 2. Le choix de préférer la précision au rappel ou à la f-mesure pour l'évaluation de la compétence se justifie par le fait que la précision mesure la probabilité *a priori* qu'une information extraite soit pertinente, c'est à dire la compétence de l'algorithme en regard d'une information donnée. Selon la stratégie choisie, cf. section 5, la confiance sera maintenant d'un niveau 2.

La date extraite pourra ensuite être comparée à d'autres informations. Supposons, par exemple, que nous cherchions à dater une manifestation dans un pays et que la date initialement extraite soit le '04/07/2008'. L'extraction va produire l'information 'Vendredi quatre juillet 2008'. Si notre connaissance du pays place une fête nationale, ou l'anniversaire d'un événement important, le 4 juillet dans le pays étudié, cette connaissance extérieure va confirmer qu'une manifestation est possible à cette date. La plausibilité de l'information sera donc estimée à 1 mais la confiance restera à son niveau actuel.

Enfin, si le mot 'vendredi' est repéré plus loin dans le même texte, cette information tendra à confirmer la date du quatre juillet, et augmentera la crédibilité de l'information et, par là même, la confiance en notre information. D'autres informations extraites pourraient, également, contredire ou corroborer la date.

Cette implémentation de la chaîne de cotation pour l'extraction d'information est illustrée par la figure 2.

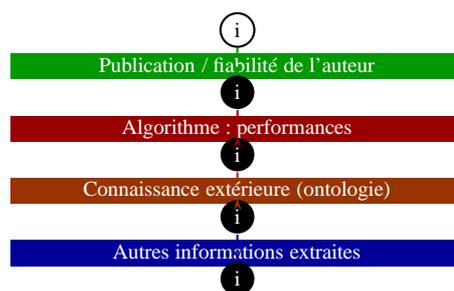


FIG. 3 – La chaîne de cotation pour l'extraction d'information

## 7 Conclusion

Parce que le contexte du renseignement militaire évolue vers le traitement de vastes quantités de données, structurées comme des mesures techniques de capteurs, ou déstructurées comme les sources ouvertes (presse, internet, rapports diplomatiques, ...) nous pensons que le besoin de systèmes automatisés augmente. Pour proposer des systèmes lisibles et utiles, il faut prévoir le filtrage de données en fonction de leur vertu informative et de leur qualité.

À partir des critères existants, nous proposons des critères additionnels pour évaluer automatiquement, un taux de confiance. Nous classons les facteurs utilisés selon leurs domaines de définitions et d'influence, intégrant deux nouveaux critères et en en redéfinissant un autre.

Nous expliquons ensuite comment nous avons construit une chaîne de cotation utilisant de manière intuitive nos nouvelles définitions.

## Références

- Besombes, J. et L. Cholvy (2007). Information fusion : using an ontology to information evaluation. In *Proceedings of the International Colloquium on Information Fusion 2007*, Xi'an, China, pp. 416–422.
- Bloch, I. (1996). Information combination operators for data fusion : A comparative review with classification. *Systems, Man and Cybernetics – Part A* 26(1), 52–67.
- LDC (2001). *Message Understanding Conference 7 (MUC7)*, Philadelphia. LDC.
- Makkonen, J. et H. Ahonen-Myka (2003). Utilizing temporal information in topic detection and tracking. In T. Koch et I. T. Solvberg (Eds.), *Proceedings of 7th European Conference on Digital Libraries (ECDL 2003)*, pp. 393–404. Springer-Verlag.
- Matheus, C. J., D. Tribble, M. M. Kokar, M. G. Ceruti, et S. C. McGirr (2005). Towards a formal pedigree ontology for level-one sensor fusion. In *Proceedings of the 10th International Command & Control Research and Technology Symposium*.
- National Institute of Standards and Technology (2007). *Sixteenth Text REtrieval Conference Proceedings (TREC 2007)*, Gaithersburg, Maryland. National Institute of Standards and Technology.
- Nimier, V. (2004). Information evaluation : a formalisation of operational recommendations. In *Proceedings of the Seventh International Conference on Information Fusion*, Stockholm, Sweden, pp. 1166–1171.
- NIST (2007). Automatic Content Extraction Evaluation (ACE07).
- OTAN (1997). Annex to stanag 2022 (edition 8).
- OTAN (2002). Allied Joint Publication – 2.1, Intelligence Procedures.
- Revault d'Allonnes, A., H. Akdag, et O. Poirel (2007). Trust-moderated information-likelihood. A multi-valued logics approach. In *Computation and Logic in the Real World, CiE 2007*, Sienna, Italy.
- Rogova, G. L. et V. Nimier (2004). Reliability in information fusion : literature survey. In *Proceedings of the Seventh International Conference on Information Fusion*, Stockholm, Sweden.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.
- Shafer, G. (1990). Perspectives on the theory and practice of belief functions. *International Journal on Approximate Reasoning* 4(5–6), 323–362.
- Smets, P., Y.-T. Hsia, A. Saffiotti, R. Kennes, et E. Umkehrer (1991). The transferable belief model. In R. Kruse et P. Siegel (Eds.), *ECSQARU*, Volume 548 of *Lecture Notes in Computer Science*, Marseille, France, pp. 91–96. Springer.
- Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning – i. *Information Sciences* 8(3), 199–249.

## Summary

Automatic processing of large quantities of data has become increasingly unavoidable in most areas. Some areas are, however, more sensitive to the quality of the produced data. Classical statistical measures are a cornerstone of algorithm evaluation, yet they rarely qualify the individual results. Our work offers criteria to evaluate these produced data and integrates them in a confidence-scoring chain. We show how to use our criteria in an automatic extraction system and how this system will score its production on the fly. This confidence score can then be used by the expert in charge of data verification either as a ranking index or as a highlighting on doubtful information. To produce the most comprehensive reliability score we can, we consider different parameters: source quality; its ability to produce considered information; the information itself in the light of expert knowledge and other constructed information.