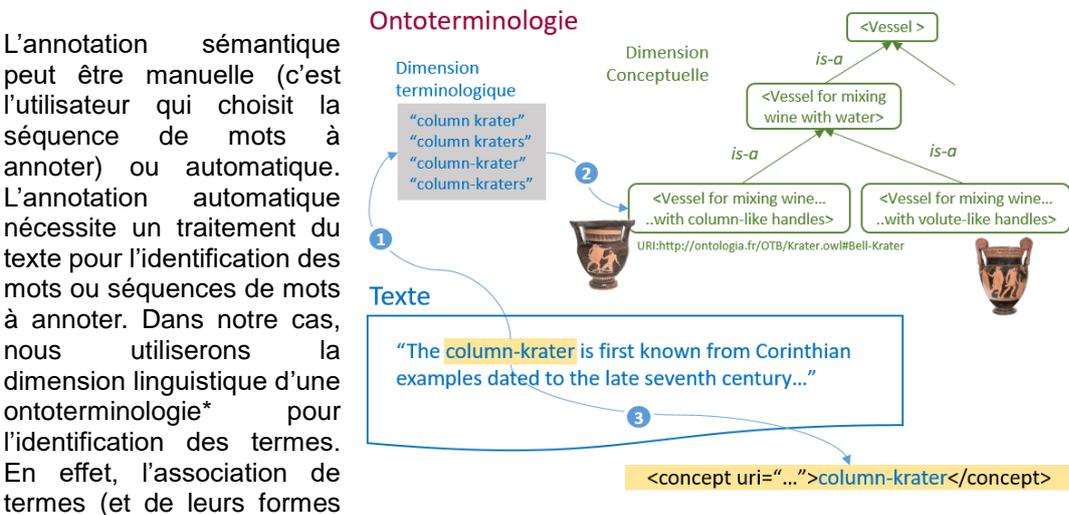


Titre	ANNOTATION SEMANTIQUE
Niveau	Ecole d'Ingénieurs 2 ^{ème} année
Dates et durée	Début : 1 ^{er} juin 2020 Fin : 31 juillet 2020 Durée : 9 semaines
Ville, Pays	Le Bourget du Lac (Campus Scientifique), France
Laboratoire	LISTIC (Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance) http://www.polytech.univ-savoie.fr/LISTIC
Description du sujet de stage	<p>Problématique : L'annotation de documents vise à enrichir un texte en associant aux mots des informations externes qui peuvent être de nature linguistique comme la catégorie morpho-syntaxique (nom, verbe, adjectif, etc.) ou sémantique, c'est-à-dire en lien avec ce que ce désigne les mots comme une personne, un lieu, un événement, un artefact. Plus précisément, et en ce qui nous concerne, l'annotation sémantique consiste à annoter un texte à l'aide d'une ontologie, c'est-à-dire à étiqueter les termes d'un document avec des liens qui pointent vers les éléments de l'ontologie en rapport avec les termes. Cet étiquetage se fait à l'aide de balises XML en lien avec le Web Sémantique.</p> <p>L'annotation sémantique peut être manuelle (c'est l'utilisateur qui choisit la séquence de mots à annoter) ou automatique. L'annotation automatique nécessite un traitement des mots ou séquences de mots à annoter. Dans notre cas, nous utiliserons la dimension linguistique d'une ontoterminologie* pour l'identification des termes. En effet, l'association de termes (et de leurs formes fléchies) à un concept permet un étiquetage automatique. La figure ci-dessus illustre ce principe.</p> <p>*Une ontoterminologie est une terminologie dont le système conceptuel est une ontologie formelle.</p> <p>Objectif du stage :</p> <p>Conception et implémentation (au choix : Java, Python, Javascript) d'un environnement d'annotation sémantique manuelle et automatique de textes. L'environnement sera paramétrable par des ontoterminologies du domaine des Humanités Numériques et appliqué à des textes relevant de ce domaine. Les ontoterminologies qui seront mises à disposition seront représentées dans un format du W3C (RDF, OWL, SKOS).</p> <p>Mots clés : Annotation sémantique, Ontologie, Terminologie, Traitement Automatique des Langues (TAL), Web Sémantique, Humanités Numériques.</p> <p>Références :</p> <ul style="list-style-type: none"> - https://fr.wikipedia.org/wiki/Annotation_s%C3%A9mantique - http://alic.paris-sorbonne.fr/PAGESPERSO/atanassova/lfa/Cours11.pdf - https://halshs.archives-ouvertes.fr/halshs-01505635/document - http://www.quaero.org/media/files/bibliographie/lipn_qpr5_ma.pdf - https://www.supinfo.com/articles/single/3380-etat-art-i-annotation-semantique - « Terminologie et ontologie pour les humanités numériques : le cas des vêtements de la Grèce antique » C. Roche, M. Papadopoulou. Revue Humanités numériques, Humanistica, n°2, 2020, 22 pages (sur demande aux auteurs) 
Compétences	Conception & Programmation (Java, Python, Javascript)
Gratification	Indemnités de stage selon la législation en vigueur
Tuteur / Contact	Pr Christophe Roche Téléphone : 04 79 75 87 79 (06 86 25 07 05) roche@univ-savoie.fr http://christophe-roche.fr/