

## ANALYSE DES ARCHIVES DU WEB POUR PREDIR LES EVENEMENTS A VENIR

### Deux chercheurs analysent les archives du web pour prédire les événements à venir

Jean-Laurent Cassely

Peut-on prédire les événements du futur avec un bon algorithme et (beaucoup) d'analyse de données? C'est l'hypothèse testée par Eric Horvitz, chercheur chez Microsoft, et Kira Radinsky, qui travaille au Technion-Israel Institute de Haïfa, dans leur article «Mining the Web to Predict Future Events», explique le site Gigaom.

En compilant les données disponibles à un instant  $t$ , on devrait pouvoir générer des alertes concernant par exemple les maladies ou les manifestations à venir, pensent les auteurs. Les deux chercheurs ont donc expérimenté leur idée sur 22 années d'archives du New York Times, de 1986 à 2007, pour tester sa validité.

L'expérience consiste à observer si un certain type de menace survient dans le cadre d'une chaîne d'événements précise, par exemple en cherchant les mot-clé «malaria» ou «choléra». Or, en suivant les articles de presse sur les catastrophes naturelles comme les tempêtes ou les sécheresses, on aurait pu prévoir un an en amont une recrudescence du choléra en Angola, écrivent les deux chercheurs.

Autre exemple parlant dans le domaine du maintien de l'ordre, celui de la prédiction des manifestations et protestations. Plusieurs cas impliquant la mort d'un innocent par la police dans un quartier à forte population immigrée ont permis aux auteurs de relier certains événements (les funérailles, le procès des policiers, etc.) à des protestations ou émeutes.

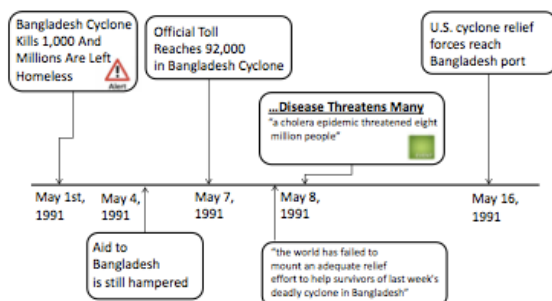


Figure 6: Example of cholera alert following storms in Bangladesh. Triangular alert icons represent inferences of significant upswings in likelihood of forthcoming cholera outbreak.

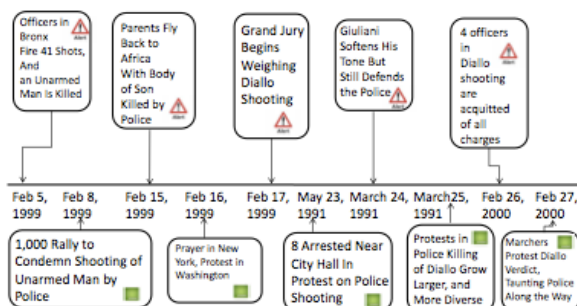


Figure 7: Example of alerts on the likelihood of forthcoming riots after shooting of unarmed minority. Triangular alert icons represent inferences of significant upswings in likelihood of a forthcoming riot.

### Des exemples de chaînes d'événements identifiées par les auteurs

Ces usages prédictifs de l'informatique se multiplient de par le monde. La start-up Recorded Future propose déjà un service de ce type à ses clients, parmi lesquels des agences gouvernementales américaines, écrit Technology Review.

Les policiers de Los Angeles utilisent eux un logiciel analysant les archives des actes de délinquance, qui leur indique les zones dans lesquelles patrouiller en priorité, avec un certain succès.

IBM a récemment annoncé que le style steam punk allait devenir une mode majeure des années à venir, en compilant des sources en ligne et notamment des discussions sur les réseaux socio-numériques. Et des chercheurs britanniques ont quant à eux créé un algorithme pour prédire où vous vous trouverez dans 24 heures... à l'aide des données de tracking (suivi géographique) de votre téléphone mobile. Leur marge d'erreur? Environ 20 mètres.