

| Appel à projets |

## **La FMSH coordonne une nouvelle ANR sur la valorisation des archives multimédias**

La chaire UNESCO ITEN « Innovation, Transmission, Édition Numériques » (FMSH / Université Paris 8) est lauréate d'une ANR avec le projet ARCHIVAL - Valorisation d'archives multimédia.

Si les tentatives pour indexer et structurer les textes, vidéos et images sont nombreuses, les interfaces de résultats de recherche offrent encore trop peu de possibilités de navigation, d'exploration et d'articulation des contenus. Le projet ARCHIVAL travaillera sur la compréhension automatique multimodale du langage pour développer de nouvelles interfaces intelligentes de médiation et de transmission des savoirs.

Le questionnement du projet est double :

- Quel rôle peuvent jouer les méthodes de compréhension par les machines dans la réinterprétation de fonds d'archives thématiques ?
- Selon quelles modalités des interfaces de médiation des contenus peuvent-elles exploiter des résultats générés par les méthodes actuelles d'Intelligence Artificielle ?

Pluridisciplinaire et multi-acteurs, le projet vise à faire collaborer des chercheurs issus des Sciences de l'Information et de la Communication et de l'Informatique autour de la valorisation des archives et du partage des savoirs pour les arts, la culture et le patrimoine.

Le projet ARCHIVAL est coordonné par la FMSH sous la direction scientifique de la Professeure Ghislaine Azémard, titulaire de la chaire UNESCO ITEN. Il réunit la Chaire UNESCO ITEN (FMSH / Université Paris 8), le Laboratoire d'Informatique et Système (LIS) de l'Université Aix-Marseille (AMU) à travers l'équipe TALEP (Traitement Automatique du Langage Ecrit et Parlé), l'équipe LINKMEDIA de l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) et Orange Labs.

## Résumé du projet

---

Le numérique transforme l'accès aux savoirs qui restent encore difficilement appréhendables par les publics, malgré de nombreuses tentatives pour indexer, structurer les textes, vidéos et images selon des référentiels. Très souvent présentées sous forme de listes exclusivement fonctionnelles, les interfaces de résultats de recherche offrent peu de possibilités de navigation et d'exploration et articulent mal les contenus entre eux. Le projet pluridisciplinaire et multi-acteurs ARCHIVAL vise à faire collaborer des chercheurs issus des Sciences de l'Information et de la Communication et de l'Informatique autour de la valorisation des archives et du partage des savoirs pour les arts, la culture et le patrimoine.

Il s'est structuré autour des questionnements suivants :

- Quel rôle peuvent jouer les méthodes de compréhension par les machines dans la réinterprétation de fonds d'archives thématiques ?
- Selon quelles modalités des interfaces de médiation des contenus peuvent-elles exploiter des résultats générés par les méthodes actuelles d'Intelligence Artificielle ?

Les méthodes de « compréhension automatique du langage » permettent de structurer l'information et de générer des cartographies de l'espace sémantique qui raisonnent comme un maillage de l'information. Les méthodes neuronales de type « end-to-end », permettent d'apprendre directement des représentations (projections dans des espaces numériques) en fonction d'une tâche sans avoir à spécifier explicitement les différents niveaux d'analyse nécessaires à sa réalisation. Les espaces vectoriels sémantiques ainsi construits permettent de définir des relations entre des mots et des documents, mais aussi d'utiliser ces représentations pour générer des sorties prenant en compte le sens, la « science latente » contenue dans les documents, dépassant en cela la simple extraction de termes (mots clés, entités, triplets RDF).

Parallèlement, les méthodes de « compréhension de document par les machines » (Machine Reading Comprehension) permettent d'interroger des documents par des questions naturelles pour en découvrir des extraits pertinents. Un document est alors appréhendé au prisme des relations qu'il entretient avec d'autres documents et des questions sur son contenu, mais aussi au travers des synthèses (résumés, réécritures, cartographies) qui peuvent être générées dynamiquement suite à son analyse. Un des objectifs scientifiques majeurs du projet ARCHIVAL est d'opérer cette compréhension de manière multimodale en exploitant des documents mêlant paroles, textes, images fixes et vidéos. Les méthodes neuronales rendent possible l'intermodalité grâce à la définition d'espaces vectoriels de représentations communs aux différentes modalités.

Exploiter cette convergence, qui a donné lieu à des résultats spectaculaires, ouvre de nombreuses possibilités pour la médiation par le numérique. Les équipes d'ARCHIVAL s'interrogeront sur les manières de structurer une collection de documents hétérogènes en faisant apparaître de manière explicite les liens implicites, de révéler la nature de ces liens et de les valoriser de manière intelligible par la médiation d'interfaces ergonomiques qui garantissent une appropriation réussie des contenus. Un périmètre de corpus a été délimité avec le fonds mixte sur l'autogestion de la FMSH, récemment labellisé CollEx, qui sera complété par le large fonds audiovisuel universitaire de Canal-U. L'analyse et la valorisation de ce fonds représente un intérêt pour les SHS dans un contexte où il devient nécessaire de repenser structurellement de nouveaux modèles de développement socioéconomiques (autonomie démocratique, économie sociale et solidaire, développement alternatif...). En termes d'impact, ARCHIVAL s'attachera à mettre en œuvre un cœur technologique générique pouvant être transposé par la suite à d'autres champs d'études, valorisable par les partenaires et dont le rayonnement pourra être porté au sein du Campus Condorcet.

## **En savoir +**

---

### **Chaire UNESCO « Innovation, Transmission, Édition Numériques » (ITEN)**

Dirigée par la Professeure Ghislaine AZEMARD, la chaire ITEN (FMSH / Université Paris 8) est dédiée à la recherche, la formation et l'expérimentation interdisciplinaires sur les nouvelles modalités de médiation, de transmission et d'édition par le numérique. Elle est spécialisée dans le domaine des médias émergents et des humanités numériques.

[www.chaire.fr](http://www.chaire.fr)

### **FMSH**

La Fondation Maison des sciences de l'homme (FMSH) est un carrefour national et international contribuant au développement, au rayonnement et à l'attractivité des sciences humaines et sociales. Elle crée les conditions favorables à la recherche et au dialogue interdisciplinaire, et initie des dynamiques où se mobilisent réseaux scientifiques et où se croisent acteurs publics et privés.

[www.fmsch.fr](http://www.fmsch.fr)

## **Contact presse**

---

Fondation Maison des sciences de l'homme

Chloé Lepart | [clepart@msh-paris.fr](mailto:clepart@msh-paris.fr) | 01 40 48 64 15 | 06 19 44 14 80