

Construction d'ontologies légères pour les types de données

Démarche

Nous avons repris les différentes étapes visant à construire une ontologie de domaine à partir du thésaurus IAU pour les appliquer aux sous domaines de l'astronomie liés aux types d'objets : galaxy, multiple object, nebulae, source et star.

Note : le type multiple objet ne correspondant à aucun concept existant dans le thésaurus, il n'a pas été exploité pour l'instant.

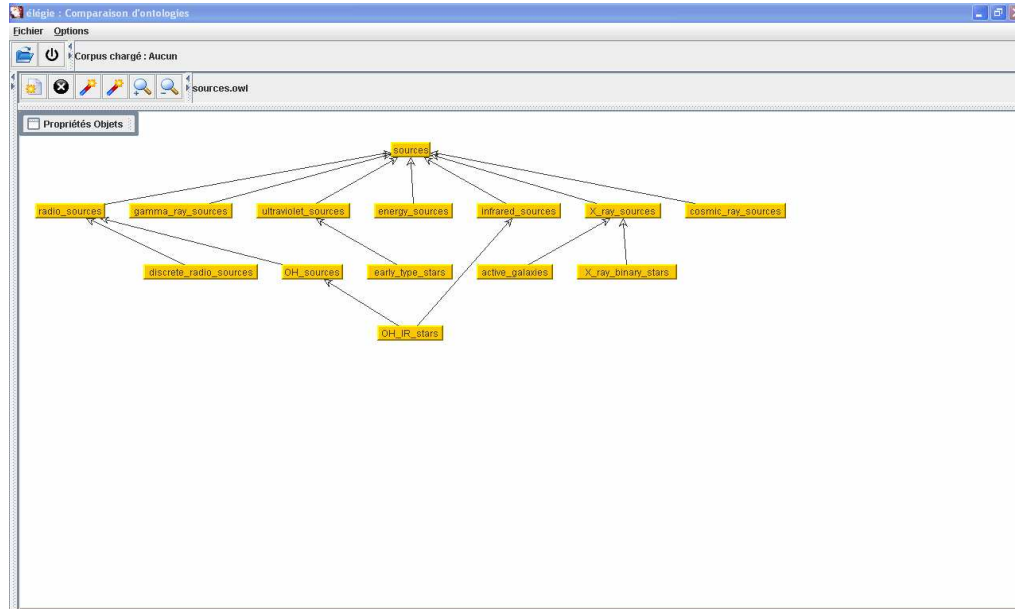
Pour cela, nous avons :

- Extrait du thésaurus
 - les concepts «galaxy, neb, source et star » et de leur hiérarchie descendante
 - les concepts « liés à » ces concepts dans le thesaurus
- Nous obtenons à ce niveau-là 4 ontologies « de base » pour chacun des 4 types.
- Proposé de nouveaux labels plus spécifiques pour les relations « liés à » (ou RT)
- Extrait à partir des corpus (réalisés par Soizic pour chaque type) des nouveaux termes en lien avec les termes existants dans les ontologies.
- Proposé des placements dans les ontologies pour les nouveaux termes extraits
- Développé un prototype d'interface pour faciliter la validation par les experts des différents éléments créés

Ce document vise à vous présenter des exemples de ce que nous obtenons de façon à ce que vous puissiez évaluer si vous souhaitez entamer la validation des différents éléments permettant de créer les ontologies finales. Ces ontologies ont pour but d'être intégrées à un système de recherche d'information permettant d'évaluer de quel type d'objet traite un document textuel.

1. Echantillon de l'ontologie de base extraite du thésaurus pour le type source

L'ontologie est visualisée à partir d'un prototype qui pourra vous être fourni pour explorer les différentes ontologies



2. Analyse des relations « est lié à » pour les concepts descendants de source

Même principe que lors de la pris en compte du domaine de l'astronomie dans son intégralité.

Concepts de type source liés à des concepts de type star

early_type_stars orion_population_stars
 radio_stars X_ray_stars
 HZ_Herculis_stars X_ray_binary_stars
 OH_IR_stars variable_stars
 A_dwarf_stars A_subdwarf_stars
 O_stars wolf_rayet_stars
 early_type_variable_stars O_stars
 X_ray_binary_stars X_ray_pulsars
 Ae_stars Be_stars
 optical_pulsars radio_pulsars
 Oef_stars Of_stars
 OB_stars runaway_stars

Concepts de type source liés à des concepts de type sources

HZ_Herculis_stars X_ray_binary_stars
 A_dwarf_stars A_subdwarf_stars
 blazars quasars
 B_stars early_type_variable_stars
 early_type_variable_stars O_stars
 X_ray_binary_stars X_ray_pulsars
 Ae_stars Be_stars
 optical_pulsars radio_pulsars
 A_stars early_type_variable_stars
 Oef_stars Of_stars

Proposition de labels pour les relations :

recording

move off
subdivide
single out
tangle
diffract
watch
wander
bridge
scatter off
upflowing

Ces propositions devront être validées pour définir dans un tableau les labels de relations possibles entre différents types d'objets.

3. Nouveaux termes extraits à partir du corpus sources

micropulses
parcel
spgc
extra delay
short micropulses
iron sulfide
pi
gb
ce
stable stage
ca xix
hook
newborn pulsar
cspn
microstructure
vertical optical depth
quasiperiodicities
mf slope
variation of outflow velocity
newborn
orbital aberration

La pertinence de ces nouveaux termes devra être validée ainsi que les propositions de placement dans l'ontologie. Afin de faciliter cette étape, un prototype a été développé. Contrairement aux premières évaluations faites, nous proposons maintenant d'accéder à l'ensemble de la connaissance représentée pour un concept (label, relation, concept père/fils) avant de valider l'intégration de nouvelles connaissances.

4. Utilisation de l'interface de validation

Nous avons développé une interface visant à faciliter la validation des nouveaux termes et leur intégration dans l'ontologie. Cette interface permet de visualiser à la fois les nouveaux termes, les concepts existants dans l'ontologie et les relations proposés entre les nouveaux termes et les concepts existants.

Le prototype a l'avantage de sauver dans un fichier les éléments validés ainsi que les éléments rejetés et de mettre à jour l'ontologie en fonction.

La fenêtre de démarrage permet de sélectionner le fichier contenant les nouveaux termes et l'ontologie existante (Figure 1).

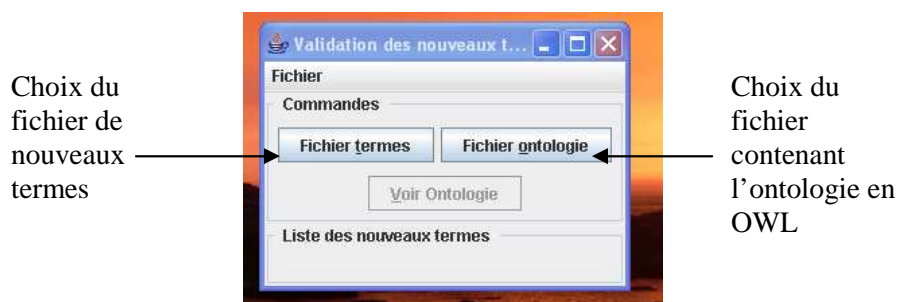


Figure 1 Copie d'écran de la fenêtre de démarrage

Une fois les deux fichiers chargés, le prototype met en place les différentes analyses pouvant être faites pour chaque nouveau terme du fichier (intégration dans l'ontologie en tant que nouveaux concepts en relation avec un concept existant, ou bien en tant que concept fils ou père de concepts existants). Pour accéder aux propositions obtenues par les analyses, l'utilisateur sélectionne un terme dans la liste de nouveaux termes (Figure 2) et l'interface est mise à jour (Figure 3). La mise à jour consiste à afficher le concept existant dans l'ontologie auquel il est proposé de rattacher le terme ainsi que le label pour la relation. Afin de donner à l'utilisateur un maximum de connaissance sur ce qui est déjà représenté dans l'ontologie, le concept et l'ensemble de la connaissance qui lui est déjà associée dans l'ontologie est présenté dans une nouvelle fenêtre. L'utilisateur peut alors choisir de valider la proposition (la proposition est intégrée dans l'ontologie) ou bien de la rejeter (reste alors à analyser les autres propositions pour le nouveau terme).

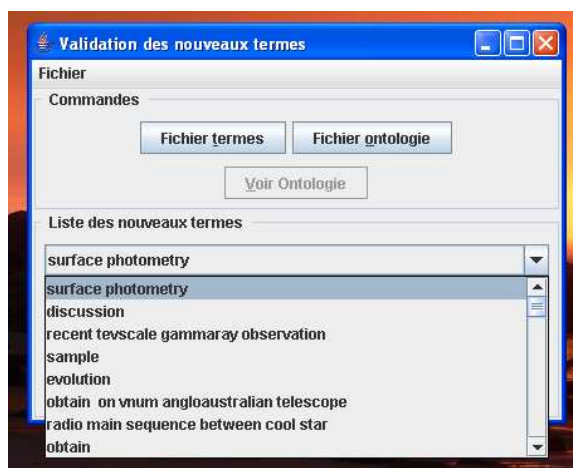


Figure 2 : visualisation des nouveaux termes

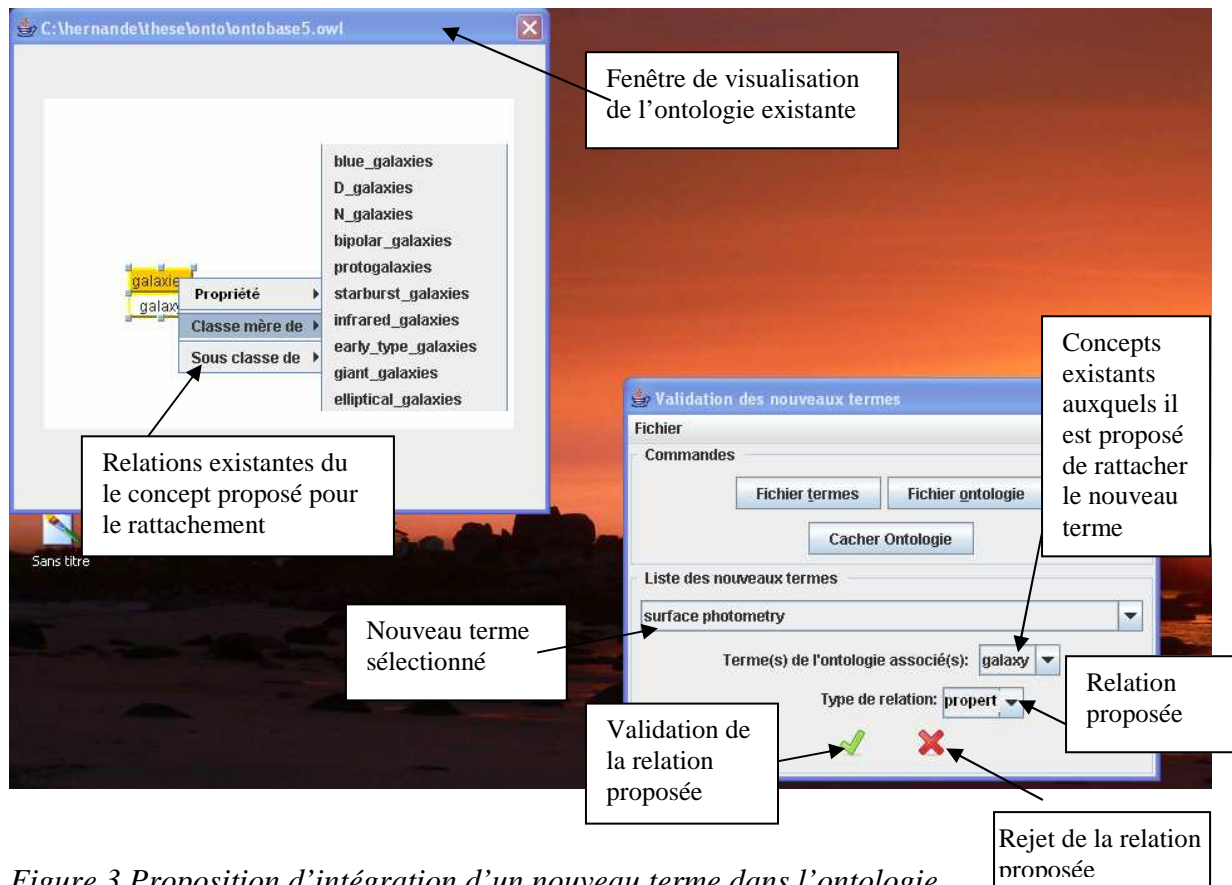


Figure 3 Proposition d'intégration d'un nouveau terme dans l'ontologie