

# Veille d'Information sur le Web avec Re-Watch

Christophe Brouard\*, Christian Pomot\*\*

\*Université Grenoble Alpes, LIG UMR 5217/Equipe AMA, France  
Christophe.Brouard@imag.fr,  
<http://ama.liglab.fr/brouard/>

\*\*Société Com&Net, 155 Cours Berriat 38028 Grenoble Cedex 1  
cpomot@com-et-net.com  
<http://com-et-net.com>

**Résumé.** Les algorithmes d'apprentissage automatique peuvent être utilisés pour créer des outils de recommandation qui permettent de prédire la pertinence d'un document pour une thématique de veille donnée en se basant sur les précédents jugements de pertinence donnés pour cette thématique pour d'autres documents. Ces outils de recommandation permettent de filtrer dans un flux entrant de documents ceux qui sont susceptibles d'être pertinents sans que l'utilisateur ait besoin de déterminer lui-même les mots clefs marquant l'adéquation d'un document pour un sujet de la veille. Bien que cette problématique de recherche ait été abondamment abordée, les outils de veille d'information pour le web intégrant un apprentissage en sont encore à leur balbutiements. Nous présentons ici l'application web Re-Watch permettant la définition d'un thème de veille, la sélection de sources d'information sur le web relatives à ce thème et l'adaptation des scores de pertinence des documents aux retours de l'utilisateur. L'application permet aussi, pour chaque thème, une auto-évaluation de la qualité du filtrage et une interrogation du moteur de recherche Google. Cette application encore en cours de développement est néanmoins actuellement fonctionnelle et accessible sur le web à l'url suivante : <http://www.specificsearch.com>.

## 1 Introduction

La veille d'information peut-être définie selon Cacaly et al. (2008) comme « un processus continu et dynamique faisant l'objet d'une mise à disposition personnalisée et périodique de données ou d'informations, traitées selon une finalité propre au destinataire, faisant appel à une expertise en rapport avec le sujet ou la nature de l'information collectée ». Plusieurs aspects importants ressortent de cette définition. D'une part, le processus est continu, c'est-à-dire qu'il s'étend sur une certaine durée (typiquement plusieurs semaines, mois ou années). Or les moteurs de recherche traditionnels comme Google, Bing, Yahoo, pour citer les plus utilisés, sont conçus pour des besoins d'information ponctuels. Les fonctions de sauvegarde des informations pertinentes et des sources associées y sont notamment très rudimentaires et la récupération des résultats des précédentes sessions de recherche n'est pas aisée. D'autre part, le processus est dynamique, des informations apparaissent et disparaissent tous les jours. Or