

Description des connaissances en astronomie: des metadonnées aux ontologies

Sebastien Derriere

CDS

derriere@astro.u-strasbg.fr

Plan de l'exposé

Résumé des travaux Ontologies dans MDA

- Bases de connaissance en astronomie
 - Etat des lieux
 - Metadonnées
 - Applications
- Prototypage sur les Ontologies
 - Application aux UCDs
 - Application au thésaurus
- Perspectives

Etat des lieux

- Bases de connaissances utiles en astronomie
 - Standardisation pour le VO
 - Application des techniques ontologies/web sémantique
- Thésaurus IAU (pas de maintenance)
- Mots-clés FITS
- Mots-clés des journaux
- UCDs (en cours d'évolution pendant MDA)

Metadonnées en astronomie

- **Ontologie Ottawa**
 - Inexploitable, trop vague... ou trop précis
 - Besoin de définir des ontologies sur des domaines précis, pour répondre à des besoins ciblés
- **Définition de métadonnées standardisées**
 - Identification d'un besoin
 - Consensus pour l'adoption (communauté)
 - Respect des standards
 - Maintenance sur le moyen/long terme

Définition de metadonnées

- Définition précise de metadonnées dans des domaines spécifiques
 - Voir exposé de M.-L. Dubernet sur les données atomiques et moléculaires
 - Retombées sur les UCDs (e.g. Objets du système solaire)

Applications

- Travail sur les UCDs
 - Collaboration CDS/LORIA
 - Post-doc Huaizhong Kou (2003/2004)
 - Stage Sébastien Derivaux (07-08/2004)
 - Stage Alexandre Richard (DEA, 2005)

Pourquoi les UCDs?

- Les UCD ont été créés à l'origine pour le service VizieR de catalogues au CDS

+ de 5000 catalogues différents

Origines diverses

- Numérisation de plaques photos
- Observations dans tout le spectre EM
- Publication électronique de tables des journaux



Les descriptions et noms de colonne varient énormément

- Plus de 140 noms différents pour une magnitude V !!

Que sont les UCDs ?

- UCD = Unified Content Descriptors
- UCD : métadonnées associées aux données
- Donner une **description sémantique** des quantités astronomiques = ce que c'est, avec un vocabulaire standardisé
- UCDs particulièrement utiles pour
 - Homogénéiser des collections de données hétérogènes
 - Interopérabilité lors des échanges de données

Première version: UCD1

- ESO–CDS data mining project

Exploration manuelle de 100,000 colonnes

creation de nouveaux mots pour chaque quantité

~1500 mots créés

- PHOT_EXTINCTION_ISM = Interstellar extinction
- POS_GAL_LAT = Galactic Latitude

- UCD donne une description sémantique du contenu des colonnes

Permet de rechercher des catalogues d'après leur contenu

Permet des comparaisons automatisées

Assignment d'UCD

- Objectif: trouver quel UCD correspond à une description textuelle (courte!)
- Travail de Huaizhong Kou sur les UCD1
 - Indexation par Fastr
 - Test de différents algorithmes
 - Baysien
 - K-NN
 - Rocchio
- Contribution à la définition des UCD1+, et corrections dans Vizier

IVOA: UCD1+

- UCD1 définis avant le VO
- L'IVOA définit des standards permettant l'interopérabilité entre les projets VO: intérêt pour les UCDs
- Migration vers UCD1+
 - re-utiliser une base existante (UCD1)
 - syntaxe plus flexible
 - vocabulaire contrôlé par IVOA

IVOA: UCD1+

- **Même but que UCD1**
 - Décrire des quantités: "ce que c'est"
 - Avec une précision raisonnable
 - Permettre l'interopérabilité
- **Nouvelle syntaxe (IVOA Rec. 19/08/2005)**
 - Pas sensible à la casse
 - Possibilité de combiner plusieurs mots dans 1 UCD
 - Le 1er mot porte l'essentiel du sens
- **Nouvelle liste de mots (<500) avec définitions (IVOA Rec. 31/12/2005)**

Exemples d'UCD1+

- `pos.eq.dec;meta.main` = déclinaison
- `phot.mag` = magnitude
- `phot.mag;em.opt.V` = magnitude mesurée dans l'optique entre 500 et 600nm
- `stat.error;phys.temperature` = erreur de mesure sur la temperature
- `phot.color;em.opt.B;em.opt.V` = couleur B-V

UCDs dans le VO

- Standards sur le site IVOA
 - <http://www.ivoa.net/Documents/latest/>
- Utilisation dans d'autres standards:
 - VOTable (attribut "ucd" de <FIELD> et <PARAM>)
 - Registry: description du contenu des colonnes <column> pour TabularSkyService
 - VOEvent: utilisation dans <param> et <group>
- Utilisation dans Cone Search
- Utilisation dans applications (Aladin, TopCat)

Outils UCD1+

- Ensemble d'outils disponibles pour:
 - Faciliter le passage UCD1 -> UCD1+
 - Manipuler les UCD1+
- Outils sous forme de scripts CGI et Web Services
 - <http://vizier.u-strasbg.fr/UCD/tools.htx>
 - <http://cdsweb.u-strasbg.fr/cdsws/ucdClient.gml>

UCD builder



UCD builder



Use this interface to find the UCD corresponding to a description:

Enter a description in natural language:

Clear Form

Reset Form

maximum value of R magnitude

Build UCD

Last updated Thu May 12 15:54:16 2005

Suggested complete UCD:

phot.mag;em.opt.R;stat.max

Refine your search:

The following words matched your query. You might force the selection of words matching precisely your query, and rebuild a UCD using these words:

Clear Form

Reset Form

word	definition	flag	score
<input type="checkbox"/> phot.mag	Photometric magnitude	E	60
<input type="checkbox"/> phot.mag.bol	Bolometric magnitude	Q	50
<input type="checkbox"/> phys.magAbs.bol	Bolometric absolute magnitude	Q	30
<input type="checkbox"/> em.opt.R	Optical R band between 600 and 750 nm	S	15
<input type="checkbox"/> stat.max	Maximum or upper limit	S	30

Rebuild UCD using selected words.

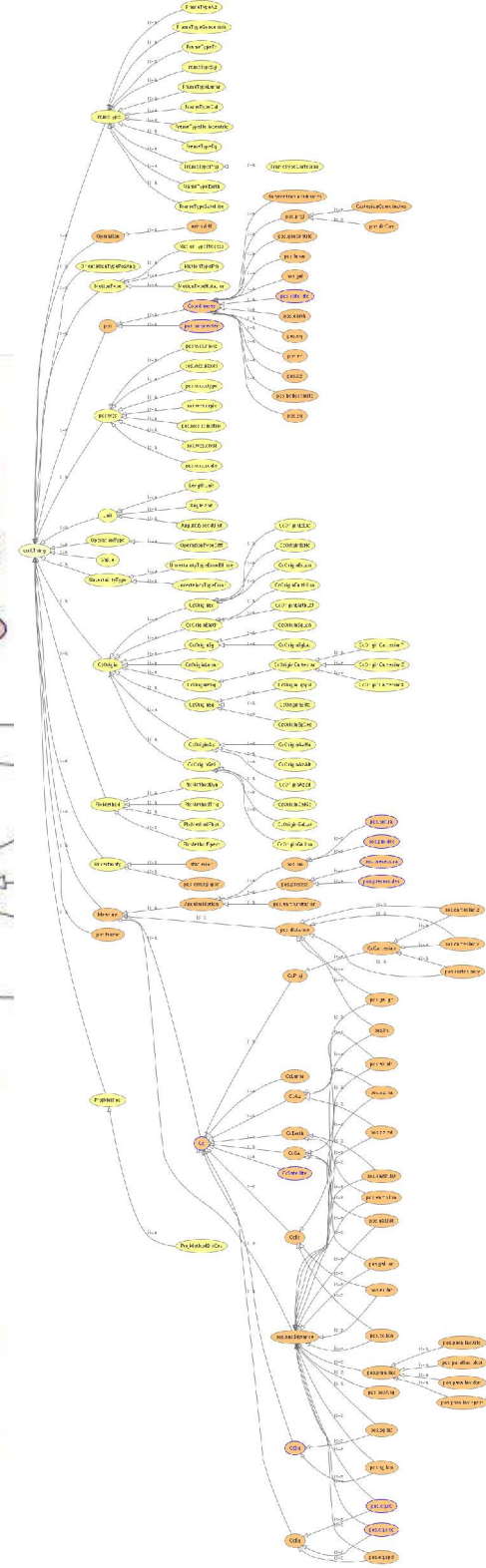
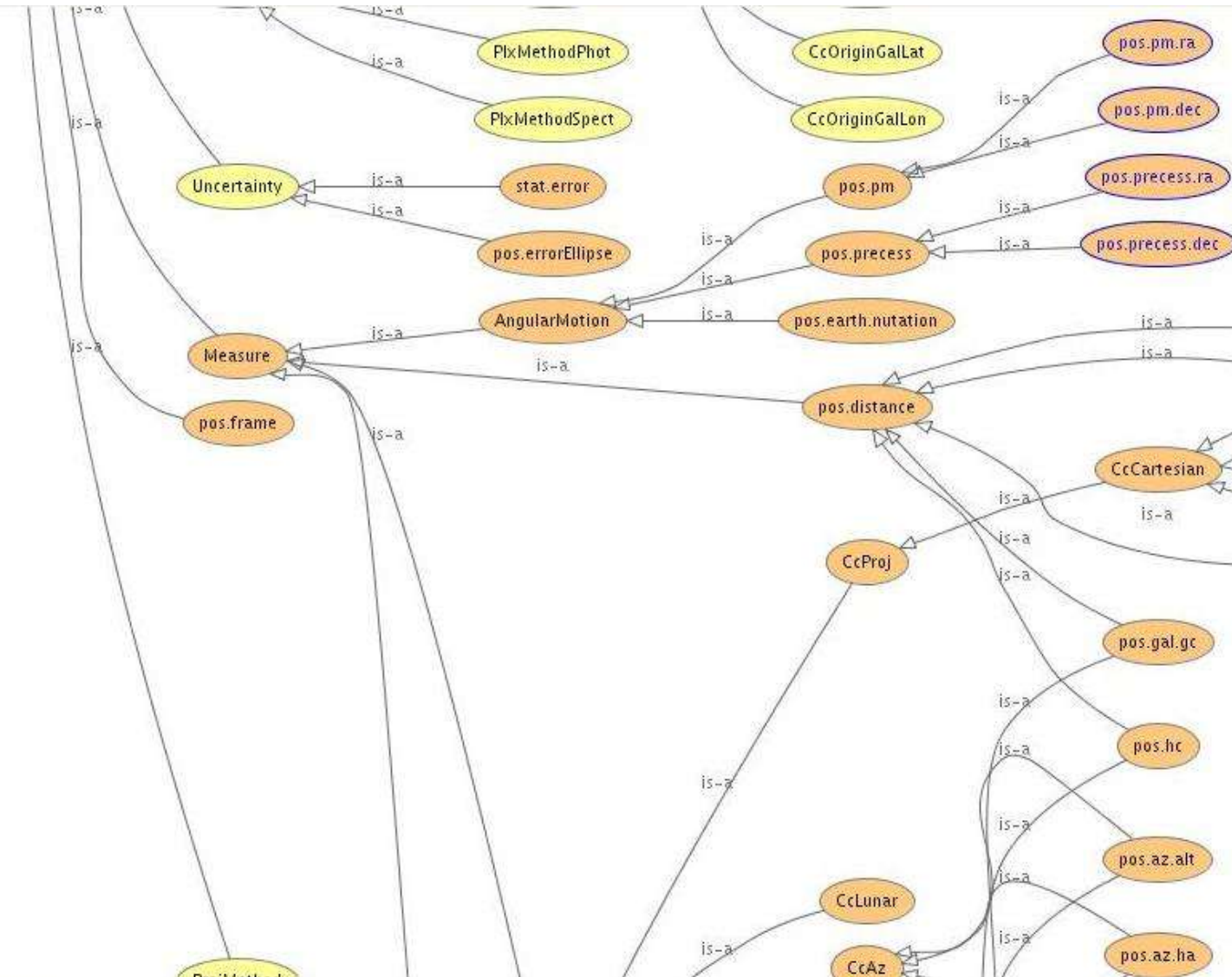
Ontologie des UCDs

- Approche exploratoire
- Stage de S. Derivaux (+ Rim Al Hulou)
 - Ingestion automatique de la liste des mots (utilisation des flags de syntaxe)
 - Utilisation de Protégé pour l'édition
 - Utilisation d'APIs pour les applications Web reposant sur la sortie OWL
- Applications:
 - Assignation d'UCDs
 - Identification de concepts similaires

Ontologie des UCDs

- Stage d'Alexandre Richard (mots pos.*)
 - Voir présentation A. Napoli
 - Hiérarchisation des concepts et propriétés

Ontologie pos



Résultats (1)

- Production de *use cases* avec une ontologie UCD
 - Attribution d'UCD
 - Normalisation et validation
- Retombées sur l'IVOA
 - Aide pour l'adoption du standard UCD
 - Retour sur l'emploi des ontologies

UCD -> sémantique

- Extension du groupe IVOA UCD à un cadre plus large: sémantique liée à l'OV
- Répondre aux nouveaux besoins, ex: VO Event
 - Décrire les types d'objets astronomiques
 - Décrire les phénomènes physiques
- Travaux en cours...

Résultats (2)

- Construction d'une ontologie à partir du thésaurus et d'un corpus de documents
 - Collaboration CDS/IRIT
 - Voir présentation N. Hernandez
- Perspectives pour la fouille de textes
 - Souligne le besoin d'une bonne annotation de l'ontologie pour faire le lien avec les textes

Perspectives

- Extension de l'ontologie des UCDs à d'autres mots que pos.*
 - Aide à la normalisation, validation des UCD
 - Identification d'analogies
- Extension à tout le thésaurus UAI
- Importance de l'annotation dans la construction d'une ontologie
- IVOA: passage de UCD à Semantique
- VOTech: construction d'une ontologie des types d'objet astronomiques