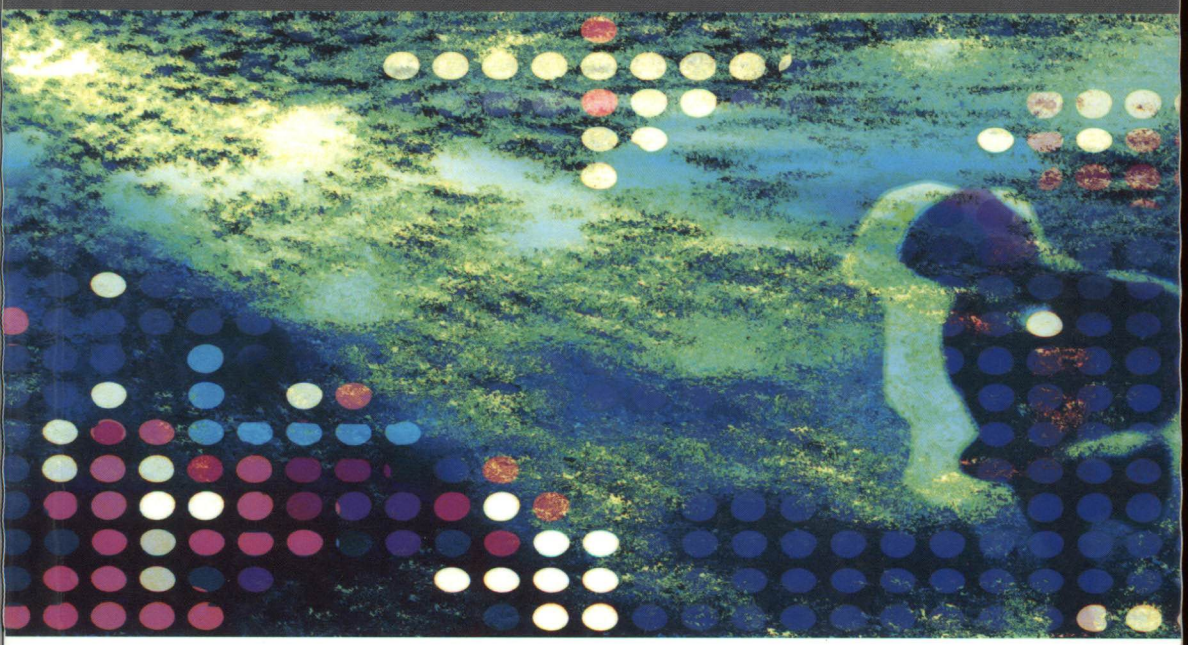
COLLECTION SYSTÈMES D'INFORMATION ET ORGANISATIONS DOCUMENTAIRES
SOUS LA DIRECTION DE STÉPHANE CHAUDIRON



Expérimentations et évaluations en fouille de textes

un panorama des campagnes DEFT

sous la direction de

Cyril Grouin
Dominic Forest

Hermes Lavoisier

Expérimentations et évaluations en fouille de textes

un panorama des campagnes DEFT

sous la direction de
Cyril Grouin
Dominic Forest

Hermes Science publications
Lavoisier

Table des matières

La fouille de textes est une activité combinant traitements informatiques et données linguistiques avec comme objectif principal l'extraction et l'organisation automatique des informations présentes dans les textes. Deux familles de méthodes permettent d'atteindre ce but : celles à base de connaissances d'experts et celles reposant sur un apprentissage automatique supervisé.

Une campagne d'évaluation consiste à confronter les systèmes développés par plusieurs équipes sur un même jeu de données et en un temps limité. Créé en 2005 à l'image des campagnes anglo-saxonnes, le défi fouille de textes (DEFT) est aujourd'hui la seule campagne d'évaluation francophone en fouille de textes.

Cet ouvrage rassemble les méthodes utilisées lors des différentes éditions du défi. Les thématiques relèvent de la classification de documents en genres et thèmes, de la fouille d'opinions et de l'identification de la période de parution d'un document.

Les coordonnateurs

Ingénieur d'Etudes CNRS au LIMSI, Cyril Grouin travaille sur l'anonymisation de documents cliniques et sur les entités nommées étendues. Il coorganise les campagnes d'évaluation DEFT sur la fouille de textes.

Dominic Forest est professeur agrégé à l'Ecole de bibliothéconomie et des sciences de l'information de l'Université de Montréal. Ses recherches portent sur la fouille de textes et les humanités numériques.

978-2-7462-3836-7

hermes Science publications

www.hermes-science.com