

Projet METAREVE : Vers l'extraction automatique de métadonnées par reconnaissance vocale

Violette Abergel^{1,2}, Van Tuan Bui³, Amine Berbagui³, Olivier Malavergne⁴, Ruven Pillay⁵, Vincent Detalle⁶

¹ MAP, UPR 2002 CNRS, Marseille, France

² CNRS, INSA Lyon, Université Claude Bernard Lyon 1, LIRIS, UMR5205, Villeurbanne, France

³ Fondation des Sciences du Patrimoine, Cergy-Pontoise, France

⁴ LRMH, CRC, UAR 3224, Champs-sur-Marne, France

⁵ C2RMF, Paris, France

⁶ SATIE, CY Cergy Paris Université

Abstract :

Pour les sciences du patrimoine, terrain de production de savoirs sociétaux et collectifs reposant sur une perpétuelle confrontation entre objets matériels et études pluridisciplinaires alliant sciences humaines et sociales et sciences expérimentales, les enjeux de gestion de données numériques se trouvent aujourd'hui confrontés à la forte hétérogénéité de sources documentaires (textes, images, vidéos, ...), de données analytiques (issues de divers capteurs, techniques d'imagerie, analyses d'échantillons, ...) et de processus de traitement mobilisés à des fins de description, analyse, suivi, ou encore conservation. Garanties de la fiabilité d'une donnée, les métadonnées et parodonnées de provenance fournissent des renseignements précieux pour rendre compte des contextes d'acquisition et des possibilités de réutilisation ultérieures. Cependant, leur production rigoureuse se révèle bien souvent chronophage voire fastidieuse, les informations à renseigner étant à la fois diverses et difficiles à retrouver *a posteriori*. Dans ce contexte, le projet interdisciplinaire METAREVE propose de simplifier cette démarche en l'accompagnant depuis la production des données sur le terrain jusqu'à leur traitement *ex situ*. La méthode proposée vise à automatiser l'extraction de ces informations à l'aide d'approches de compréhension automatique de la parole empruntées au domaine de l'intelligence artificielle, en s'appuyant pour cela sur des thésaurus construits par les différentes communautés mobilisées dans les sciences du patrimoine. Il a en particulier pour objectif de fournir une brique logicielle qui pourra alimenter l'écosystème développé dans le cadre de l'EquipEx+ ESPADON.