Intelligent Content Acquisition in Web Archiving

Mots clés :
- Directeur de thèse : PIERRE SENELLART
- Co-encadrant(s) : 
- Unité de recherche : Laboratoire Traitement et Communication de l'Information
- École doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal : Divers

Résumé du projet de recherche (Langue 1)

Background ---------- Since its beginnings in the 1990s, the set of hyperlinked resources known as the World Wide Web has become an extremely precious source of information. For easy access to this content, search engines such as Google or Bing crawl the Web and index pages. But crawling technologies also have an important application in the field of Web archiving [1]: the storage of (versioned) archives of the Web, to ensure continuous access to its resources despite data deletion, to carry out temporal analyzes of its content, or to provide a user the possibility to see a Web page or a Web site as it was at a given date in the past. This PhD topic is part of the ARCOMEM European project (2011-2013). This project, involving 12 institutions around Europe, aims at leveraging the wisdom of the crowds for intelligent preservation of Web content. PhD Topic Description