



**INFO**  
Département Informatique  
IUT Belfort-Montbéliard



# Chatter en langage naturel avec les services de données astronomiques du CDS

## RAPPORT DE STAGE

*Du 8 avril au 14 juin 2019*

**Centre de Données astronomiques  
de Strasbourg  
Observatoire astronomique de  
Strasbourg**

André SCHAAFF, Ingénieur de recherche  
Sébastien DERRIERE, Astrophysicien  
Thomas BOCH, Ingénieur de recherche

**Antoine HERKENS S4 A1**

**DUT Informatique 2ème année**

Michel SALOMON, Professeur à l'IUT  
de Belfort



Observatoire astronomique  
de Strasbourg





# Remerciements

Je remercie chaleureusement mon maître de stage André SCHAAFF, ingénieur de recherche, pour m'avoir guidé au cours de ma mission à l'observatoire. Par la même occasion, je tiens à remercier mes co-encadrants Sébastien DERRIERE, astrophysicien et Thomas BOCH, ingénieur de recherche, dont les indications et les connaissances m'ont permis de mieux comprendre les notions astronomiques tout comme le fonctionnement des outils informatiques présents au CDS et qui ont de fait contribué à la réussite de mon stage.

Je suis également reconnaissant envers M. Pierre-Alain DUC, directeur de l'observatoire astronomique de Strasbourg, ainsi qu'envers M. Mark ALLEN, directeur du Centre de Données astronomiques de Strasbourg, pour m'avoir permis d'effectuer mon stage au sein de cette unité mixte de recherche. En plus du cadre de travail très agréable, j'ai eu la chance de pouvoir assister à tous les évènements organisés pour les équipes de l'établissement (visites, séance de planétarium, réunions, séminaires).

Je remercie également Michel Salomon, enseignant tuteur à l'IUT qui a su répondre à mes questions et m'a guidé au cours de mon stage, ce qui a aussi conduit à sa bonne réalisation.

# SOMMAIRE

Remerciements.....	7
Introduction.....	9
<b>1 Présentation de l'organisme.....</b>	<b>10</b>
<b>1.1 Historique et structure de l'observatoire.....</b>	<b>10</b>
<b>1.2 Centre de Données astronomiques de Strasbourg.....</b>	<b>11</b>
1.2.1 Simbad.....	11
1.2.2 VizieR.....	12
1.2.3 Aladin.....	12
<b>1.3 Ma place au sein de l'observatoire.....</b>	<b>13</b>
<b>2 Présentation du sujet.....</b>	<b>14</b>
2.1 Contexte et utilité d'un chatbot pour le CDS.....	14
2.2 Fonctionnement schématique du chatbot.....	15
2.3 Présentation de Dialogflow.....	15
2.4 Objectifs et enjeux du stage.....	17
2.5 Problèmes et avantages.....	19
<b>3 Cahier des charges.....</b>	<b>20</b>
3.1 Descriptif fonctionnel.....	20
3.2 Contraintes techniques.....	20
<b>4 Mise en œuvre.....</b>	<b>21</b>
4.1 Missions.....	21
4.2 Organisation.....	23
<b>5 Bilan.....</b>	<b>24</b>
5.1 Bilan pour le CDS.....	24
5.2 Bilan humain.....	24
5.3 Bilan pédagogique.....	24
Conclusion.....	26
Sitographie.....	28

# Introduction

Dans le monde de la recherche en astronomie, l'accès aux données scientifiques est crucial. La majorité des chercheurs ne font plus d'observations eux-mêmes mais s'appuient sur la quantité colossale de données et publications pour faire leurs propres recherches. Effectivement, les projets réalisant des observations nécessitent des budgets colossaux, qu'elles soient depuis la Terre ou l'espace. Les données obtenues par ces projets sont par la suite publiées, informatisées et stockées, pour que la communauté scientifique internationale puisse y avoir accès librement et gratuitement. Ceci ne se fait que très peu dans les autres domaines scientifiques. La réutilisation des données permet d'une part de rentabiliser les projets, mais surtout de faire progresser le domaine plus rapidement. L'enjeu majeur est donc d'avoir un accès rapide et précis pour obtenir les données souhaitées.

L'observatoire astronomique de Strasbourg, avec le Centre de Données astronomiques de Strasbourg possède une quantité très importante de données grâce à leur ancienneté. Le Centre de Données astronomiques n'échappe pas au fait que l'accès à l'information doit être le plus rapide et facile possible. Jusqu'à maintenant, cela se fait en remplissant des formulaires parfois assez long précis.

L'objectif de mon stage est donc de faciliter cet accès à l'information en continuant le développement d'un chatbot, ou assistant virtuel en français. Il permettra à un astronome d'interroger les services du Centre de Données astronomiques à travers cet assistant, en ayant une conversation naturelle. L'avantage de l'assistant est donc la simplicité d'utilisation, car il suffit d'écrire en langage parlé ce que l'on souhaite obtenir contrairement à un formulaire avec beaucoup de paramètres à préciser.

L'offre de stage m'a fortement intéressée, car les assistants virtuels sont courants et toujours plus utilisés actuellement. Cela m'a fortement motivé de contribuer à un projet moderne et novateur car spécifique à l'astronomie en tant que développeur.

Je m'intéresserai d'abord au contexte de l'institut de recherche et à la présentation de mon sujet, puis je verrai quels sont les besoins du cahier des charges avant d'analyser la mise en œuvre et de dresser le bilan de mon stage.

# 1 Présentation de l'organisme

## 1.1 Historique et structure de l'observatoire

L'observatoire astronomique de Strasbourg, situé en plein cœur du campus universitaire, est un établissement construit en 1881 par l'empire allemand, dans le but d'exposer son intérêt et son avancement en astronomie. Actuellement, il n'y a plus d'observations scientifiques faites à l'observatoire bien qu'il héberge la troisième plus grosse lunette astronomique de France.

L'observatoire est reconnu en tant qu'unité mixte de recherche (UMR), partagée entre le CNRS et l'Université de Strasbourg. Au total, environ 80 personnes travaillent à l'observatoire, principalement des chercheurs (enseignants ou non), des documentalistes, des ingénieurs et des astronomes. L'observatoire héberge le Centre de Données astronomiques de Strasbourg (CDS), très utilisé mondialement et qui contribue fortement à la renommée de l'établissement grâce aux services gratuits libres d'accès.

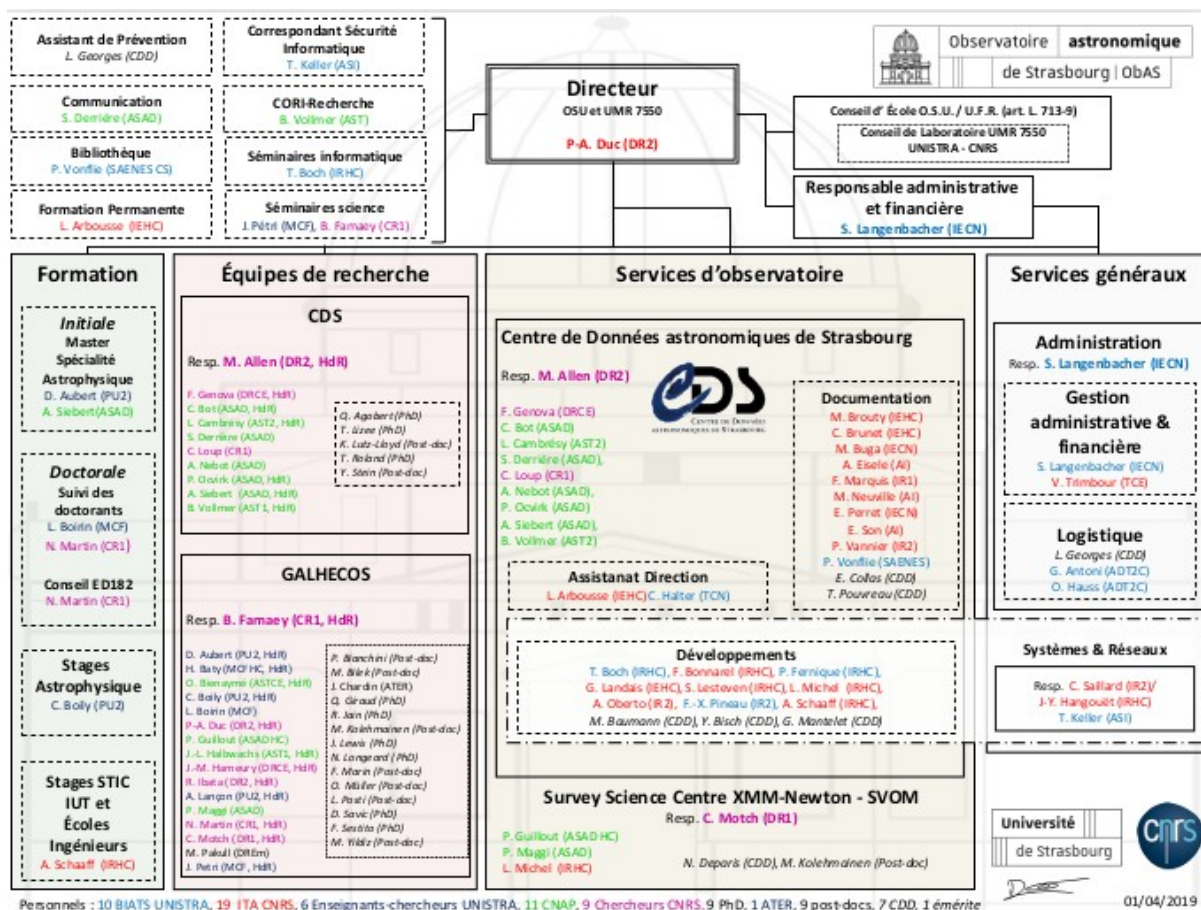


Illustration 1: Organigramme de l'observatoire de Strasbourg

## 1.2 Centre de Données astronomiques de Strasbourg

Le Centre de Données astronomiques de Strasbourg, appelé dans la suite du rapport «CDS», désigne l'équipe de recherche travaillant à l'archivage, le traitement et la mise à disposition de données astronomiques à l'international. Ayant commencé dès 1972 à informatiser les données et à en accumuler une quantité colossale, les services du CDS sont très utilisés à l'international de nos jours. Par ailleurs, en France, il n'existe que quelques centres de données astronomiques, mais bien moins importants que celui de Strasbourg.

Le CDS est composé d'environ un tiers d'informaticiens, d'un tiers de documentalistes et d'un tiers d'astronomes. Ensemble, ils veillent au bon fonctionnement des trois principales plateformes mises à disposition par l'observatoire à savoir Simbad, VizieR et Aladin, mais aussi des nombreux outils gravitant autour.

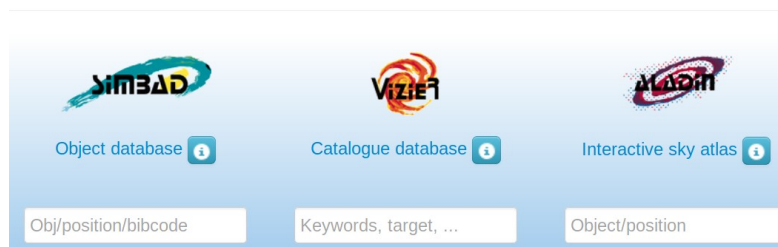


Illustration 2: Les trois principaux services du CDS : Simbad, VizieR et Aladin

Le CDS s'impose comme un acteur très important de l'IVOA, l'International Virtual Observatory Alliance (association internationale développant des standards et des normalisations dans le traitement et la diffusion des données afin d'assurer l'interopérabilité des services astronomiques).

### 1.2.1 Simbad

Simbad est une base de données créée, alimentée et maintenue par le CDS. Elle contient en 2019 plus de 10 millions d'objets astronomiques (étoiles, galaxies, nébuleuses, ...) situés en dehors du système solaire et est requêtée environ 500 000 fois par jour. Quand un objet est entré dans Simbad, les différentes mesures faites sur lui sont renseignées et sont donc accessibles rapidement pour les astronomes. Pour accéder à celles-ci, une interface web est proposée. Grâce à plusieurs critères qui lui sont proposés (nom de l'objet, position dans le ciel, référence bibliographique, etc), l'utilisateur peut récupérer les mesures entrées dans la base. Il existe également un moyen plus technique

utilisant l'ADQL, une version modifiée du SQL<sup>1</sup> adaptée aux besoins de l'astronomie. Grâce à l'ADQL, on peut récupérer dans une application les données de Simbad structurées dans un format compréhensible par la machine (comme le JSON ou le XML).

Simbad est notamment utilisée par d'autres services, comme NED (base de données de la NASA) ou bien des moteurs de recherche scientifiques comme Wolframalpha tant la quantité de données est importante et pertinente.

### M 31 -- Galaxy

```

Other object types:          LIN (), G (Ref,LEDA,...), QSO (Ref,[W2006],...), AGN ([W2000c],[W2003c],...), Rac
ICRS coord. (ep=J2000) :    00 42 44.330 +41 16 07.50 (Infrared) [ ] C 2006AJ....131.1163S
FK4 coord. (ep=B1950 eq=1950) : 00 40 00.092 +40 59 41.73 [ ]
Gal coord. (ep=J2000) :    121.174322 -21.573311 [ ]
Radial velocity / Redshift / cz : V(km/s) -300.0 [4.0] / z(spectroscopic) -0.001000 [0.000013] / cz -299.85 [4.00]
                               C 2012AJ....144....4M
Morphological type:        SA(s)b D 2013AJ....146...67B
Angular size (arcmin):    44.673 25.062 45 (IR) D -
Fluxes (6) :              U 4.86 [0.03] D 2007ApJS..173..185G
                               B 4.36 [0.02] D 2007ApJS..173..185G
                               V 3.44 [0.03] D 2007ApJS..173..185G
                               J 2.094 [0.016] C 2006AJ....131.1163S
                               H 1.283 [0.017] C 2006AJ....131.1163S
                               K 0.984 [0.017] C 2006AJ....131.1163S

```

Illustration 3: Exemple et extrait de résultats obtenus avec Simbad lorsqu'on entre la galaxie « m31 »

## 1.2.2 VizieR

VizieR correspond à une base de données contenant plus de 18 000 catalogues (publications) astronomiques informatisés. C'est comparable à une bibliothèque où les livres sont synthétisés pour ne garder que les données et leurs descriptions. Cette base de données est très complexe. Pour imager, elle est composée elle-même de base de données, qui répertorient les informations importantes des articles scientifiques. Ces derniers ayant un contenu pouvant beaucoup varier dans leur forme, cela rend leurs traitements très compliqués. Par conséquent, une requête à VizieR est plus longue qu'une à Simbad, c'est pourquoi Simbad est requêtée en premier par défaut, puis ensuite VizieR si l'information n'a pas été trouvée.

## 1.2.3 Aladin

Aladin est un atlas interactif du ciel donnant accès à plus de 5 téraoctet d'images et servant de portail à l'observatoire virtuel. Ce service est très visuel mais également compliqué à mettre en place à cause de la taille des images qu'il faut traiter de nombreuses fois avant de pouvoir les afficher dans un navigateur web.



Illustration 4: Image de la galaxie "m31", affichée avec Aladin

1 Langage informatique permettant de récupérer des données dans une base de données « classique »



## 1.3 Ma place au sein de l'observatoire

En tant que stagiaire, j'ai pu travailler au sein de l'observatoire durant 10 semaines. Étant une UMR, il y a donc de la recherche et du développement, esprit bien représenté par le bureau dans lequel j'ai travaillé avec André SCHAAFF (ingénieur de recherche) et Sébastien DERRIERE (astrophysicien).

L'endroit était idéal, car je pouvais communiquer facilement avec mes collègues qui ont su me transmettre leurs connaissances pour M. DERRIERE plus particulièrement en astronomie et M. SCHAAFF en informatique. Thomas BOCH était lui joignable par mail et a pu me donner des précisions sur le fonctionnement d'Aladin.

Ainsi, j'ai pu me familiariser rapidement avec les notions astronomiques nécessaires au développement du chatbot et comprendre les besoins qu'un chercheur aurait à utiliser mon outil. Mon stage s'est déroulé dans un atmosphère conviviale et professionnelle.

## 2 Présentation du sujet

### 2.1 Contexte et utilité d'un chatbot pour le CDS

Le CDS propose donc de nombreux services, tous sophistiqués et plus ou moins complexes à utiliser. En plus de Simbad, Vizier et Aladin, il existe beaucoup d'autres services tels que Sesame (pour résoudre des noms d'objets), X-match (similitudes entre différents catalogues), alasky (serveur pouvant prendre en compte en multitude de requêtes variées) etc.

Dans le but de rendre l'utilisation des services plus simple et centralisée, une partie de l'équipe de recherche et développement (constituée d'André SCHAAFF, Sébastien DERRIERE et Thomas BOCH) envisage l'idée d'un chatbot. Une fois abouti, il serait mis sur le site public du CDS et permettrait à tout utilisateur d'avoir un accès facilité à l'information des services regroupée à un endroit, en plus d'être naturel et moderne.

Le développement a débuté en 2018 par un stagiaire d'IUT (Alexis Guyot) et a été continué par moi-même. Nous avons tous deux été encadrés par nos tuteurs cités auparavant, qui nous ont donné les objectifs du chatbot et transmis les notions d'astronomie nécessaires au bon développement.

Le travail que j'ai fournis pour l'observatoire lui est bénéfique et aussi pour moi personnellement. Effectivement, les employés sont très occupés et peuvent se concentrer sur leurs projets en cours prioritaires tandis que je développe, teste, et améliore un projet encore jeune. De mon côté, j'acquiers de l'expérience professionnelle et approfondis mes connaissances informatiques.

L'objectif du chatbot est de réunir tous les services du CDS (principalement Simbad, Vizier et Aladin) au sein d'une application web. L'astronome peut donc poser sa demande en langage naturel (comme s'il parlait ou rédigeait un texte) à l'assistant virtuel, qui s'assure de trouver le.s service.s adapté.s pour renvoyer une réponse à l'utilisateur.

L'avantage de cet outil est sa facilité d'utilisation. Effectivement, l'utilisateur n'a qu'à écrire sa demande dans un champ de texte, et tout le traitement (savoir quel service demander, paramètres particuliers...) est fait par le chatbot. Cela évite donc d'utiliser des formulaires parfois longs et complexes (voir en annexe), qui prennent du temps à remplir à chaque nouvelle demande.

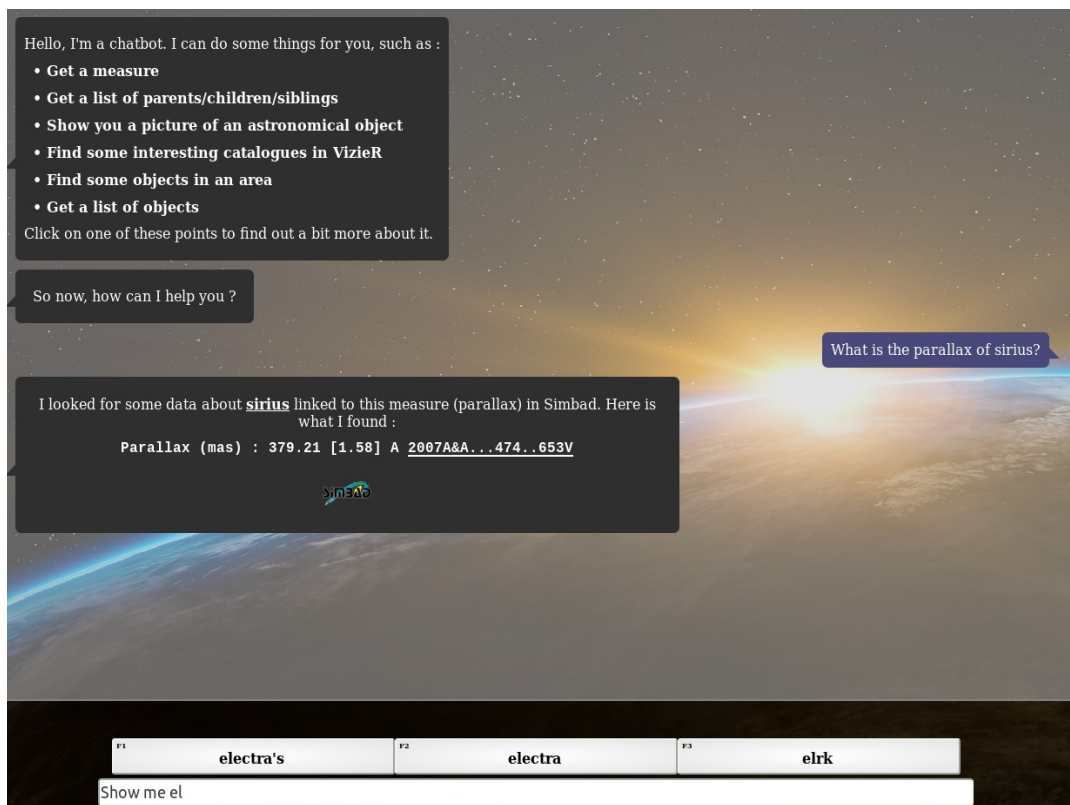
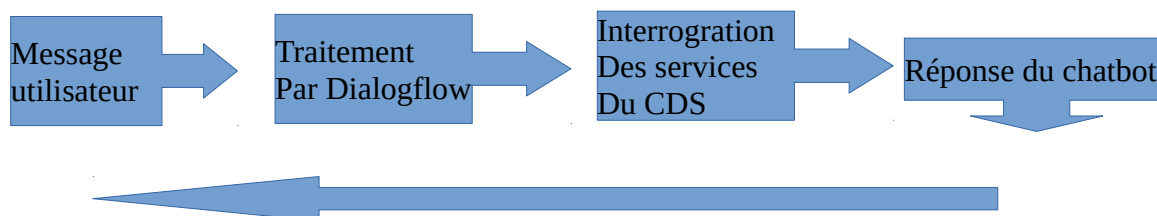


Illustration 5: Exemple d'utilisation du chatbot

À la fin du stage d'Alexis Guyot en 2018, le chatbot pouvait traiter une partie des requêtes possiblement demandées et avait les fonctionnalités de bases pour répondre sans trop de précisions aux demandes d'astronomes, tout cela avec une interface graphique simple et efficace.

## 2.2 Fonctionnement schématique du chatbot



## 2.3 Présentation de Dialogflow

Comme pratiquement tout moteur de traitement du langage naturel, Dialogflow propose une compréhension du langage basée sur un système d'intentions et d'entités. L'intention, comme son nom le suggère bien, est l'attente qui a poussé l'utilisateur à formuler

sa requête à l'oral. Par exemple, l'intention cachée derrière une phrase du type « Quelle est la température effective de Sirius? » est d'obtenir une mesure concernant un objet. Les entités correspondent aux éléments clés de la demande, ceux qui auront une importance pour formuler une réponse pertinente et cohérente. Dans notre précédent exemple, on discerne deux entités : « température effective » qui correspond à la mesure attendue, et « Sirius » qui est le nom de l'objet à étudier. Ces deux termes sont indispensables pour la formulation d'une réponse. Dialogflow décrit donc une phrase en catégorisant son intention, et en mettant de côté ses entités. Après traitement de la phrase d'exemple, le bot gardera seulement à l'esprit que la personne qui lui parle veut obtenir une mesure, la température effective, d'un objet qui s'appelle Sirius. Il stockera ensuite ce constat dans une structure JSON qu'il retournera à l'utilisateur.

Au départ, il demande au développeur de lui montrer comment reconnaître une intention et ses entités sur quelques exemples. A partir de cela, il va commencer le processus de *machine learning*. Afin de minimiser son taux d'erreur, il va également essayer de reconnaître des mots souvent utilisés dans les exemples donnés par le développeur, jusqu'à ce que son algorithme de *machine learning* soit assez performant pour ne plus avoir à baser une grosse partie de son jugement sur le *matching*. La plateforme en ligne de Dialogflow permet d'effectuer l'étape dite « d'entraînement » durant laquelle le développeur donne son avis sur les réponses de l'intelligence artificielle.

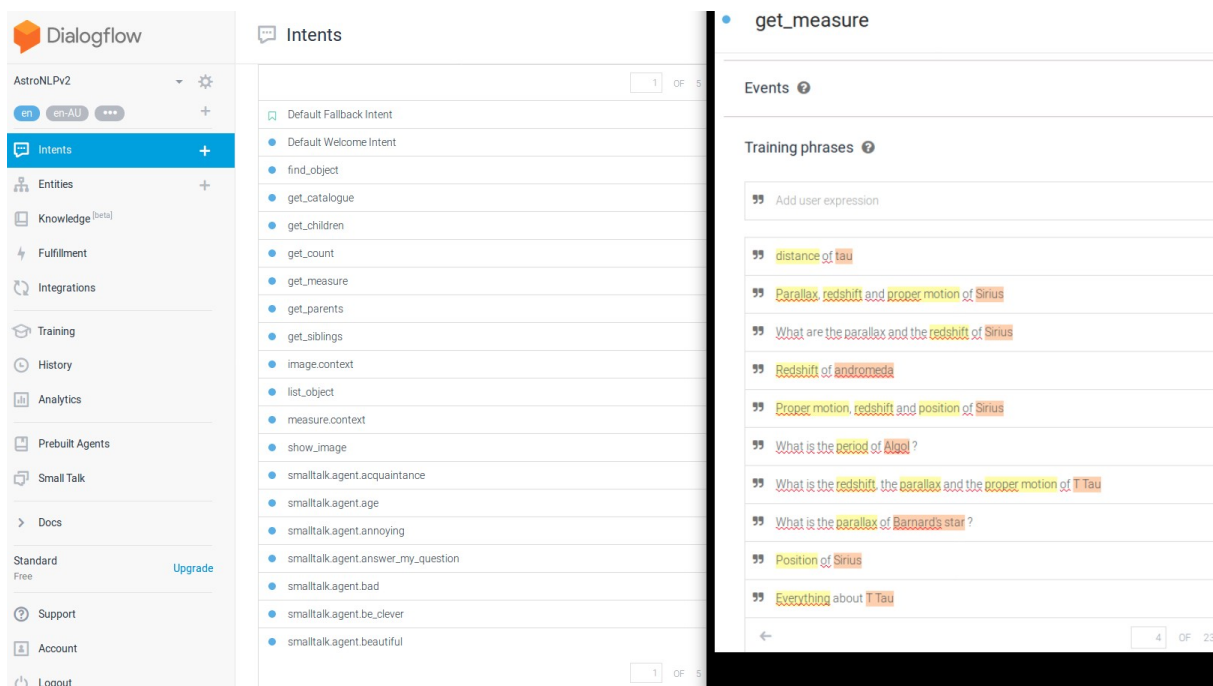


Illustration 6: Agent dialogflow créé en fonction de nos propres besoins

À gauche, les intentions. À droite, un exemple de phrases d'entraînements associées à une intention. Grâce aux phrases d'apprentissage, l'agent s'entraîne automatiquement grâce à une intelligence artificielle gérée par Google (aucun accès à l'intelligence artificielle n'est possible).

Dans un second temps, une fois toutes les informations récupérées, il faut interroger les services adéquats du CDS pour pouvoir retourner la réponse souhaitée à l'utilisateur. C'est cette étape qui est la plus longue et complexe, car pour obtenir une information, il faut interroger plusieurs services. Cela implique donc beaucoup de traitements asynchrones, des requêtes imbriquées et très diversifiées (TAP, ADQL, ajax, xmlhttprequest...), avec en plus un comportement le plus cordial et humain possible envers l'utilisateur.

## 2.4 Objectifs et enjeux du stage

La qualité de mon stage réside dans le fait que les assistants virtuels sont très présents de nos jours et donc en développer un est moderne, particulièrement dans le domaine de l'astronomie. De plus, le fait de travailler dans un domaine scientifique que je connaissais très peu est personnellement très enrichissant car je dois me familiariser rapidement avec les termes, les besoins des astronomes/astrophysiciens afin de débiter le développement avec les bonnes notions.

La faculté du chatbot à comprendre les intentions des scientifiques, tout comme la faculté à présenter des résultats pertinents relèvent du défi : si l'assistant n'est pas efficace et pratique, il peut conduire soit à fausser les recherches, soit à abandonner l'outil car pas assez compétent. En tant que développeur, c'est donc une responsabilité importante que j'ai de produire un chatbot fonctionnel. Ce défi constitue aussi la difficulté du stage, et en développer la solution le rend très passionnant.

Mon rôle pour ce stage est de perfectionner l'outil qu'à développé Alexis, pour rendre l'outil plus fonctionnel. Les objectifs de mon stage déterminés par André SCHAAFF sont donc les suivants :

- effectuer une migration importante de version de l'outil Dialogflow (de la version 1 à 2), impliquant un changement majeur d'architecture dans le code. Cela permet de garder le chatbot à jour, car les services d'anciennes versions de Google ne sont plus opérationnels à partir d'une date limite.
- Tester le chatbot pour qu'il reste actuel et fonctionnel,

- raffiner la compréhension des messages utilisateur ainsi que la pertinence des réponses,
- ajouter de nouvelles fonctionnalités en fonction du temps restant pour le rendre plus complet.

Ces missions permettront de garantir la robustesse du chatbot et ainsi d'offrir une « conversation » entre chatbot et utilisateur naturelle/humaine, mais également enrichissante scientifiquement pour l'utilisateur.

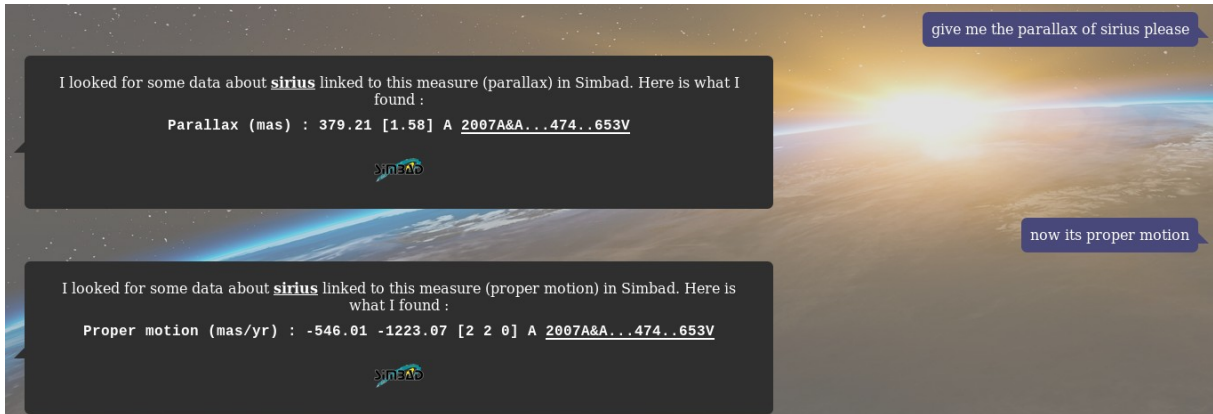


Illustration 7: Exemple d'une conversation (début)

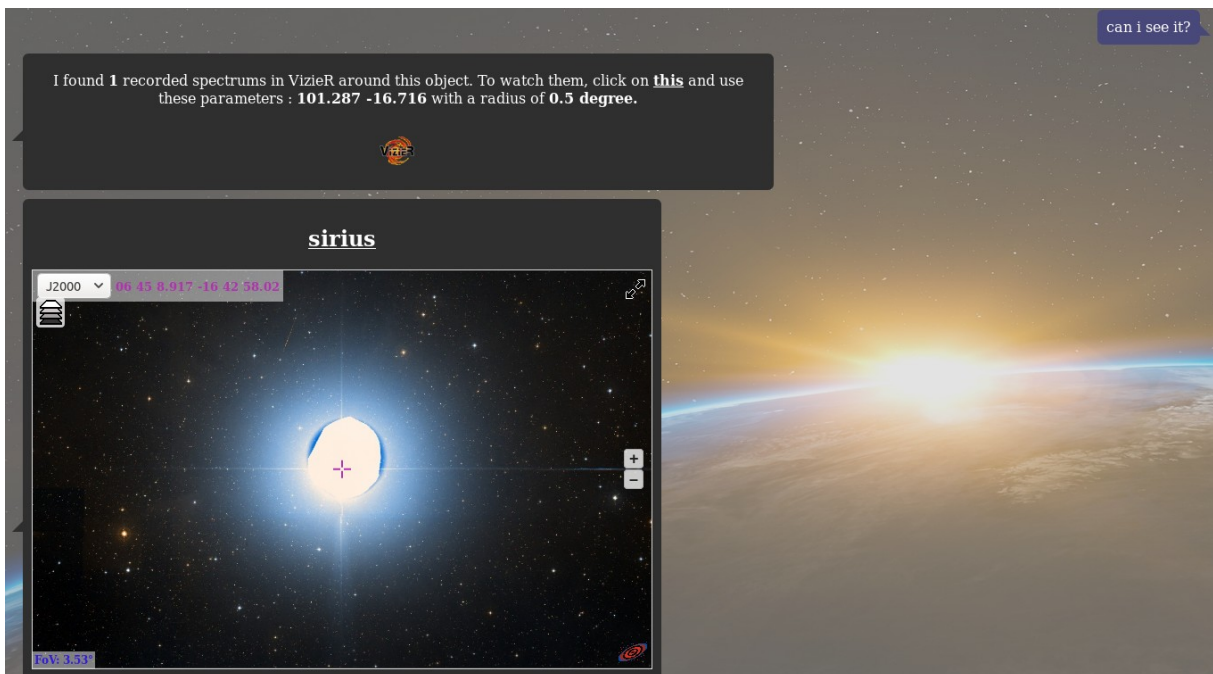


Illustration 8: Exemple d'une "conversation" (suite)

## 2.5 Problèmes et avantages

Utiliser Dialogflow est un réel avantage, car cela permet de gagner du temps de développement car même s'il faut configurer l'outil avec vigilance et précision, la qualité du service gratuit est tout de même appréciée et sûrement plus robuste que si on le développait soit même. Cependant, cela pose aussi un problème : comme énoncé précédemment, une de mes missions est de migrer de la version 1 à 2 de Dialogflow. L'observatoire est donc totalement dépendant de Google et ne peut qu'accepter les changements de l'entreprise américaine s'il souhaite continuer à utiliser leurs services.

D'autre part, continuer un développement en cours est une contrainte. Effectivement, il faut comprendre l'intégralité du fonctionnement de l'application développée jusqu'à maintenant pour pouvoir la continuer, la modifier et l'améliorer.

# 3 Cahier des charges

## 3.1 Descriptif fonctionnel

Le projet du chatbot durera plusieurs années et est actuellement à son début. La priorité pour l'observatoire est donc d'avoir une base solide pour ajouter le plus de fonctionnalités étape par étape.

Le chatbot possède à l'issue du stage les fonctionnalités principales suivantes :

- une capacité à répondre à une vingtaine de questions astronomiques types (voir exemple en annexe 2), le tout avec rapidité et souplesse. Ces questions constituent un corpus que l'on enrichit au fur et à mesure et permet également de vérifier qu'il n'y a pas de régressions lors des évolutions.
- un mode « *debug* » destiné à l'équipe de développement qui continuera le projet du chatbot. Ce mode permet de tracer chronologiquement les différents appels avec les données associées aux services du CDS pour comprendre le chemin de données en cas de bug. Le traçage est envoyé et conservé sur un serveur pour pouvoir y accéder facilement. Un mode « *comment* » permet aux utilisateurs de donner un avis sur une réponse fournie, une critique. Cela sera lu par l'équipe de développement et permettra d'améliorer certains points.
- une gestion d'erreurs précise permettant de guider l'utilisateur afin qu'il reformule correctement sa demande

## 3.2 Contraintes techniques

Les contraintes techniques du stage sont les suivantes :

1. comprendre le code existant (Javascript, HTML5, CSS) pour pouvoir le continuer et l'améliorer, se familiariser avec le fonctionnement de Dialogflow (son paramétrage) ;
2. rapidement acquérir les notions d'astronomies nécessaires pour comprendre les divers services du CDS (Table Access Protocol, ADQL...) et leur utilisations dans le code pour pouvoir les réutiliser;
3. changer l'architecture existante pour pouvoir migrer la version de Dialogflow. L'application web uniquement cliente est maintenant cliente/serveur et a subi une modification de code conséquente. S'en sont suivis de nombreux tests pour tester le



bon fonctionnement de la nouvelle version. Le serveur de Dialogflow, comme celui des logs et commentaires ont été développés en Node.js.

# 4 Mise en œuvre

## 4.1 Missions

Tout d'abord, ma première mission a consisté à migrer la version de l'outil Dialogflow de la version 1 à 2, car Google a annoncé que la version 1 sera dépréciée dans quelques mois et par conséquent cessera d'être fonctionnelle. Cette migration a été longue et complexe pour les raisons suivantes :

- Dialogflow est un outil récent donc encore assez peu documenté (même la documentation officielle de Google est légère). N'étant pas détaillé, cela a nécessité beaucoup de recherches (tutoriels, documentation, forums) et de l'initiative pour développer sa propre solution.
- La migration nécessite donc plusieurs compétences : capacité de recherches autonomes, administration système (pour configurer l'installation de l'outil), changement d'architecture majeur dans le code.
- Le changement d'architecture est le suivant : il faut désormais faire les appels aux services de Dialogflow par un serveur Node.js et plus par une API que l'on pouvait appeler côté client. Il faut donc également gérer les appels entre client et le serveur. Le JSON retourné par Dialogflow est à forme variable est sa structure est totalement différente que la première version. Il faut donc refaire l'intégralité de la déconstruction (ou *parsing*) pour pouvoir récupérer les données voulues pour interroger les services du CDS.
- Suite au changement d'architecture, il faut s'assurer que le chatbot ait le même comportement qu'avant. Il a fallu faire de nombreux tests pour assurer son bon fonctionnement. Les tests ont permis de constater le bon fonctionnement du chatbot après la migration.

Ma seconde mission a été d'ajouter des fonctionnalités et d'assurer la robustesse de l'outil. Concernant les fonctionnalités ajoutées, j'ai modifié la base d'entraînement existante de Dialogflow pour que l'utilisateur puisse utiliser le chatbot comme s'il avait un humain en face. Cela implique de comprendre des demandes plus implicites ou concises, comme « *informations arcturus* », à l'inverse d'un équivalent « *tell me the parallax, redshift, proper motion and spectral type of arcturus* » bien plus claire et explicite, mais moins agréable pour

l'utilisateur. Il est important de choisir les phrases d'entraînement avec précaution, car sinon la détection d'intention est biaisée et les informations ne seront pas utilisées dans les services adéquats.

Une autre fonctionnalité ajoutée est la possibilité de changer de relevé d'image pour un objet. Cela nécessite quatre requêtes qui se succèdent : chercher si l'objet existe, chercher les coordonnées de l'objet, une première recherche pour mettre un filtre par défaut, une seconde qui va chercher les autres filtres pertinents en fonction de critères précis pour les trier et les afficher sous l'image. Ce type de fonctionnalité est long à mettre en place car il requiert une bonne gestion de plusieurs tâches asynchrones et appels à différentes API. En revanche, c'est agréable pour l'utilisateur qui se voit la possibilité de choisir différents relevés, et donc peut les comparer pour mener sa recherche scientifique à bien.

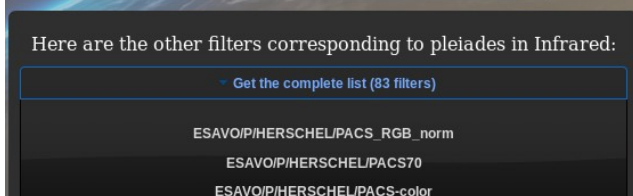


Illustration 9: Relevé d'image en infrarouge, avec la possibilité de le changer en dessous

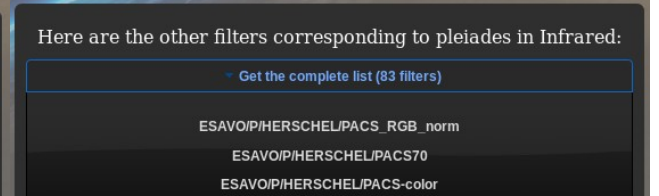
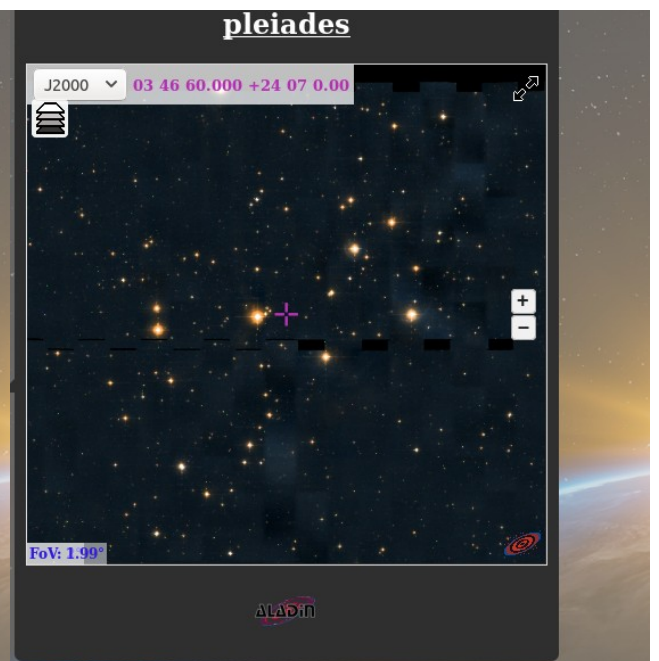


Illustration 10: Le troisième relevé à été sélectionné

## 4.2 Organisation

La totalité du stage s'est déroulé sur une méthode de travail *SCRUM*. Mon maître de stage André SCHAAFF ainsi que mon co-encadrant Sébastien DERRIERE m'ont donnés les tâches globales que j'avais à faire lors de réunions avec un certain ordre de priorité. Ensuite, j'étais libre de m'organiser comme je le souhaitais et de choisir mes méthodes de développement. Cela m'a permis un travail très autonome où j'ai également pu inclure mes initiatives pour ajouter une plus-value à la fonctionnalité attendue (exemple : gestion d'erreurs, assouplissement de l'outil Dialogflow pour une meilleure compréhension). En cas de blocage persistant ou d'incompréhension liée à l'astronomie, j'ai demandé leur aide et ils me l'ont donnée avec sympathie.

D'un point de vu critique, je pense avoir fourni un développement de bonne qualité. J'ai fourni un travail complexe pour la plupart des aspects en respectant les bonnes normes de programmation (nomenclatures, codestyle, commentaires...). Certaines tâches m'ont posé problème, comme la configuration de la version 2 de l'outil Dialogflow, mais j'ai réussi à m'en sortir par moi-même et en cas de blocage persistant, mes tuteurs m'ont aidé. J'ai trouvé les missions du stage très intéressantes, car elles m'ont permis de consolider et agrandir mes connaissances en Javascript, apprendre le Node.js. Par ailleurs, j'ai maintenant des notions d'astronomies et un exemple d'utilisation de traitement de langage naturel grâce à Dialogflow. Enfin, travailler dans un organisme regroupant développement et recherche m'a permis d'acquérir une véritable autonomie et rigueur.

# 5 Bilan

## 5.1 Bilan pour le CDS

Grâce à mon stage, le chatbot est désormais plus fonctionnel et robuste. Sa maintenance, ses tests, débogages et fonctionnalités ajoutées ont permis de l'améliorer. Mon équipe encadrante a été satisfaite de mon travail car j'ai remplis mes missions.

Cependant, le projet pour l'observatoire reste à ses débuts et ne sera mis en service que dans quelques années, le temps qu'il faudra pour assurer sans failles « l'intelligence » du chatbot, en d'autres termes sa capacité à comprendre tout type de requêtes, notamment nuancées ou imbriquées, permettant par exemple de répondre que la demande utilisateur n'a pas de sens même si elle est syntaxiquement correcte<sup>2</sup>. Le chatbot n'en est pas encore à ce stade.

## 5.2 Bilan humain

Mon stage à l'observatoire a été ma première expérience professionnelle dans le domaine de l'informatique. Je garde un très bon souvenir de cette première expérience car elle s'est très bien déroulée. J'ai pu travailler au sein d'une UMR, avec une atmosphère de travail détendue et cordiale. Les rapports (verbaux, par écrit...) avec mon maître de stage et mes co-encadrants, ainsi qu'avec le personnel de l'observatoire ont été professionnels et ont donc permis un travail efficace de ma part.

## 5.3 Bilan pédagogique

Durant 10 semaines entièrement consacrées au développement, je dresse un bilan pédagogique solide. Effectivement, les acquis de l'IUT m'ont permis d'avoir la base de connaissances nécessaires pour pouvoir comprendre le code existant. Mais ce stage m'a surtout permis d'approfondir mes qualités d'informaticien, notamment la polyvalence. Le chatbot est certes une application web mais a regroupé plusieurs domaines : le web côté client essentiellement, mais également de la programmation de serveurs, du système pour l'installation de l'outil Dialogflow, et évidemment une part importante de recherches, lectures de documentations...

---

2 Par exemple, demander « what is the redshift of sirius » n'a pas de sens scientifiquement parlant, car une étoile n'a pas cette mesure. Voir l'illustration 12 pour plus de détails.

La principale différence avec l'IUT est donc la quasi totale autonomie de mon stage. L'organisation était totalement libre durant ce stage, ainsi que la méthode de développement. Par ailleurs, la complexité du travail était naturellement plus grande qu'à l'IUT (où les projets sont plus simples car à visée éducative), car il fallait développer un outil d'une part destiné à une communauté scientifique et d'autre part faisant recours aux nombreux services du CDS (Simbad, Vizier, Aladin, MocServer...).

# Conclusion

Pour conclure, mon stage à l'observatoire astronomique de Strasbourg a été une expérience professionnelle personnellement enrichissante. J'ai pu découvrir comment une UMR fonctionne dans le domaine de l'astronomie et contribuer activement au développement de l'un de leurs projets.

J'ai approfondi et élargi mes connaissances en informatique, particulièrement en technologies web, ce qui me sera utile plus tard. En développant ce chatbot, j'ai pu découvrir les enjeux de l'UMR à Strasbourg et y apporter ma contribution qui permettra à la fin de sa programmation de répondre aux besoins de la communauté astronomique scientifique internationale. Une évolution du chatbot serait par exemple la reconnaissance vocale, qui serait très agréable pour les utilisateurs.

Développer un chatbot m'a beaucoup intéressé car c'est un outil déjà popularisé dans de nombreux domaines tels que les services après ventes, mais encore très rarement dans l'astronomie. Ainsi, j'ai pu participer à un projet moderne et original, car il se distingue nettement des projets typiques (services de boutique en ligne par exemple) présents partout.

Ce stage a été un défi technique car beaucoup de notions étaient nouvelles à mon arrivée : les notions d'astronomie, le traitement de langage naturel, les services du CDS spécifiques à l'astronomie... Grâce mon travail autonome, j'ai continué ma formation d'informaticien en développant et en me documentant beaucoup, ce qui m'a permis de réussir mon stage. De plus, cette expérience m'a fait prendre goût au développement web et à la notion de traitement de langage naturel.

Cette première expérience professionnelle m'a confronté aux notions d'apprentissage intelligent (par conséquent *machine & deep learning* également), puisqu'il fallait entraîner la reconnaissance d'informations dans le langage. Cela m'a fortement intéressé à ce domaine et je pense m'y orienter dans la suite de mes études et ma carrière professionnelle.

## Table des illustrations

Illustration 1: Organigramme de l'observatoire de Strasbourg.....	10
Illustration 2: Les trois principaux services du CDS : Simbad, VizieR et Aladin.....	11
Illustration 3: Exemple et extrait de résultats obtenus avec Simbad lorsqu'on entre la galaxie « m31 ».....	12
Illustration 4: Image de la galaxie "m31", affichée avec Aladin.....	12
Illustration 5: Exemple d'utilisation du chatbot.....	15
Illustration 6: Agent dialogflow créé en fonction de nos propres besoins.....	16
Illustration 7: Exemple d'une conversation (début).....	18
Illustration 8: Exemple d'une "conversation" (suite).....	18
Illustration 9: Relevé d'image en infrarouge, avec la possibilité de le changer en dessous. .	22
Illustration 10: Le troisième relevé à été sélectionné.....	22
Illustration 11: Exemple de formulaire pour VizieR long à remplir.....	1
Illustration 12: Les objets astronomiques n'ont pas tous les mêmes mesures selon leur type	11



# Sitographie

## **Dialogflow :**

Documentation de Dialogflow

<https://dialogflow.com/docs/reference>

## **CDS :**

Site web du Centre de Données astronomiques de Strasbourg

<http://cds.u-strasbg.fr/>

Simbad, base de données principale du CDS

<http://simbad.u-strasbg.fr/simbad/>

Documentation d'Aladin, atlas du ciel du CDS

<https://aladin.u-strasbg.fr/AladinLite/doc/API/>

MocServer du CDS, permettant différents types de requêtes

<http://alasky.unistra.fr/MocServer/query>

Page Wikipédia du CDS

[https://fr.wikipedia.org/wiki/Centre\\_de\\_donn%C3%A9es\\_astronomiques\\_de\\_Strasbourg](https://fr.wikipedia.org/wiki/Centre_de_donn%C3%A9es_astronomiques_de_Strasbourg)

## **Observatoire astronomique de Strasbourg :**

Page Wikipédia de l'observatoire astronomique de Strasbourg

[https://fr.wikipedia.org/wiki/Observatoire\\_astronomique\\_de\\_Strasbourg](https://fr.wikipedia.org/wiki/Observatoire_astronomique_de_Strasbourg)

## **Autres :**

Stackoverflow, forum centré sur les questions réponses en programmation

<https://stackoverflow.com>

# SOMMAIRE DES ANNEXES

Annexe I Notions astronomiques.....	I
Annexe II Questions types.....	III

# Annexe I

## Notions astronomiques

### I Formulaire VizieR



**Find catalogs among 18559 available**

Clear  Find...

Expand search

**?** *Catalog*, author's name, word(s) from title, description, etc.  
e.g.: AGN, Veron, I/239, or bibcodes...

► **Search for catalogs by column descriptions (UCD)** **?**

► **Search for catalogs containing additional data**

**Search by Position across 20092 tables**

Target Name (resolved by [Sesame](#)) or Position:  J2000  2 Target dimension:  2 arcmin  Go!

Radius  Box size

[More about VizieR](#)  

Wavelength	Mission	Astronomy
Radio	AKARI	Abundances
Millimeter	ANS	Ages
IR	ASCA	AGN
optical	BeppoSAX	Associations
UV	CGRO	Asteroseismology
EUV	Chandra	Atomic_Data
X-ray	COBE	Binaries:cataclysmic

Illustration 11: Exemple de formulaire pour VizieR long à remplir

De plus, si l'on cherche une mesure, il faudra effectuer un X-match pour croiser les données entre plusieurs catalogues et enfin obtenir le résultat désiré. C'est donc assez fastidieux.

## II Mesures associées au type d'objet

Quelles mesures ont un sens pour différents types d'objets :

Mesure	Etoiles	Etoiles doubles	Nébuleuses planétaires	Pulsar	Galaxies	Quasar
Position	x	x	x	x	x	x
Type (ambigü !), object type, or nature	x	x	x	x	x	x
Spectral type	x	x				
Morphological type					x	
Redshift					x	x
Radial velocity	x	x	x	x	(1)	(1)
Size ambigü, choisir un des deux suivants						
Angular diameter			x		x	
Diameter, Radius	x		x		x	
Parallax	x	x	x	x		
Distance	x	x	x	x	(2)	(2)
Proper motion	x	x		x		
(effective) Temperature	x					
Period	(3)	(3)				
Magnitude (4)	x	x	x	x	x	x
Flux (4)	?	?	?	?	x	x

(1) On peut traduire le redshift en radial velocity... (2) Pour les étoiles la distance peut être calculée comme l'inverse de la parallaxe. Pour les galaxies et QSO, cela peut se mesurer autrement, c'est moins fréquent. (3) Pour les étoiles ou étoiles doubles variables et périodiques (4) Le plus souvent, on va demander une magnitude spécifique: magnitude B, magnitude V, magnitude G Gaia, infrared magnitude... idem pour un flux, on va chercher le flux radio, ou en rayons X...

Illustration 12: Les objets astronomiques n'ont pas tous les mêmes mesures selon leur type

Ce tableau créé par Sébastien DERRIERE m'a permis de vérifier sémantiquement si ce que je produisais était correct.

# Annexe II

## Questions types

Voici un exemple de questions types que le chatbot peut traiter avec succès.

"Hello !",

"How are you ?",

"What is the parallax, the spectral type, the position, the proper motion, the magnitudes and the distance of Sirius ?",

"Galactic coordinates and spectral type of Arcturus ?",

"What about its parallax ?",

"Could you check in Vizier ?",

"Tell me everything about T Tau",

"Could you search the proper motion of pulsars", "Try again",

"Galaxies members of the Pleiades",

"Which cluster is Electra's parent ?",

"Siblings of T Tau",

"Show me Arcturus",

"I want to see the Pleiades in IR", "Now in optical",

"Tell me everything about M 31", "I want to see it please",

"Find some galaxies 10' around the Virgo Cluster",

"Query the latest Veron catalogue",

"Catalogues measuring redshift and parallax",

"Most popular catalogue talking about M31",

"Catalogues talking about QSOs",

"Find me the last Veron catalogue measuring redshift and parallax and speaking about QSOs",

"List some galaxies at redshift  $\leq 0$ ", "Find some objects looking like T Tau",

"Thank you very much !", "Bye bye"