

Extraction d'évènements au sein d'une plateforme de veille

Capucine Antoine*, Leila Khouas*

*Bertin IT R&D
396, rue du Mas de Verchant, Montpellier
capucine.antoine@outlook.com,
leila.khouas@bertin.fr
<https://www.bertin-it.com/>

1 Contexte et problématique

Dans le contexte d'une plateforme de veille stratégique, nous souhaitons intégrer un outil d'extraction d'évènements qui permettrait de présenter au veilleur les évènements pertinents décrits dans les textes collectés par le processus de veille mis en place. Cet outil devrait idéalement être capable d'extraire des évènements en plusieurs langues, de s'adapter à des types de textes hétérogènes et de traiter efficacement des volumes importants de données. Ce travail décrit notre approche pour réaliser un tel outil.

2 Extraction d'évènements

Un évènement se produit, implique des participants et provoque un changement d'état. En général, l'extraction d'évènements est envisagée comme une tâche de remplissage de formulaire.



FIG. 1 – Évènement de type **Attack** déclenché par **threw** et impliquant cinq arguments.

Pour répondre à cette problématique, différentes approches sont proposées telles que des approches par extraction de motifs ou par apprentissage automatique, notamment à base de réseaux de neurones. Des campagnes d'évaluation ont permis de fournir des corpus éponymes (MUC, ACE, TAC, ...) et d'établir des typologies d'évènements.

3 Approche utilisée

Nous utilisons une approche semi-supervisée basée d'une part, sur la construction automatique de dictionnaires de motifs à partir de corpus annotés, et d'autre part, sur des aménagements manuels du dictionnaire obtenu. Notre démarche repose sur l'utilisation de modèles de motifs de plus en plus complexes et riches, et leur évaluation en termes des indicateurs usuels de précision, rappel et F-mesure. La première étape consiste à construire automatiquement un dictionnaire associant un couple (trigger-POS) à un type d'événement ainsi qu'à des motifs caractéristiques. Ce dictionnaire est ensuite analysé et nettoyé manuellement pour écarter les triggers trop généralistes. L'extraction consiste à suspecter un type d'événement dès que le couple correspondant dans le dictionnaire est détecté. Nous avons utilisé les composants train et dev du

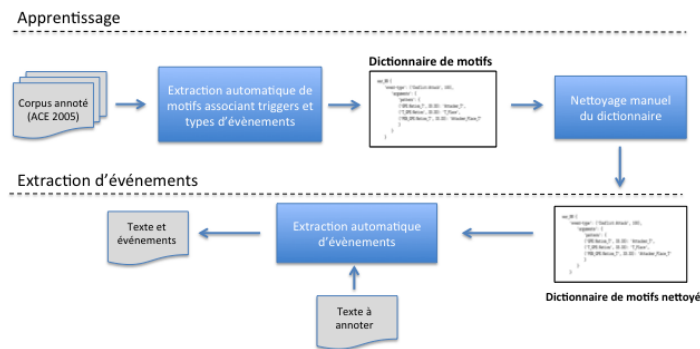


FIG. 2 – Approche générale utilisée.

corpus ACE 2005 pour l'apprentissage, et la composante test pour l'évaluation. Voici les résultats obtenus avec les premiers modèles :

	Précision	Rappel	F-Mesure
Identification du trigger	0.45	0.65	0.53
Classification du trigger	0.44	0.63	0.54

TAB. 1 – Résultats des premiers modèles.

Summary

Event extraction consists in extracting structured information about unstructured events in texts. This work focuses on event extraction in the context of business intelligence. We propose the model of an event extraction tool based on a semi-supervised pattern-matching method.