**Brunel TCHEKELI**



Note réflexive



**NB : Une partie de la note réflexive est incluse dans la rédaction du mémoire**

Tout au long de mon stage, j'ai été immergé dans le projet AGODA, une initiative visant à redéfinir la manière dont nous interagissons avec les archives numérisées des débats parlementaires français. Cette immersion m'a non seulement permis de comprendre l'importance de ces archives pour notre patrimoine culturel et historique, mais aussi les défis techniques et méthodologiques associés à leur exploitation et à leur enrichissement structurel.

**Missions demandées :**

La mission qui m’a été confiée lors de ce stage était à la fois unique et passionnante : explorer le projet AGODA et mettre en lumière la manière dont les archives numérisées des débats parlementaires français pourraient être rendues plus exploitables. Pour AGODA, Il ne s'agissait pas seulement de se contenter des données numérisées par la BNF, mais de les enrichir, de les segmenter et de les annoter pour garantir leur pertinence à travers le temps mais aussi d’ouvrir sur d’autres possibilités d’analyses. Pour ce faire, l'objectif principal du stage était d'automatiser le processus d'encodage de l'ensemble du corpus des débats parlementaires, afin d'éviter l'insertion manuelle des balises. L'idée était d'élaborer un programme capable de gérer cet encodage, avec pour finalité la génération d'un fichier XML. Une fois le programme ajusté en fonction des résultats, il serait ensuite testé avec d'autres ensembles de données regroupés par séance pour assurer son adaptabilité.

**Compréhension :**

**La France a une histoire parlementaire riche et a toujours mis l'accent sur la conservation de ses débats et documents parlementaires. Ces archives offrent un aperçu de l'évolution politique, sociale et économique du pays.**

**Avec la numérisation massive des archives historiques, de nouvelles opportunités se présentent pour la recherche en histoire et en sciences humaines. Cependant, leur utilité est parfois limitée en raison de leur format d'accès (comme des images ou des textes bruts).**

**Le projet AGODA vise à améliorer l'accès à ces archives. Il s'agit d'une collaboration entre différents laboratoires et institutions et vise à créer une plateforme innovante pour consulter et explorer les débats parlementaires.**

**L'application TEI Publisher a été choisie pour le projet en raison de sa capacité à gérer des bases de données XML. La TEI (Text Encoding Initiative) est un standard international pour la représentation des textes numériques. La norme XML-TEI n'était pas seulement un outil, mais un moyen de préserver la richesse sémantique des archives. La mission dépassait la technique ; elle touchait à la préservation du patrimoine, la mise à disposition de ces archives et envisager de nouvelles manières d’analyse de ces débats. Dans le cadre de cette mission, l'accent est mis sur l'importance du balisage automatique et la vérification de la conformité aux normes TEI.**

**L'ensemble des débats parlementaires a été numérisé et converti en un format JSON. Chaque fichier est structuré en éléments, et chaque élément possède plusieurs attributs ou clés. Ces clés donnent des informations sur la nature, la structure et le contenu de l'élément. L'une des clés, "comment", est essentielle car elle guide la manière dont l'information doit être encodée.**

**Réalisation et Apprentissages :**

**Pour répondre aux missions du stage, un programme python a été développé pour faciliter le processus d'encodage et de conversion. Ce script utilise les commentaires comme indications pour le balisage, transformant ces commentaires en balises XML appropriées. Le script mis en place a grandement simplifié le processus d'encodage, mais il n’était pas sans les défis.**

**L'intersection de la technologie et de l'histoire est fascinante. J'ai découvert comment l'automatisation, la numérisation et la collaboration interdisciplinaire peuvent révolutionner la manière dont nous valorisons notre patrimoine historique. De plus, la méthodologie et l'importance de l'automatisation dans le traitement des données m'ont fourni une perspective précieuse sur la manière dont la technologie peut être utilisée efficacement dans le domaine des sciences humaines.**

**L'une des compétences majeures que j'ai acquises a été l'automatisation des balisages. J'ai appris à utiliser des outils tels que lxml, ce qui m'a permis de manipuler et de traiter efficacement des documents XML. Plus important encore, j'ai appris à vérifier la conformité d'un document, garantissant ainsi son intégrité et sa précision. Au-delà des compétences techniques, j'ai eu l'opportunité de résoudre à la fois une problématique métier et une problématique technique en humanités numériques. Cette double problématique m'a offert une perspective unique sur la manière dont la technologie peut être appliquée pour répondre à des besoins réels dans un environnement professionnel.**

**Ce mémoire et le travail sur le projet AGODA ont été parmi les expériences les plus enrichissantes de ma vie académique. J'ai non seulement développé des compétences techniques, mais j'ai également acquis une compréhension approfondie de la manière dont la technologie peut servir des objectifs historiques et professionnels concrets.**

**Difficultés rencontrées :**

Si la tâche a été enrichissante, elle n'a pas été sans défis. La manipulation des éléments "seg" s'est révélée être une tâche difficile, nécessitant une grande attention aux détails. Les incidents et citations, malgré leur importance pour capturer le contexte, ont ajouté une couche de complexité. La transformation de simples commentaires en balises appropriées, tout en évitant les redondances et erreurs, a été un défi constant.

Mais ces défis techniques n'étaient que la partie visible de l'iceberg. Sur le plan personnel, chaque obstacle a été une occasion d'apprendre et de m'adapter. La familiarité avec l'XML, la TEI, et les subtilités de l'encodage a nécessité de plonger profondément dans le domaine des humanités numériques, un voyage d'apprentissage passionnant en soi.