

Nouvelle interface VizieR

Gilles Landais (CDS)

10 décembre 2010





- 1 Objectif de la nouvelle version
- 2 Critère d'accès aux données et statistiques d'utilisation
- 3 Les nouveautés
 - Accès aux données (navigation dans le service)
 - La recherche textuelle
 - Evolutions et nouvelles fonctionnalités
 - Futures évolutions
- 4 Démo



Une nouvelle version pour une meilleure ergonomie

- Simplifier l'existant, notamment pour un public non avertit
- Améliorer la recherche de données (*recherche par contenu (UCD)*)
- Renouveler l'interface utilisateur (vitrine du service)
- Poursuivre/améliorer des fonctionnalités existantes et apporter de nouvelles fonctionnalités (*requêtes par liste, crossmatch*)



Ce n'est pas un VizieR2.0

(le coeur, la base de données et le processus d'installation de nouveaux catalogues sont inchangés)



Garder en l'esprit

- Une recherche idoine au langage scientifique précis
- Clareté dans la recherche

Statistiques générales



Statistiques sur les 28 derniers jours

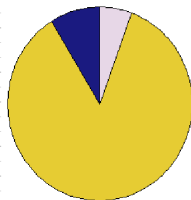
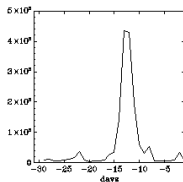
nombre total de requêtes	16 330 000
moyenne par jour	583 000
nombre minimal de requêtes	53 000
nombre maximal de requêtes	4 356 000
nombre moyen de sites	1100

Requêtes par liste : (sur 5 jours)

par position	3189
par contraintes	122

Statistiques pour une journée type

VizieR	30508
VOTable	295620
ASU	19149
vizExec	112

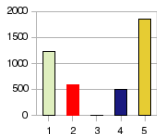
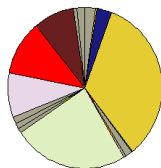


Critères d'accès

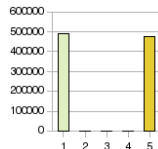
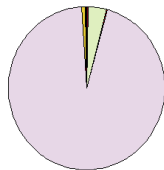
Répartition des critères : *en provenance de HTTP pris sur 5 jours*

center	1109
keyword	103
source	790
words	364
keyword center	276
source center	289
words center	72
source words center	52
words keyword	49
source words	36
source keyword	25
words keyword center	13
source words keyword center	8
source keyword center	30
source words keyword	2
ucd	1

sorties HTML issuent du 1er formulaire



toutes sorties confondues



source words ucd keyword center

Accès aux données ... (suite)



On distingue trois méthodes de recherches :

1 la recherche quantitative

recherche par position, par source (dimensions dans VizieR)

2 la recherche thématique

guidée par des mots clés prédéfinis

3 la recherche textuelle

plus large, mais soumis aux ambiguïtés du langage



Différents moyens de recherche pour un même mot clé :

- la position et un mot clé (ex : M31)
- un nom de catalogue et un mot clé (ex : 2MASS)



- 1 L'accès aux données (navigation dans le service)
- 2 La recherche textuelle
- 3 Evolutions et nouvelles fonctionnalités



Nature de la recherche dans la base de données VizieR :

1 Recherche de catalogues

(puis visualisation des données l'intérieur de ceux-ci)

2 Recherche de données *(quelqu'en soit l'origine)*

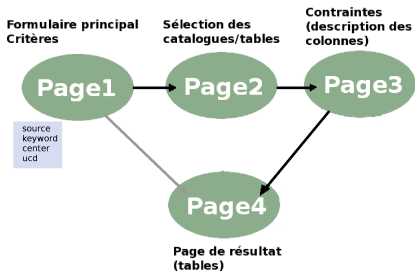
Nécessite une recherche de type quantitative sur des données unifiées

⇒ essentiellement limitée à la recherche positionnelle dans VizieR

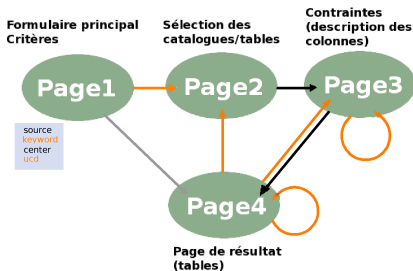


L'hétérogénéité des catalogues ont comme conséquence une dominance de la recherche de catalogues

Navigation dans le service



Navigation dans le service



1 refonte du formulaire de la page d'accès a VizieR :

Formulaire dynamique estimant le nombre de catalogues correspondant aux éléments sélectionnés, aperçu des catalogues les plus populaires (Ajax)

2 la barre de navigation

(récapitulatif des choix antérieur, adaptation de la page en cours)



Synthèse sur une recherche par mot clé

- 1 recherche si c'est un catalogue
- 2 recherche dans les acronymes
(DM = Durchmusterung (I/119), 2MASS = II/241,II/246,VII/233, Pleiades = I/90)
- 3 recherche dans les mots clés, auteurs, ... des **readme**
fournit par les auteurs ou par le CDS (les documentalistes)
- 4 recherche textuelle dans la description des catalogues
- 5 **nouveau** : recherche textuelle dans la description des colonnes

Recherche textuelle par UCD



UCD - Définition

Unified Content Descriptors

(définition de l'IVOA) Glossaire hiérarchique de la signification scientifique contenues dans les tables astronomiques

Historique :

- **UCD1** : créé au CDS et répondant à une logique ensembliste

ex : PHOT_JHN_K (magnitude K)

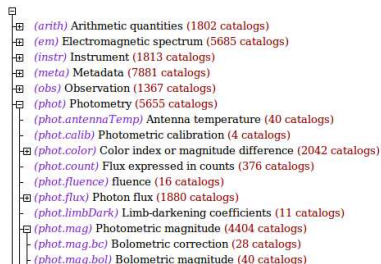
- **UCD1+** : logique linguistique, plus précise

ex : phot.mag ; em.IR.K

L'organisation des UCD :
hiérarchique à deux niveaux :

ucd = entité1[;entité2[;entité3]]

où : entitéX=mot1[mot2[...]]



Recherche textuelle par UCD ...(suite)



Traduction du texte libre en définition standardisé (UCD) :

⇒ **Dictionnaire**

Dictionnaires utilisés par ordre de pertinance :

- 1** dictionnaire d'acronyme
- 2** les descriptions des UCD
- 3** les descriptions des colonnes

Recherche textuelle par UCD ...(suite)



Traduction du texte libre en définition standardisé (UCD) :

⇒ Dictionnaire

Dictionnaires utilisés par ordre de pertinence :

- 1 dictionnaire d'acronyme
- 2 les descriptions des UCD
- 3 les descriptions des colonnes
 - 1 Recherche du texte dans les colonnes
 - 2 regroupement des catalogues trouvés par UCD
 - 3 comparaison du nombre de catalogues trouvé pour un UCD avec le nombre total de catalogue incluant cet UCD (pertinence)

Recherche textuelle par UCD ...(suite)



Traduction du texte libre en définition standardisé (UCD) :

⇒ Dictionnaire

Dictionnaires utilisés par ordre de pertinence :

- 1 dictionnaire d'acronyme
- 2 les descriptions des UCD
- 3 les descriptions des colonnes
 - 1 Recherche du texte dans les colonnes
 - 2 regroupement des catalogues trouvés par UCD
 - 3 comparaison du nombre de catalogues trouvé pour un UCD avec le nombre total de catalogue incluant cet UCD (pertinence)



La liste de définition standardisée est classée par pertinence. Elle peut aboutir en fin de liste à des résultats « étrange »

Evolutions et nouvelles fonctionnalités



■ annotation

peut se faire au niveau du catalogue, de la table ou de la donnée

■ liste

nouvelle possibilité d'affichage en une seule table

⇒ fusion (jointure) lorsque plusieurs tables sont sélectionnées

■ crossmatch

limité, intéressant pour de petites volumétries (vitesse de croisières : 100 lignes par seconde pour un crossmatch contenant le catalogue tycho (2500000 sources))

■ samp (VO)

module développé à Paris, permet d'envoyer simplement le résultat affiché à une autre application du VO

Futures évolutions



- ajout des **SED** (*spectral energy distribution*) pour uniformiser les tables incluant des magnitudes (\Rightarrow nouvelle dimension dans VizieR)
- **TAP** (*Table Access Protocol (VO)*), avec **ADQL** (*langage de requêtes SQL incluant des fonctions astronomiques*)



- URL b ta : <http://viz-beta.u-strasbg.fr>
- Mise en production pr vue pour le mois de janvier *(apr s l'AAS)*
- Remont es d'anomalies ou suggestions :
 - mail gilles.landais@astro.unistra.fr
 - francois.ochsenbein@astro.unistra.fr



1 Réutilisation des résultats dans VizieR :

Recherche d'une zone HII au centre galactique (IR)

→ récupération des résultats dans le catalogue IRAS

(catalogue J/A+AS/115/81)

- recherche par position
- recherche par identifiant
- crossmatch des deux catalogues

2 Utilisation des UCD :

Recherche à la position $0 + 0$ (J2000) de catalogues contenant des informations de redshift et vitesse (en optique)

(catalogue VII/259)

- crossmatch (2deg)
- jointure dans la même zone
- envoi du a Aladin (samp)