

NotaryPedia: Représentation graphique des connaissances et visualisation des textes du patrimoine culturel

Département d'intelligence artificielle
Université de Malte, Malte

Charlene Ellul, Joel Azzopardi, Charlie Abela

Les Archives notariales de La Valette sont fréquentées par différents utilisateurs : notaires, chercheurs en histoire, étudiants, artistes et parfois aussi des civils intéressés par l'histoire de leur famille, leur généalogie ou leurs biens. La conservation de plus de 20 000 manuscrits historiques n'est pas une mince affaire et représente un défi constant.

NotaryPedia est un projet passionnant visant à extraire automatiquement les données des actes notariés historiques et utilise une approche évolutive pour stocker ces données. Les entités et les relations extraites fourniront un meilleur aperçu de notre passé et l'enrichissement multidisciplinaire des données aidera les chercheurs à compléter le casse-tête de nos ancêtres. Ainsi, en plus de préserver les informations contenues dans les actes notariés, NotaryPedia vise à faire revivre le passé en reliant les personnes, les lieux et les mots clés.

A partir des registres les plus anciens et les seuls transcrits de la collection¹, remontant au 15^{ème} siècle, des techniques d'apprentissage automatique sont utilisées pour extraire automatiquement les dates, les personnes, les lieux et les mots clés. Ces registres sont principalement écrits en latin, mais on peut aussi trouver des mots médiévaux siciliens et maltais où l'absence d'un meilleur mot ne pouvait être trouvée en latin. L'annotation des documents se fait à l'aide des indices des publications et des modèles de d'apprentissage automatique ont été formés sur ces données. Les résultats ont été remarquables en ce qui concerne la reconnaissance automatique des noms de personnes, des lieux, des mots-clés populaires et des phrases clés. Des mots-clés et des phrases-clés tels que faldellas, olej, asinj montrent le sujet du document et donnent un aperçu de ce que les citoyens maltais commerçaient. La méthode d'indiction pour représenter les dates est courante dans ces documents et une approche fondée sur des règles a été utilisée pour extraire ces dates.

De plus, la classification des actes a été jugée importante surtout lorsque le scribe a omis le type d'acte ou lorsque seul un texte partiel de l'acte est disponible ou lisible. Les actes sont classés comme Apoca, Debitum, Dos.

Les informations extraites sont stockées à l'aide d'une approche évolutive où aucun schéma prédéfini n'est requis. Cela permet d'enrichir davantage les données des documents eux-mêmes, d'autres archives liées à différentes disciplines (comme la paléographie, la conservation, la linguistique

maltaise), des données ouvertes disponibles sur le Web et finalement aussi des résultats des historiens eux-mêmes.

Visuellement, les entités extraites peuvent être représentées comme des bulles tandis que les relations entre ces entités peuvent être représentées comme des liens entre elles. Cette structure est mieux connue sous le nom de « *Knowledge Graph* » (« graphique des connaissances ») et est illustrée à la figure 1.

Les relations entre les entités représentent des relations généalogiques telles que la mère, le fils, la veuve, l'origine de personnes telles que « Jacobo vit à Hal Manin » et les limites de zones géographiques telles que « Hal Leu-limites de-Hal Manin ». D'autres techniques d'enrichissement sont à l'étude pour trouver d'autres relations entre acheteurs, vendeurs et objets ou services échangés.

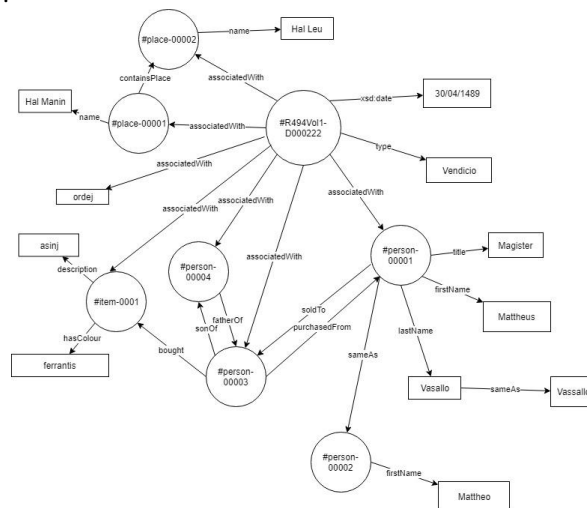


Fig. 1. Un sous-graphique de la représentation graphique des connaissances avec les relations entre entités

Une interface Web conviviale sera bientôt disponible pour visualiser les données et permettre aux utilisateurs de naviguer dans la représentation graphique des connaissances. De plus, le crowd sourcing contrôlé va être progressivement mis en place, ce qui permettra aux utilisateurs de contribuer directement à la croissance de cet espace.

Suivez-nous sur nos médias sociaux :

<https://www.facebook.com/notarialarchives/>

<https://www.instagram.com/notarialarchivesfoundation/>

* Ce travail est partiellement financé par le projet E-18LO28-01 dans le cadre de la collaboration entre les Archives Notariales de La Valette et l'Université de Malte.

¹Compilé par Dr Stanley Fiorini dans la série « *Documentary Sources of Maltese History* »