

---

# Transcrire le passé à l'aide d'outils du présent

Stéphane Nicolas<sup>\*†1</sup>

<sup>1</sup>Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes (LITIS) – Université de Rouen : EA4108, Université du Havre, Institut National des Sciences Appliquées de Rouen – Avenue de l'Université UFR des Sciences et Techniques 76800 Saint-Etienne du Rouvray, France

## Résumé

```
< !-[if gte mso 9]> 0 0 1 266 1469 LITIS 12 3 1732 14.0 < ![endif]-> < !-[if gte mso 9]> Normal 0 21 false false false FR JA X-NONE < ![endif]-> < !-[if gte mso 9]> < ![endif]-> < !-[if gte mso 10]> /* Style Definitions */ table.MsoNormalTable {mso-style-name:"Tableau Normal"; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:""; mso-padding-alt:0cm 5.4pt 0cm 5.4pt; mso-para-margin:0cm; mso-para-margin-bottom:.0001pt; mso-pagination:widow-orphan; font-size:12.0pt; font-family:Cambria; mso-ascii-font-family:Cambria; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Cambria; mso-hansi-theme-font:minor-latin;} < ![endif]-> < !-StartFragment->
```

Qu'il s'agisse de transcriptions diplomatiques ou semi-diplomatiques de manuscrits de Gustave Flaubert, ou de la transcription d'un ouvrage médiéval constitué d'une collection de manuscrits écrits entre le X<sup>e</sup> et le XIII<sup>e</sup> siècle, le travail de description et d'interprétation de documents modernes ou anciens commence souvent pour les chercheurs par un travail fastidieux et difficile de transcription, réalisé manuellement à l'aide d'outils non nécessairement adaptés à cette tâche. Ce travail de description et de transcription est également important pour permettre l'indexation et la recherche d'information dans les images numériques de ces documents dans le cadre de la constitution de vastes bibliothèques numériques et d'archives numérisées afin de transposer ces témoins de notre Histoire, dans le monde numérique qui est le notre aujourd'hui. Dans ce contexte d'accès et de partage de la connaissance, l'enjeu des laboratoires de recherche travaillant dans le domaine du document numérique et de l'analyse des documents numérisés, est de développer en collaboration avec les acteurs de la valorisation du patrimoine des outils adaptés répondant aux besoins actuels dans ce domaine. Nous présentons dans cette intervention l'état actuel des avancées dans ce domaine, et nous montrons, à travers l'exemple de différents projets menés au LITIS, que l'analyse automatique de documents peut apporter une aide adaptée à cette problématique, en permettant d'analyser la structure des documents (agencement des parties textuelles ou détection de ratures dans les manuscrits de Flaubert, ou structuration logique en articles dans des journaux anciens), et de détecter automatiquement des entités textuelles ou graphiques par exemple.

```
< !-EndFragment->
```

---

\*Intervenant

†Auteur correspondant: [Stephane.Nicolas@litislab.eu](mailto:Stephane.Nicolas@litislab.eu)