

Veille Technologique assistée par la Fouille de Textes

François Jacquenet, Christine Largeron, Stéphanie Chapaux

Université Jean Monnet de Saint-Etienne
EURISE

23 rue du docteur Paul Michelon
42023 Saint-Etienne Cedex 2

Francois.Jacquenet@univ-st-etienne.fr
Christine.Largeron@univ-st-etienne.fr

Résumé. Le domaine de la veille technologique vise à récolter, traiter, et analyser des informations scientifiques et techniques utiles aux acteurs économiques. Dans cet article, nous proposons d'utiliser des techniques de fouille de textes pour automatiser le processus de traitement des données issues de bases de textes scientifiques. Toutefois, la veille introduit une difficulté inhabituelle par rapport aux domaines d'application classiques des techniques de fouille de textes, puisqu'au lieu de rechercher de la connaissance fréquente cachée dans les données, il faut rechercher de la connaissance inattendue. Les mesures usuelles d'extraction de la connaissance à partir de textes doivent de ce fait être revues. Pour ce faire, nous avons développé le système UnexpectedMiner dans lequel de nouvelles mesures permettent d'estimer le caractère inattendu d'un document. Notre système est évalué sur une base de résumés d'articles dans le domaine de l'apprentissage automatique.

1 Introduction

Depuis quelques années le secteur économique a pris conscience des enjeux liés à la maîtrise de l'information stratégique. Toutefois, les entreprises sont de plus en plus submergées d'informations. Elles ont de grandes difficultés à dégager les données stratégiques dont elles ont besoin pour anticiper les marchés, prendre des décisions et agir sur leur environnement socio-économique [Samier et Samoval, 2001, Martinet et Marti, 2001, Revelli, 2000]. Ceci a conduit à l'émergence de l'intelligence économique définie par H. Martre comme "l'ensemble des actions de recherche, de traitement, de distribution et de protection de l'information obtenue légalement et utile aux acteurs économiques" [Martre, 1994]. Lorsque les informations à analyser sont de nature scientifique et technique, on parle plus spécifiquement de veille technologique pour désigner la surveillance des brevets et de la documentation scientifique (articles, thèses, ...) [Desvals et Dou, 1992, Jakobiak, 1990, Jakobiak, 1994].

Le processus de veille peut être décomposé en quatre phases principales : l'audit des besoins, la collecte des données, le traitement des données et la synthèse et la diffusion des résultats. Dans cet article, nous nous intéressons principalement à la troisième phase. Pour automatiser le traitement des données collectées, les techniques de fouille