

Rapport Projet Web Sémantique

EL KATEB Sami, PAUL Thomas

Évolution de notre modélisation de rapports médicaux

Pour faire évoluer notre modélisation de rapports médicaux, nous avons commencé par intégrer les retours reçus sur notre premier projet. Ainsi, nous avons remplacé la propriété *gender* par les classes Man et Woman. Puis nous avons ajouté les numéros de sécurité sociale des patients et les numéros RPPS des médecins. Nous avons défini ces propriétés comme owl:InverseFunctionalProperty. En effet, si deux entités ont la même valeur pour l'une de ces deux propriétés, celles-ci sont identiques. Pour cette même raison, ssn et rpps sont les valeurs de owl:hasKey pour les patients et les médecins respectivement.

Puisque nous avons commencé à développer une ontologie OWL, nous avons mis à jour la déclaration de notre ontologie en y ajoutant la référence à la version précédente. Nous avons également corrigé nos erreurs précédentes où certaines propriétés définies comme ObjectProperty auraient dû être des DataProperty.

Nous avons enrichi notre ontologie par des liens de parenté, tels que hasAncestor, une propriété asymétrique, irréflexive et transitive, et sa sous-propriété hasParent, qui n'est pas transitive. Nous avons aussi ajouté hasChild, l'inverse de hasParent, qui possède les mêmes types que celle-ci. Enfin, nous avons ajouté hasBrother et hasSister, toutes deux disjointes, irréflexives et sous-propriétés de hasSibling. Ces liens de parenté nous ont permis de mettre en évidence les prédispositions génétiques. Ainsi, nous avons créé la classe PersonWithGeneticDiabetesPredisposition, incluant les personnes ayant un ancêtre ou un frère/sœur diabétique. Nous avons choisi de ne pas exclure les personnes déjà diagnostiquées diabétiques de cet ensemble, considérant qu'elles restent prédisposées génétiquement au diabète.

Ensuite, nous avons défini l'ensemble des maladies respiratoires (RespiratoryCondition) comme l'union des maladies présentant des symptômes tels que la toux, l'essoufflement ou le mal de gorge. Nous avons aussi défini l'ensemble des maladies infectieuses et, son complément, celui des maladies non infectieuses.

Finalement, nous avons défini une équipe (Team), puis une équipe médicale (MedicalTeam) comme étant une équipe comprenant au moins un médecin.

Nous avons aussi intégré trois thésaurus SKOS dans nos données : un thésaurus de médicaments, un thésaurus de symptômes et un autre de maladies.

Ces thésaurus se prêtent bien à une organisation hiérarchique, facilitant la classification des différentes familles de médicaments, types de symptômes et de maladies. Ils permettent également d'établir des relations entre les différents termes. La création de ces thésaurus aide à normaliser la terminologie. De plus, nous avons créé une collection de maladies contagieuses, permettant de regrouper les maladies selon leur contagiosité, contrairement au thésaurus organisé par organe atteint.

OWL Entailment

Lors de l'élaboration de notre ontologie OWL, nous avons défini les clés des patients comme étant leur numéro de sécurité sociale, en utilisant la propriété `hasKey`.

Cette propriété stipule que deux instances partageant le même numéro de sécurité sociale sont considérées comme une seule et même instance. Nous pouvons observer l'effet de cette propriété en recherchant le poids du Patient7. Bien que l'instance Patient7 ne possède pas de propriété `weight`, le Patient11, ayant le même numéro de sécurité sociale, possède cette propriété. Ainsi, les deux instances étant traitées comme identiques, il est possible de récupérer le poids du Patient7 (**mcr_query_entailment 1**). De même, définir la propriété `ssn` comme `InverseFunctionalProperty` produit le même effet.

Nous avons également défini les `PersonsWithGeneticDiabetesPredisposition` comme étant les individus ayant un ancêtre ou un sibling (frère/sœur) atteint de diabète. Toutefois, les seules relations explicitement définies dans nos données sont `hasChild` et `hasBrother`. La relation `hasParent` est inférée à partir de la relation `hasChild`, dont elle est l'inverse, ce qui permet ensuite d'inférer la relation `hasAncestor`, dont `hasParent` est la sous-classe.

La relation `hasAncestor` est transitive et s'étend sur plusieurs générations, ce qui permet d'inférer des prédispositions génétiques à partir des diagnostics posés sur les grands-parents. Ainsi, en incluant uniquement les relations `hasChild` dans nos données, le raisonneur OWL RL infère correctement que les patients 2, 8 et 9, bien qu'ils ne soient pas diagnostiqués diabétiques, ont une prédisposition génétique au diabète (**mcr_query_entailment 2**).

Les maladies respiratoires sont quant à elles inférées comme étant l'union des maladies présentant l'un des différents symptômes respiratoires. Nous pouvons les identifier en recherchant les entités de type `RespiratoryCondition` (**mcr_query_entailment 3**). Quant aux équipes médicales, il s'agit d'équipes comprenant au moins un médecin. Nous observons que les équipes contenant des médecins sont classées comme équipes médicales, alors que l'équipe 3, ne comprenant aucun médecin, ne l'est pas (**mcr_query_entailment 4**).

Contraintes SHACL

Lors de notre projet initial, nous avons défini la classe “rapport de cas” (CaseReport) qui comprenait un médecin, un patient, un examen clinique et, éventuellement, une thérapie. Cependant, cette modélisation a été modifiée pour ne plus inclure de patient afin d’éviter la duplication des données déjà présentes dans l’examen clinique. Pour affiner et contrôler cette modélisation, nous avons utilisé une contrainte SHACL. Ainsi, nous avons créé `shapes:case_report`, une contrainte exigeant que chaque rapport de cas comporte au moins un médecin et un examen clinique.

De la même manière, nous avons défini les examens cliniques comme étant associés à un seul patient et à au moins un médecin, conformément à la contrainte `shapes:clinical_examination`. La validité de ces contraintes peut être facilement vérifiée en retirant le médecin de l’examen clinique ou du rapport de cas, ou en augmentant le nombre de patients.

Une autre évolution de notre projet Web Sémantique a été l’introduction de trois thésaurus. Pour standardiser les entités de ces thésaurus, nous avons créé des contraintes SHACL pour les médicaments, les signes et symptômes, et les maladies. Cela nous a permis de contraindre les instances de ces classes à être des `skos:Concept` et à contenir au moins un `prefLabel` et une définition. De plus, nous avons exigé que les instances de la classe médicament incluent au moins une substance active, une unité de médicament et une forme galénique. De même, les maladies doivent présenter au moins un signe clinique ou un symptôme. La validité de ces contraintes peut être testée en retirant les propriétés requises ou en incluant des langues non prises en charge.

Enfin, nous avons défini des contraintes spécifiques pour les médecins et les patients. Notamment, les instances de ces deux classes doivent être identifiées par un numéro RPPS pour les médecins et un numéro de sécurité sociale pour les patients. Le format de ces numéros est vérifié à l’aide d’expressions régulières. Nous pouvons tester le bon fonctionnement de ces contraintes en utilisant des numéros invalides.