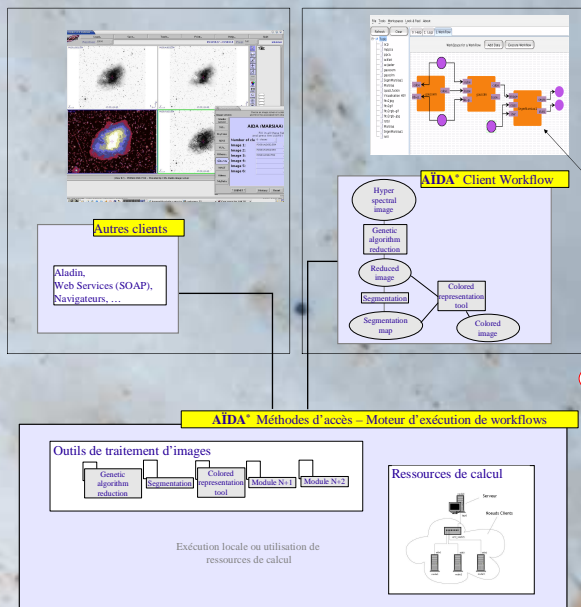


MDA: workflows et calculs distribués

A. Schaaff¹, F. Bonnarel¹, J.-J. Claudon¹, C. Pestel¹, C. Collet², M. Louys², C. Wolf², R. David², S. Genau², E. Slézak³
¹ Observatoire de Strasbourg - CDS, ² LSiIT Illkirch, ³ Observatoire de la Côte d'Azur - Cassiopee

De nombreuses personnes développent des outils pour le traitement d'images dans divers langages (C, C++, FORTRAN, MATLAB, etc.) mais ne les diffusent pas pour diverses raisons. D'une part pour des problèmes de portabilité et d'autre part en raison d'un manque d'interopérabilité avec d'autres outils. Pour apporter une solution à ces problèmes dans le contexte du projet MDA nous avons développé AÏDA (Astronomical Image processing Distribution Architecture), une architecture dans laquelle ces outils peuvent être intégrés, accédés suivant des protocoles standards (CGI, Web Services, ...) et assemblés dans des workflows.

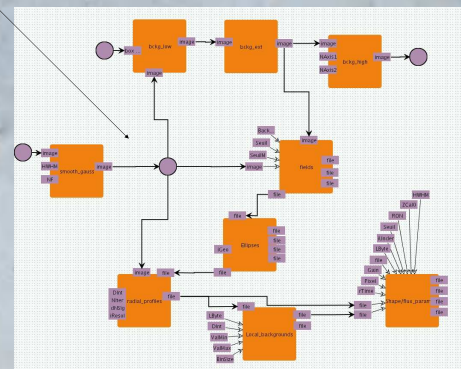


Un exemple de requête simple en imagerie monobande...

- **opérations** : détection et évaluation d'objets connexes dans une image mono-bande.
- **modèle sous-jacent** : tâches diffuses disjointes en émission sur un fond lentement variable sans défauts.
- **méthode adoptée** :
 - cartographie du fond ;
 - seuillage par segmentation ;
 - ajustement d'une ellipse de forme ;
 - évaluation du profil azimutal de brillance ;
 - calcul de mesures de forme et de flux.

JLOW (Java Libraries For Workflow)
Bibliothèque Java pour la représentation graphique des workflows
<http://cdsweb.u-strasbg.fr/cdsdevcomer/jlow.html>

...et sa solution avec l'outil AÏDA*...



... un pas vers la pérennisation des connaissances en traitement d'images des spécialistes de la communauté.

Continuité et perspective dans le cadre de l'Action Spécifique « Observatoire Virtuel » :

Initié en 2005, le groupe de travail Workflow a pour but de produire des cas d'utilisation dans plusieurs domaines (image, spectre, simulation, etc.) et de les confronter avec des outils, en particulier avec ceux de l'Observatoire Virtuel. Le développement de collaborations avec la communauté STIC (liens avec les grilles e.g. EGEE, ...) est également un objectif essentiel. Les premières réunions de travail ont été organisées conjointement avec le groupe de travail MDA. Un groupe de travail Grilles débutera en janvier 2007 par un recensement des utilisations des Grilles dans le domaine astronomique et des liens avec la communauté STIC.



Réunion thématique Grilles avec des intervenants des communautés STIC et astronomique (UK, Japon, Italie, etc.) organisée en 2005
http://cdsweb.u-strasbg.fr/meeting5/planning_en.htm

* développé dans le cadre du projet Masses de Données en Astronomie

MDA a été financé dans le cadre de l'ACI Masses de Données 2003

http://cdsweb.u-strasbg.fr/MDA/mda_en.html

Ontologies pour l'astronomie

S. Lesteven (1), S. Derriere (1), P. Dubois (1), F. Genova (1), N. Hernandez (2), J. Mothe (2), A. Napoli (3), A. Preite Martinez (4), A. Richard (4), Y. Toussaint (3)

(1) CDS – Strasbourg, (2) IRIT – Toulouse, (3) LORIA – Nancy, (4) INAF - Rome

Qu'est-ce qu'une ontologie ?

En informatique, une ontologie est un ensemble de concepts (avec une intension et une extension) et un ensemble de relations binaires entre les concepts (avec des domaines et des co-domaines) représentant les éléments du domaine, organisés hiérarchiquement dans une architecture reposant sur la relation de subsomption (classification partielle).

Les concepts et les relations sont codés dans un formalisme de représentation des connaissances (knowledge representation: KR) fournissant une syntaxe et une sémantique pour représenter les éléments du domaine et permettant des inférences.

Objectifs : aide aux astronomes et documentalistes

- Outil de recherche d'information et de ressources pertinentes, dans un volume croissant de données hétérogènes (catalogues, logs, articles, journaux, images ...) et dispersées sur le WEB, en tenant compte de contraintes complexes.

- Aide à l'annotation de ressources et de textes, pour améliorer les outils de la fouille de données.

Bénéfices d'une ontologie

Les ontologies permettent la mise au point d'une « terminologie commune » entre les humains et les agents logiciels, assurant une compréhension partagée des termes utilisés.

Le fondement logique (KR, logique de description) offre des possibilités d'inférence sur le contenu décrit.

Applications

Le thésaurus de l'UAI, les UCD et les types d'objets astronomiques de SIMBAD sont des points de départ intéressants pour la construction d'ontologies liées à l'astronomie.

Notre approche part d'ontologies sur des domaines restreints, avec des applications simples et ciblées. La mise en relation de ces ontologies permettra l'application à des problèmes plus complexes. Plusieurs stratégies de développement ont été suivies.

Thesaurus → Ontologie

Le Thésaurus Astronomique de l'UAI

- Créé en 1995 (Shobbrook R. and Shobbrook R.)
- 2863 termes, dont 2222 expressions
- 5000 relations (BT, NT, U, UF, RF)

MAGELLANIC CLOUDS

- UF clouds of magellan
- BT GALAXIES
- IRREGULAR GALAXIES
- LOCAL OF GALAXIES
- NEARBY GALAXIES
- NT LARGE MAGELLANIC CLOUD
- SMALL MAGELLANIC CLOUD
- RT MAGELLANIC STREAM

- PT : Primary term**
U : USE
 UF : use for
BT : BRODER TERM
NT : NARROWER TERM
RT : RELATED TERM
SN : Scope Note

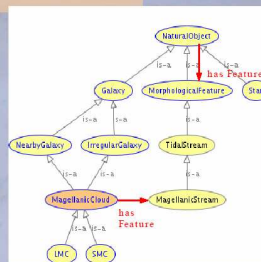
BT and NT : expriment la subsomption (+validation)

RT : détaillé avec des propriétés spécifiques

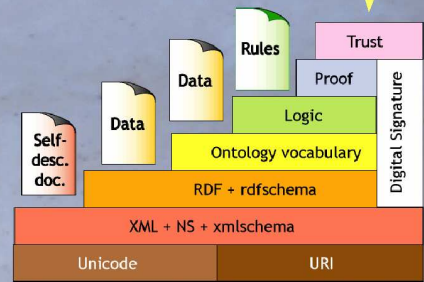
(creation de nouvelles relations entre concepts pour affiner la relation générale « is related to »)

VALIDATION DES 12 Concepts principaux

PROPERTY	RELATION
PHENOMENON	ANGLE
EVENT	PLANE
SCIENCE	REGION
INSTRUMENTATION	NATURAL OBJECT
SUBSTANCE	ARTEFAC



Knowledge description	Users	Types of relations	Technical tools
Thesaurus	Humans	. Equivalence . Hierarchy . Association . Notes	Relational Databases
Taxonomies	Humans Machines	Hierarchy	WEB
Ontologies	Machines Humans	. Concepts organized hierarchically (subsumption) and semantically . Creation of properties used to characterize concepts . Restrictions on properties give "defined" concepts (set of necessary and sufficient conditions)	Semantic WEB



The « Semantic Web cake »

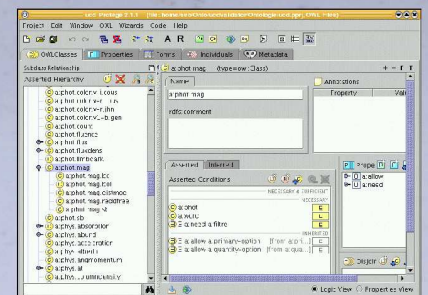
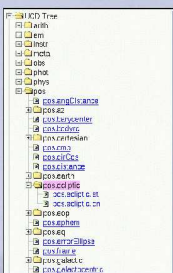


UCDs → Ontologie

UCDs : Unified Content Descriptors

- Description standardisée du contenu des tables.
- Proviennent de l'analyse d'un grand nombre de descriptions de catalogues astronomiques.
- Organisation hiérarchique, ~ 450 mots.
- Révisé par l'IVOA pour les projets d'Observatoire Virtuel.

- Règles syntaxiques pour construire des UCD à partir des mots
- Validation des UCD
- Application à l'assignation des UCD : trouver l'UCD correspondant à une description

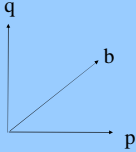


Réduction et segmentation de cubes de données hyperspectraux

F. Bonnarel¹, B.vollmer¹, C. Collet², F.Flütt², M.Louys², M.Petremand^{1,2}, A.Bijaoui³, C.Benoist³, E.Slězak³
¹ Observatoire de Strasbourg - CDS, ² LSiIT Illkirch, ³ Observatoire de la Côte d'Azur - Cassiopée

- Considérant un cube de données hyperspectral comme une image de spectres, le but de ce travail est d'extraire des données des comportements spectraux communs et de combiner l'information spatiale et spectrale. Lors d'une telle analyse sur les cubes de données, dont les dimensions atteignent des valeurs drastiques (typiquement 1000x1000x1000 échantillons) on est confronté à la malédiction de dimensionalité. C'est pourquoi nous proposons d'effectuer dans une première étape une procédure itérative de réduction de dimensionalité basée sur l'approche des « Meanshift ».
- Connaissance *a priori* → une base de référence de spectres. Après avoir projeté chaque spectre de l'observation sur cette base, on obtient un ensemble de poids de projection. L'algorithme des Meanshift est alors utilisé pour déterminer les modes de la distribution des poids de projection. Ces modes composent ainsi une nouvelle base de projection et le processus est réitéré. Cette approche a été testée avec succès en utilisant une mesure d'angle spectral au lieu d'une projection.
- Cette nouvelle méthode est en cours de validation sur des champs de galaxies simulées → GALICS (Formation hiérarchique des galaxies- IAP)

Image Hyperspectrale



- L'observation hyperspectrale Y est composée de $p \times q \times b$ pixels. Chaque site s définit un spectre Y_s .
- n spectres choisis au hasard dans l'observation composent la première base de projection B .
- Lorsque les comportements spectraux présents dans l'observation sont connus *a priori*, la base peut alors être constituée de spectres de référence choisis au sein d'une collection de spectres.

• $B = \{\beta_1 \dots \beta_n\}$ est la base de projection. Y_s est alors projeté sur B

$$Y_s = \sum_{i=1}^n \omega_{si} \times \beta_i$$

- Les poids ω_{si} définissent une distribution de points dans l'espace de projection.

• Comme la projection dépend de la luminosité, on peut aussi utiliser une mesure d'angle entre les spectres :

$$\alpha(\beta_i, Y_s) = \arccos \left(\frac{\langle \beta_i, Y_s \rangle}{|\beta_i| |Y_s|} \right)$$

- Cette mesure est invariante en luminosité et permet une discrimination des spectre basée sur leur signature spectrale.

Projection ou angle de chaque spectre sur une base de référence

ω_{si}

Nouvelle base de projection

Meanshift

modes

- Méthode des meanshift : une méthode non paramétrique pour déterminer les modes (ou maxima) de la densité de probabilité f d'une distribution de points. Cette méthode est basée sur l'estimation du gradient de f : $\nabla f(x)$

- Utilisée dans l'espace des poids de projections ou angles pour chaque vecteur de poids/angle → Chaque vecteur est associé à un mode.
- Les spectres de l'observation associés aux modes produisent une nouvelle base de projection. La projection et l'étape des meanshift sont réitérés jusqu'à la convergence (même base durant deux étapes).
- Ces modes sont caractéristiques des comportements spectraux présents dans l'image → nombre de classes spectrales.

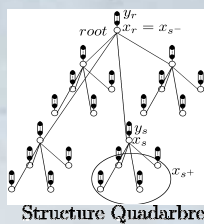
number of classes ω_{si}

- Lorsque $n < 10$ on réalise une segmentation dans l'espace de projection à l'aide du quadarbre markovien conduisant ainsi à une carte de segmentation X . Cette méthode introduit une régularisation spatiale et regroupe les poids ou les angles en régions homogènes.

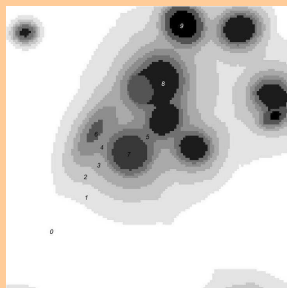
- Quand $n > 10$ on utilise l'algorithme de regroupement des K-moyennes pour la segmentation. En effet le modèle Markovien ne peut être mis en oeuvre au delà de $n > 10$ en raison de la "malédiction de dimensionalité".

- Le nombre de spectres dans la base finale est une estimation du nombre de classes pour l'étape de segmentation. Ce nombre peut également être fixé par l'utilisateur.

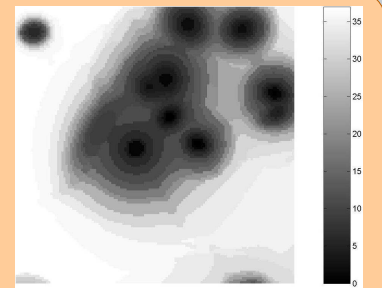
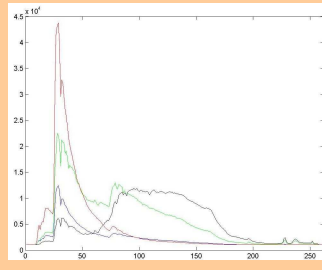
Segmentation multi-composante dans l'espace de Projection : K-Moyennes ou segmentation markovienne



2 bandes tirées du cube hyperspectral original (128 x 128 x 262 pixels). La luminosité de l'objet décroît radialement. Chaque objet présente une variation de flux suivant la longueur d'onde.



La carte de segmentation (10 classes) obtenues avec la méthode des angles. La mesure de l'angle spectral extrait seulement les classes spectrales. Au centre, la classe *bulbe* domine alors qu'à la périphérie on trouve la classe *disque*. A droite de la carte de segmentation on peut voir les spectres moyens pour les classes 7 (bleu), 8 (noir), 9 (vert), 10 (rouge).



La carte de segmentation (38 classes) obtenues avec la méthode de projection. Dans un objet, l'ensemble des classes est constitué d'une séquence de classes de luminosité et d'une composition spectrale bulbe-disque.

MDA a été financé dans le cadre de l'ACI Masses de Données

2003

http://cdsweb.u-strasbg.fr/MDA/mda_en.html