



Université de Nice - Sophia Antipolis

IUT de Nice - Côte d'Azur

Département Informatique

41 bd. Napoléon III - 06200 Nice



**RAPPORT DE STAGE POUR L'OBTENTION DU
DIPLOME UNIVERSITAIRE DE TECHNOLOGIE**

Session 2013-2014

Réalisation d'un service WEB pour la mise à disposition conviviale des données de
l'instrumentation PicardSol

Présenté par Adrien Boucher

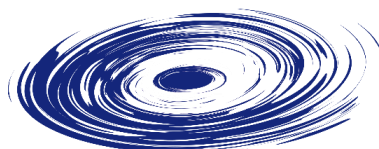
Observatoire de la Côte d'Azur

Boulevard de l'Observatoire

06304 NICE



Observatoire
de la CÔTE d'AZUR



LAGRANGE

Tuteur entreprise

Thierry Corbard

Enseignant responsable

Michel Syska

Remerciements

Je tiens à remercier M. Thierry Corbard mon tuteur en entreprise qui m'a aidé à améliorer mon site web et m'a guidé dans sa conception, Mme Catherine Renaud qui m'a suivi durant tout mon stage, son aide a été précieuse et ses conseils très utiles, et enfin M. Frédéric Morand qui m'a fait visiter le site d'observation de Calern et m'a expliqué ce qui y est fait.

Tous trois font partie de l'équipe dans laquelle j'ai travaillé et c'est grâce à eux que mon stage a été très agréable et enrichissant.

Je souhaite également remercier tout le personnel de l'OCA avec qui j'ai été en contact pour l'accueil chaleureux et l'ambiance sympathique dans laquelle mon stage s'est déroulé.

Enfin, je remercie M. Michel Syska pour ses conseils, et pour l'intérêt qu'il a porté au bon déroulement de mon stage.

Résumé

Pour obtenir un Diplôme Universitaire de Technologie en informatique, un stage de fin de parcours, d'une durée de dix semaines est obligatoire. En ce qui me concerne, je l'ai réalisé du 22 avril au 27 juin, au sein de l'équipe PicardSol de l'Observatoire de la Côte d'Azur.

Dans le cadre de ses recherches l'équipe effectue des observations du soleil tout au long de l'année et en recueille de nombreuses données qui sont stockées sur les serveurs de l'Observatoire.

La mission consistait en la création d'une interface web couplée à une base de données pour faciliter l'accès aux données des images du soleil pour les scientifiques en les classant par critères et en offrant la possibilité de télécharger une sélection d'images.

Cette interface web doit être couplée avec un système de mise à jour automatique des données afin de prendre en compte les dernières observations faites par les membres de l'équipe.

A la fin du stage, le site web et l'outil de mise à jour des données ont été développés et testés.

Abstract

To obtain the University Diploma in Computer Science, an internship of ten weeks duration at least is mandatory. In my case it was carried out from 22 April 2014 to 27 June 2014 in the PicardSol team of the Côte d'Azur Observatory.

As part of its research, the team carries on measurements of the sun and the resulting data are stored on the observatory server.

The mission consisted in creating a web interface coupled with a database to simplify the access to the solar data for the scientist by sorting them using criterias and allowing the user to download a selection of pictures.

This interface comes with an update system to always have the last measurement data.

At the end, the website and the update tool have been both developed and tested successfully.

Table des matières

INTRODUCTION	6
1. L'ENTREPRISE.....	7
1.1. L'OBSERVATOIRE DE LA COTE D'AZUR	7
1.2. LE LABORATOIRE LAGRANGE	8
1.3. LE PROJET PICARDSOL	9
2. PRESENTATION DU PROJET.....	12
2.1 SUJET DU STAGE DEFINI PAR L'ENTREPRISE	12
2.2 CAHIER DES CHARGES	13
2.3 PLANNING PREVISIONNEL.....	14
3. ANALYSE ET CONCEPTION.....	14
3.1. CHOIX DE LA SOLUTION.....	14
3.2. CONCEPTION DE LA SOLUTION A REALISER	15
4. DEVELOPPEMENT	18
4.1 BASE DE DONNEES.....	18
4.2 SCRIPTS.....	19
4.3 INTERFACE WEB	22
4.4. DOCUMENTATION	26
5. TESTS	28
6. RESULTATS.....	29
6.1 ÉTAT FINAL DU PRODUIT PAR RAPPORT AU CAHIER DES CHARGES.....	29
6.2 DIFFICULTES RENCONTREES	29
6.3 QUANTIFICATIONS.....	30
6.4 PLANNING EFFECTIF	30
CONCLUSION	31
GLOSSAIRE.....	33
BIBLIOGRAPHIE.....	34
ANNEXES	35

Introduction

Ce rapport porte sur le stage que j'ai effectué à l'Observatoire de la Côte d'Azur du 22 avril au 27 juin, au sein de l'équipe PicardSol du laboratoire Lagrange. L'équipe travaille sur l'observation du soleil depuis la Terre ainsi qu'à l'interprétation des données recueillies.

Depuis 2011 l'instrumentation PicardSol observe le soleil depuis le site de Calern durant toute l'année afin d'obtenir des mesures du diamètre solaire et d'étudier les évolutions de ce dernier. Par conséquent un nombre important de fichiers sont générés chaque jour d'observation et sont stockés sur les serveurs de l'Observatoire par date et par longueur d'onde à laquelle le soleil a été observé.

Le but de ce stage a été de réaliser une interface Web afin d'accéder aux données de façon conviviale pour les scientifiques. Un système a été également mis en place afin d'ajouter les données des dernières observations de façon régulière.

La première partie du rapport est consacrée à l'Observatoire, le laboratoire Lagrange et le projet PicardSol ainsi que l'objet du stage et son contexte. La deuxième partie présente l'analyse du projet. Et pour finir sont détaillés le développement du projet ainsi que les tests réalisés.

1. L'entreprise

1.1. L'Observatoire de la Côte d'Azur

L'Observatoire de la Côte d'Azur (OCA) est un établissement public de recherche autonome lié à l'Université de Nice-Sophia Antipolis (UNS). Il compte plus de 450 agents qui sont répartis sur 4 sites :

- Mont Gros, site historique.
- Campus de Valrose de l'UNS à Nice.
- Sophia Antipolis
- Le plateau de Calern qui est un site d'observation.

L'OCA constitue le pôle de recherche azuréen en matière d'astrophysique et de géosciences. Il fait partie des 25 Observatoires des Sciences de l'Univers français qui ont pour mission de recueillir en continu et systématiquement des données observationnelles de la Terre et de l'Univers.

Il permet ainsi d'explorer et comprendre les processus physiques qui se déroulent du cœur de notre planète aux confins de l'Univers et de valoriser les sciences de la Terre et de l'Univers.

Ses missions sont d'explorer, observer, former et diffuser des connaissances.

Il regroupe 3 unités de recherche sous la tutelle de l'UNS, du CNRS, de l'IRD et de l'OCA :

Artémis, Lagrange et Géoazur. Il héberge aussi l'unité LISE du Collège de France.

L'unité de services Galilée assure la maintenance des infrastructures et permet de mettre en œuvre la politique scientifique de l'établissement.

Pour produire et exploiter les données recueillies par les grands instruments au sol et dans l'espace, l'OCA met en œuvre une approche globale combinant théorie, modélisation, observation, analyse et constitution de banques de données. L'OCA soutient notamment les grands projets instrumentaux dans lesquels les laboratoires sont impliqués au travers de consortiums internationaux, et fournit aux équipes de recherche les moyens techniques mutualisés requis.

Les grands projets instrumentaux :

- RESIF – EPOS : Réseaux sismologiques et géodésiques
→ Structure de la Terre interne et risques associés.
- MeO : Plateforme de métrologie laser (Calern).
→ Géodésie spatiale et orbitographie.
- MATISSE : Recombinateur imageur de 4 télescopes du VLT (ESO, Chili).
→ Environnements stellaires, origine et évolution des systèmes planétaires.
- Gaia : Mission spatiale d'astrométrie (ESA).
→ Catalogue d'un milliards d'objets astrophysiques.
- VIRGO : Antenne gravitationnelle (Italie).
→ Détection des ondes gravitationnelles.

1.2. Le laboratoire Lagrange

Le laboratoire J.-L. Lagrange a été créé le 1er janvier 2012 par la fusion des laboratoires Cassiopée et Fizeau, il a maintenant un effectif de 170 personnes.

C'est une Unité Mixte de Recherche de l'Observatoire de la Côte d'Azur, du CNRS et de l'UNS.

C'est un laboratoire pluridisciplinaire qui regroupe différentes équipes :

- En astrophysique
 - Planétologie
 - Physique stellaire et solaire (projet PicardSol)
 - Galaxies et cosmologie
- Mécanique des fluides
- Traitement du signal et images, composée essentiellement de maîtres de conférences.

Les équipes sont composées d'astronomes, de maîtres de conférences et de chercheurs.

Les ingénieurs sont regroupés au sein du Groupe Ingénierie et Recherche (GIR) qui compte 32 ingénieurs et techniciens.

Afin de développer de nouvelles théories ainsi que des modèles et de les confronter à des observations acquises depuis le sol et l'espace, des compétences en instrumentation pour l'observation astronomique à haute résolution spatiale et en calcul à haute performance sont particulièrement mises en valeur au sein des équipes.

Le laboratoire est à la tête d'un consortium européen développant MATISSE, un instrument de seconde génération pour l'interféromètre VLTI de l'ESO.

Une station de contrôle à distance de l'interféromètre CHARA/VEGA situé au Mont Wilson en Californie est disponible à l'Observatoire de Nice. Le laboratoire est également impliqué dans plusieurs missions spatiales, particulièrement dans la mission Gaia de l'ESA avec un rôle majeur dans le traitement et l'analyse des données. Un fort programme de R&D forme la base des futurs programmes instrumentaux. Le laboratoire met un accent important sur la formation, pilotant le nouveau Master IMAG2E et accueillant une trentaine de doctorants, sur les collaborations internationales avec de nombreux visiteurs présents à long terme au laboratoire, et organisant plusieurs conférences internationales chaque année.

1.3. Le projet PicardSol

PicardSol est un ensemble d'instruments sur le site de Calern. Cet ensemble a fonctionné en parallèle avec la mission Picard conduite dans l'espace jusqu'en avril 2014. L'objectif est de déterminer la forme du soleil simultanément depuis l'espace et le sol afin d'apprendre à interpréter les mesures faites au sol grâce à une analyse détaillée des effets atmosphériques.

L'équipe est composée de chercheurs, d'ingénieurs et d'un doctorant qui collaborent afin de mettre en œuvre des outils logiciels pour la gestion des instruments d'observation et des données collectées de façon à répondre au mieux aux besoins des scientifiques.

Le projet se fait en collaboration avec le CRAAG à Alger.

Motivations

Dans le but de mieux comprendre le fonctionnement des étoiles, notre soleil étant l'étoile la plus proche et donc la plus facilement observable est étudié avec attention.

Mais les mesures faites depuis le sol du soleil à différents endroits montrent des résultats contradictoires. Ceci est dû à l'influence de l'atmosphère qui perturbe les observations.

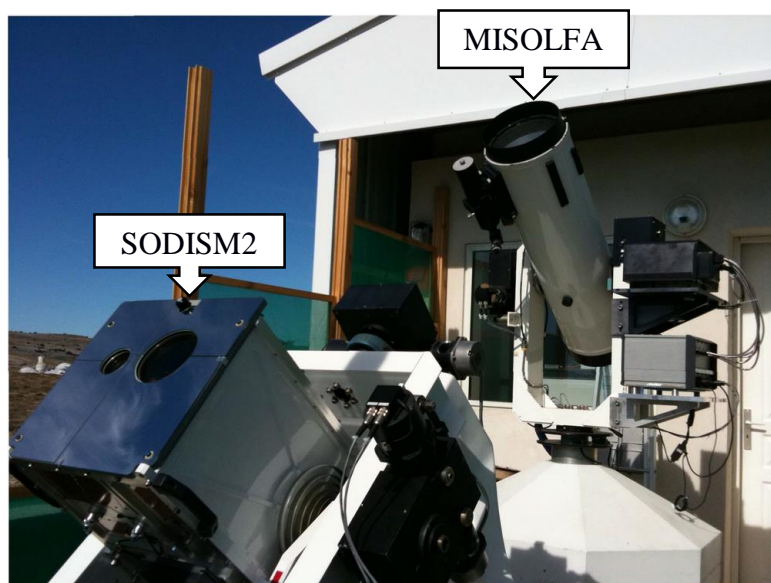
Pour pallier ce problème des mesures ont été faites en orbite, notamment grâce au satellite Picard.

Cependant faire des observations depuis l'espace est coûteux et il y a de nombreuses contraintes. De plus la plus grande série d'observations disponible pour l'instant a été réalisée depuis le sol. Il est donc essentiel de pouvoir comprendre et interpréter les mesures faites depuis le sol, d'où la mise en place d'un important programme de mesures depuis l'espace associées à celles du projet PicardSol. Ce dernier continue encore à ce jour après la fin de la mission spatiale. Une référence du soleil dans l'espace peut être tout de même obtenue via des satellites d'observation actuellement en fonctionnement pour comparer avec les données obtenues au sol.

Instruments

Durant la mission spatiale Picard (2010-2013) les observations au sol ont été faites par les deux instruments suivants, installés sur le site d'observation de Calern :

- SODISM2 : réplique de l'instrument spatial de mesure du diamètre solaire. Il acquiert des images pleines du soleil dans 5 longueurs d'ondes différentes.
- MISOLFA : télescope Cassegrain Coudé de 26cm de diamètre servant à caractériser la turbulence atmosphérique.



Les objectifs du projet

- En physique Solaire : étude des variations du soleil à long terme.
- En physique atmosphérique : permettre une meilleure connaissance des propriétés optiques de l'atmosphère et test de la validité des modèles atmosphériques existants grâce à la

comparaison sol-espace.

- En climatologie : étude de l'influence éventuelle de la variabilité solaire sur le climat.
- SODISM2 sert aussi de banc de tests pour démontrer la faisabilité (ou non) des mesures de rayon au sol pour le long terme, sans observations spatiales.
- Déterminer s'il est possible de faire un lien avec les observations passées.

Principes d'analyse

SODISM2 capture des images complètes du soleil à 5 longueurs d'onde différentes (393nm, 2 à 535.7 nm avec un temps d'exposition différent, 607nm, 782nm, 1025nm) au rythme d'une image par minute. Certaines de ces images ont été prises quasiment au même instant que celles de SODISM prises dans l'espace au rythme de 2 images par orbite.

Au même moment, MISOLFA prend en continu des images du limbe en haute résolution à un rythme de 32 images par seconde sur deux bords opposés du disque solaire. Il enregistre également les fluctuations temporelles sur la voie pupille.

Les mesures faites par MISOLFA servent à obtenir les paramètres de la turbulence atmosphérique. Ces paramètres devraient être ajoutés aux en-têtes des images prises par SODISM2.

L'équipe du projet sur le site de Calern



2. Présentation du projet

2.1. Sujet du stage défini par l'entreprise

L'objectif de ce stage est de faciliter dans un premier temps l'accès pour les scientifiques aux données de l'instrument SODISM2 via un serveur WEB – PHP MySQL. Il doit être possible de rajouter d'autres instruments.

Le stage se décompose en trois phases.

Dans un premier temps, définir la base de données MySQL sur le serveur WEB **solar-data** en utilisant les mots clefs définis dans les en-têtes des fichiers fits.

Elle doit répertorier les en-têtes des fichiers fits qui sont rangés par longueur d'onde dans laquelle l'image a été prise et par jour (nommé AAAAMMJJ pour année+mois+jour).

Les données sont archivées sur le serveur PICARD et sont accessibles uniquement via un ftp anonyme depuis le serveur solar-data. Prévoir la mise à jour automatique de la base de données MySQL compte tenu des nouvelles images générées au fur et à mesure des observations.

Construire une interface WEB en PHP pour l'interrogation de la base de données. Cette interface devant être dans un premier temps protégée par un mot de passe.

Elle doit permettre l'affichage de la liste des fichiers depuis la base de données selon des critères donnés.

L'affichage doit comprendre des vignettes permettant de prévisualiser les images pour ainsi sélectionner les données souhaitées facilement.

Il est également demandé de générer une archive des images sélectionnées pour téléchargement sachant que l'archive peut être volumineuse. La contrainte est que les fichiers fits ne doivent pas être dupliqués pour le téléchargement ni encombrer le réseau.

2.2. Cahier des charges

Un cahier des charges fonctionnel a été rédigé par mes soins afin de clarifier les besoins du projet. Ceci a été fait grâce aux discussions avec mon tuteur M. Corbard, qui est le premier utilisateur du site web qui a été développé, pour répondre au mieux aux attentes des utilisateurs de l'outil.

Il a été rédigé à partir des besoins suivants :

- Définir une base de données MySQL sur le serveur web solar-data d'après les entêtes des fichiers fits.
- Mettre à jour automatiquement la base de données avec les nouvelles images.
- Construire une interface Web en PHP pour l'interrogation de la base de données selon des critères
- Générer une archive des images sélectionnées à télécharger.
- Permettre l'implémentation d'autres instruments.

En plus de ces besoins le lien d'un site exemple avait été fourni pour qu'il puisse éventuellement servir de référence : <http://sidc.oma.be/uset/searchForm.php>

Il en est ressorti l'organisation suivante pour le déroulement du stage :

Tâche	Durée prévisionnelle (heures)
Compréhension du sujet	10
Recherche	35
Cahier des charges	15
Spécifications	30
Développement et documentation	100
Tests unitaires	45
Tests de validation	45
Rédaction du rapport de stage et préparation de la soutenance	70
Total	350

2.3. Planning prévisionnel

Tâche \ Semaine	Avril		Mai				Juin			
	17	18	19	20	21	22	23	24	25	26
Compréhension du sujet										
Recherche										
Cahier des charges										
Spécifications										
Développement et documentation										
Tests unitaires										
Tests de validation										
Rapport de stage / Soutenance										

Ce tableau, que j'ai réalisé après le cahier des charges avec mon responsable, représente la charge de travail prévisionnelle. Le développement commence dès la deuxième semaine avec la création de la base de données. J'ai réalisé des tests parallèlement au développement afin de m'assurer du bon fonctionnement du code produit.

3. Analyse et conception

3.1. Choix de la solution

Au début du stage, deux solutions m'ont été proposées pour la gestion de la base de données :

- Utiliser le logiciel FitHAS (*Fits Header Archiving System*) qui permet de stocker les en-têtes des fichiers fits dans une base de données MySQL.
- Développer mon propre moyen de gestion de la base de données.

J'ai en premier lieu étudié le fonctionnement du logiciel FitHAS développé par un laboratoire chinois pour l'Observatoire Virtuel chinois (<http://www.china-vo.org/fithas.html>). Mais celui-ci n'est pas documenté et les codes sources ne sont pas accessibles.

Le logiciel permet de créer une table dans une base de données avec les en-têtes d'un fichier fits ou

des fichiers fits contenus dans un dossier.

Malheureusement toutes les données sont stockées dans des champs de type 'text' dans la base de données, ce qui est gênant pour le traitement.

De plus la fonction ajout de fichiers dans la base de données ne fonctionne pas. Il est impossible de choisir la table dans laquelle ajouter les données, et l'interface graphique n'est pas tout à fait au point.

J'ai trouvé qu'il était tout de même possible d'ajouter les données à une table en utilisant la fonction de création d'une table. Pour cela il faut spécifier le nom de la table existante et ignorer le message d'erreur et la table est mise à jour.

Étant donné qu'en plus de ces défauts le logiciel ne permet pas l'ajout automatisé des en-têtes dans une base de données j'ai décidé avec l'accord de mon tuteur de l'abandonner.

J'ai donc opté pour la création de mon propre outil afin de pouvoir rendre la tâche automatique et de pouvoir y incorporer d'autres fonctionnalités.

Travaillant sous un environnement Linux exclusivement, j'ai donc opté pour la création d'un script en Bash qui est un langage rapide, efficace et qui offre de grandes possibilités.

3.2. Conception de la solution à réaliser

3.2.1. Base de données

La conception de la base de données a été faite sur la base des résultats donnés par la génération de table du logiciel FitHAS.

La base de données a été faite sous MySQL.

Chaque table sert à stocker les données d'un instrument différent, car les mots-clés varient en fonction de ce dernier.

Field	Type	Null	Key	Default	Extra
id	bigint(20)	NO	PRI	NULL	auto_increment
SIMPLE	text	YES		NULL	
BITPIX	bigint(20)	YES		NULL	
NAXIS	bigint(20)	YES		NULL	
NAXIS1	bigint(20)	YES		NULL	
NAXIS2	bigint(20)	YES		NULL	
EXTEND	text	YES		NULL	
FILENAME	text	YES		NULL	
REVISION	text	YES		NULL	
DATE	text	YES		NULL	
INSTITUT	text	YES		NULL	
ORIGIN	text	YES		NULL	
OBSERVER	text	YES		NULL	
CONTACT	text	YES		NULL	
RESCIENT	text	YES		NULL	
INSTRUME	text	YES		NULL	
CMP_NAME	text	YES		NULL	
OBJECT	text	YES		NULL	
BUNIT	text	YES		NULL	
PHYS PARA	text	YES		NULL	
DATE_OBS	text	YES		NULL	

3.2.2. Scripts

Afin de répondre aux besoins du projet, j'ai décidé de créer trois scripts.

Pour la gestion de la base de données, un premier script gère la mise à jour des données.

Il a pour mission :

- De chercher les nouveaux fichiers fits.
- D'extraire les données des en-têtes des fichiers trouvés.
- D'ajouter les données dans la base de données.

En parallèle, un second script doit générer les images tirées des fichiers fits au format jpeg afin de les afficher sur le site web.

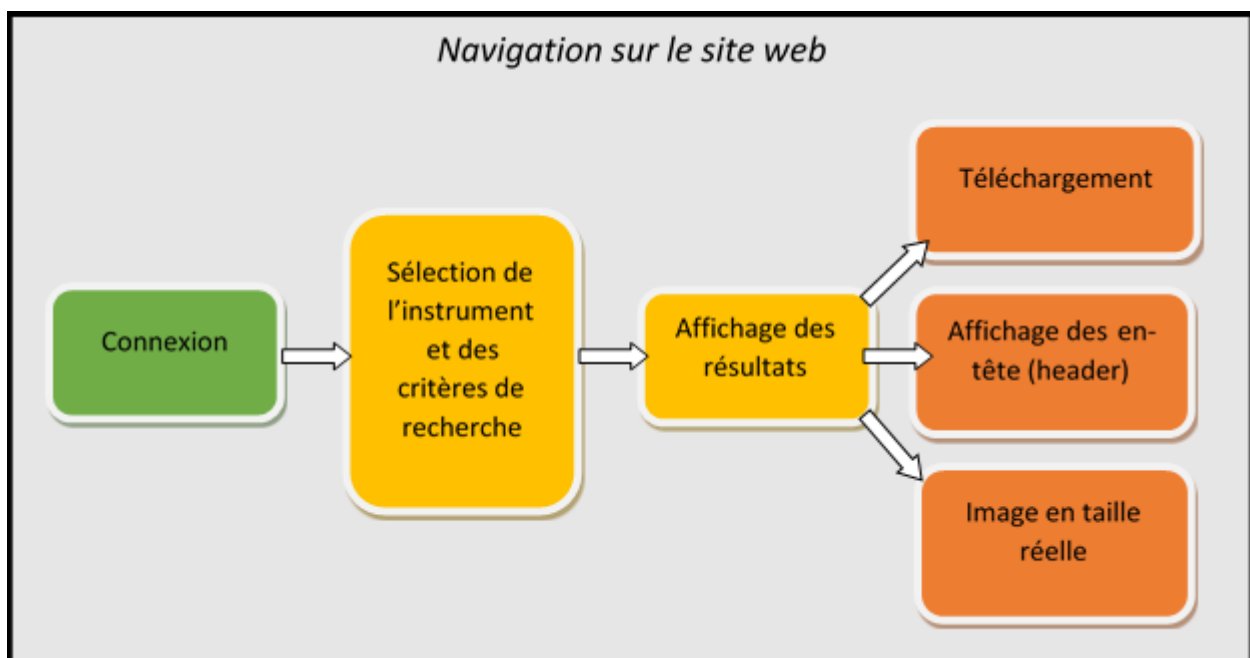
Il y a également un troisième script pour créer une table associée à un nouvel instrument dans la base de données à partir de l'en-tête d'un fichier fits.

3.2.3. Site web

Le site web a été organisé de façon à faciliter l'utilisation en réduisant le nombre de pages auxquelles il faut accéder pour avoir les données. Le but est d'optimiser le confort.

Il est composé de 3 parties :

- Page d'accueil / login pour le mot de passe de sécurité et le mail de l'utilisateur. Il a été ensuite décidé d'utiliser Htaccess pour protéger le site si la sécurité doit être renforcée. Cela permet de choisir qui peut y avoir accès. Le but est de pouvoir accéder au site de n'importe où, mais avec une forte protection.
- Page de sélection des critères d'affichage des images.
 - Sélection instruments.
 - Critères de sélection adaptés à l'instrument sélectionné.
- Page d'affichage des images.
 - Téléchargement.
 - Affichage des en-têtes.
 - Affichage des images en taille réelle.



Afin d'éviter un temps de chargement de la page trop long, le nombre de résultats affichés a été limité.

Le nombre de fichiers à télécharger en une fois est également limité. Un nombre important de

fichiers volumineux à télécharger poserait des problèmes et pour le réseau et pour l'utilisateur car le téléchargement serait trop long étant donné que chaque fichier pèse 16Mo.

Ces valeurs sont définies dans le fichier de configuration du site.

4. Développement

Le développement a été découpé en 3 parties : la base de données, les scripts de gestion de la base de données et le site web qui permet l'affichage et la récupération des données.

4.1. Base de données

Dans un premier temps la BD a été créée à l'aide du logiciel FitHAS.

Chaque en-tête possède un ID unique et chaque mot-clé correspond à une colonne de la table. Pour chaque mot-clé le commentaire a été ajouté dans la structure de la table à la création. Les valeurs sont stockées sous forme de texte et le premier commentaire trouvé est ajouté dans la base de données en plus des autres valeurs

Elle a été par la suite refaite en utilisant le script que j'ai développé et qui sera décrit dans la prochaine partie. Le changement majeur est que les valeurs numériques sont maintenant de type INT pour les entiers et DOUBLE pour les réels et que toutes les lignes de commentaires présentes dans les en-têtes sont ignorées car elles ont toutes le même mot-clé : 'COMMENT'.

La base de données contient à ce jour les données de près de 102000 fichiers fits générés par SODISM2 et elle est régulièrement mise à jour de façon automatique grâce à un script Bash qui tourne une fois par jour, de nuit pour ne gêner personne.

4.2. Scripts

Afin de faciliter et permettre l'automatisation de la mise à jour de la base de données, j'ai créé trois scripts en langage Bash.

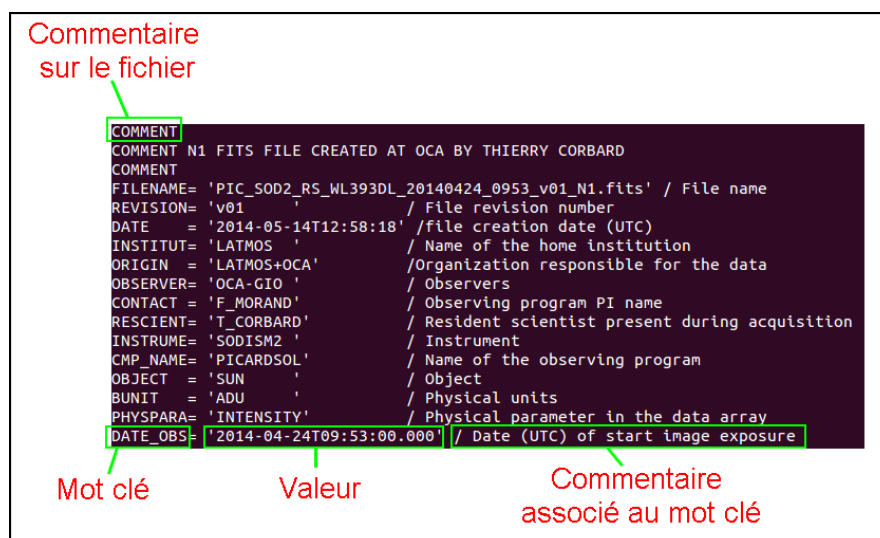
➤ **update_fits.sh**

C'est le premier script que j'ai réalisé et c'est de loin le plus important en termes de travail fourni pour le développer ainsi qu'en nombre de lignes de codes. C'est ce script qui permet la mise à jour automatique de la base de données afin d'y ajouter les dernières données des observations faites par les instruments après que celles-ci ont été prétraitées.

Je me suis tout d'abord occupé de la recherche des fichiers fits à traiter. La liste des fichiers trouvés est comparée à la liste des fichiers présents dans la base de données afin de ne garder que les nouveaux fichiers.

Vient ensuite le traitement des en-têtes des fichiers fits c'est à dire l'extraction des données utiles des en-têtes. La commande « funhead » du package funtool d'Ubuntu permet de récupérer l'en-tête à l'état brut (en ASCII) d'un fichier fits passé en paramètre, il a donc fallu un moyen de le traiter.

Une ligne d'un en-tête contenant une valeur est construite de la façon suivante :

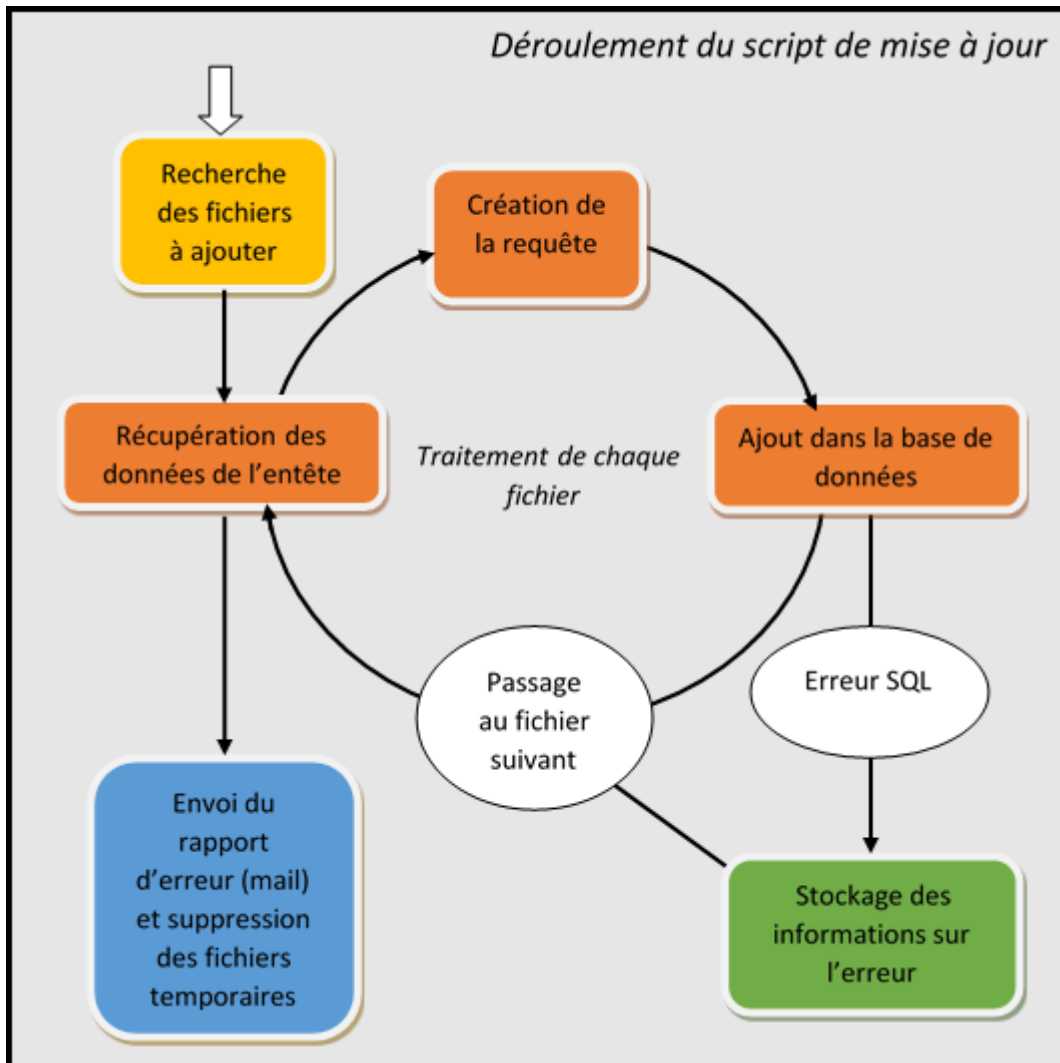


```
COMMENT
COMMENT N1 FITS FILE CREATED AT OCA BY THIERRY CORBARD
COMMENT
FILENAME= 'PIC_S0D2_RS_WL393DL_20140424_0953_v01_N1.fits' / File name
REVISION= 'v01' / File revision number
DATE = '2014-05-14T12:58:18' /file creation date (UTC)
INSTITUT= 'LATMOS' / Name of the home institution
ORIGIN = 'LATMOS+OCA' /Organization responsible for the data
OBSERVER= 'OCA-GIO' / Observers
CONTACT = 'F_MORAND' / Observing program PI name
RESCIENT= 'T_CORBARD' / Resident scientist present during acquisition
INSTRUME= 'SODISM2' / Instrument
CMP_NAME= 'PICARDSOL' / Name of the observing program
OBJECT = 'SUN' / Object
BUNIT = 'ADU' / Physical units
PHYSPARA= 'INTENSITY' / Physical parameter in the data array
DATE_OBS= '2014-04-24T09:53:00.000' / Date (UTC) of start image exposure
```

Le commentaire du mot-clé est facultatif mais le slash pour le délimiter est toujours présent.

Si la valeur n'est pas entre guillemets c'est soit un nombre soit un booléen.

Pour tous les fichiers fits à ajouter, les mots-clés et valeurs sont extraits et stockés afin de générer une requête SQL. Une fois l'en-tête traité, la requête est exécutée.



En cas d'erreur SQL lors de l'ajout des données du fichier les sont stockés dans un fichier temporaire : le nom du fichier, la requête, le message d'erreur MySQL et les requêtes à exécuter si le problème vient d'un mot-clé supplémentaire dans le fichier générant l'erreur. Une fois tous les fichiers traités, un mail est envoyé aux personnes concernées avec ces informations afin d'avertir qu'il y a une anomalie et de permettre de la corriger. Naturellement, les données du fichier ne sont alors pas stockées.

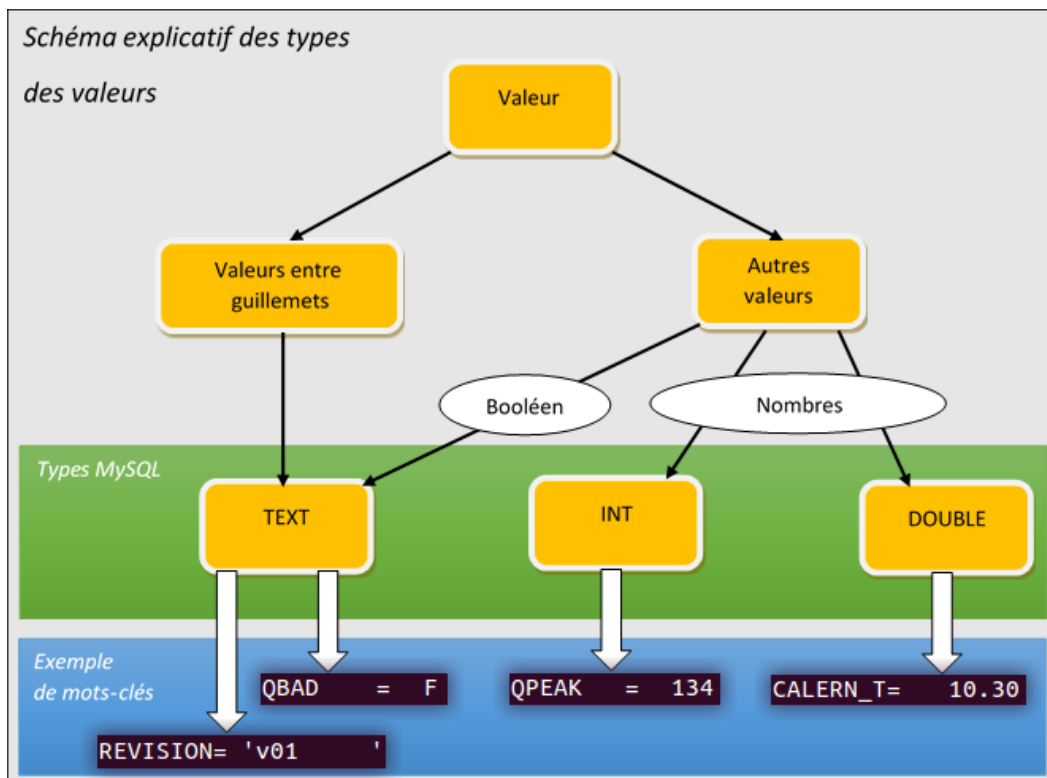
➤ **create_table.sh**

Comme son nom l'indique, ce script permet de créer une table dans la base de données afin de rajouter un nouvel instrument. Pour cela il faut spécifier le nom de la table et un fichier fits qui va servir de référence.

De façon similaire au script précédent, l'en-tête est traité à la différence près que le commentaire associé à chaque mot-clé est aussi récupéré pour la structure de la table.

Pour tous les mots-clés, la valeur est analysée afin de déterminer le type approprié dans la base de données :

- Pour les nombres à virgules, le type associé est 'DOUBLE'.
- Pour les entiers, le type associé est 'BIGINT', c'est un type dérivé de 'INT' qui permet de stocker de grandes valeurs.
- Les autres valeurs sont stockées en tant que texte (type 'TEXT').



Une fois l'en-tête traité, la requête est assemblée et exécutée. La table ainsi est créée dans la base de données.

➤ **fits_preview_generator.sh**

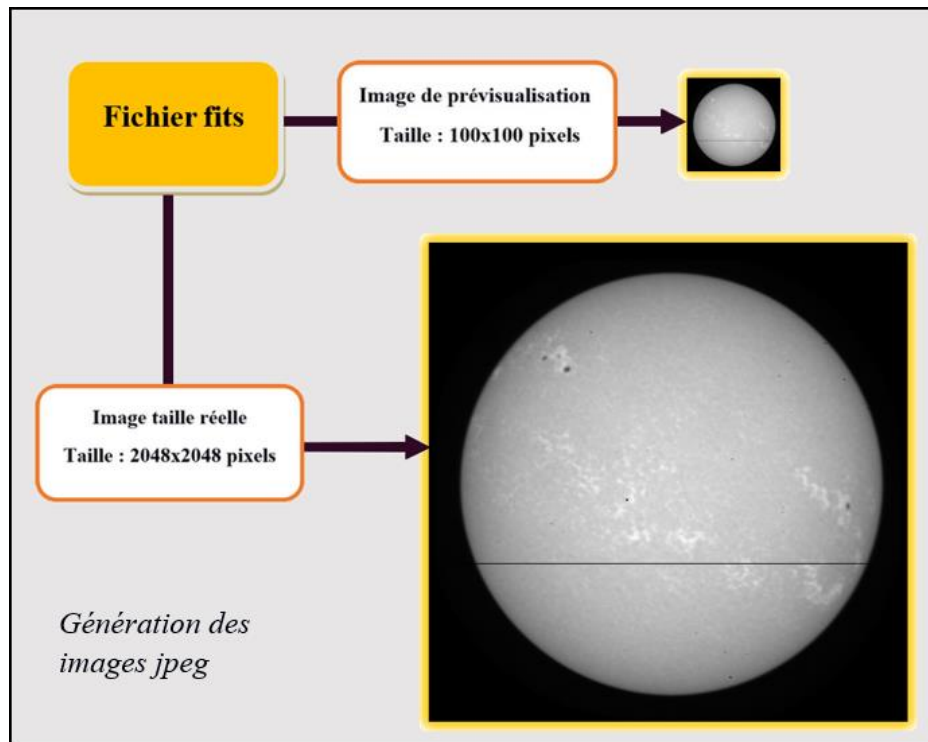
Ce script permet de générer les deux images jpeg de chaque fichier fits utiles au site web car les images des fichiers fits ne sont lisibles qu'avec un logiciel spécialisé.

Les images sont stockées dans 2 répertoires différents, un pour les images à taille réduite et un pour

celles à taille réelle.

Le script liste tous les fichiers fits dans le répertoire de recherche qu'on lui donne. Si une image a déjà été générée le script passe à la suivante, sinon les deux images jpeg sont créées.

Afin de mieux organiser les données j'ai fait en sorte que les images jpeg soient rangées comme les fichiers fits. L'arborescence des dossiers à partir du dossier racine contenant ceux-ci est recréée et les images gardent le même nom excepté l'extension qui devient '.jpg'. Il faut donc impérativement stocker les images dans deux dossier différents (un pour chaque taille d'image).



4.3. Interface web

➤ Accueil : [index.php](#)

J'ai tout d'abord créé une page d'accueil simple où l'utilisateur doit rentrer un mot de passe et son mail pour lui envoyer les informations permettant de récupérer les données qu'il souhaite avoir.

➤ Sélection des critères de recherche : [selection.php](#)

Je me suis ensuite occupé de la page de sélection des critères. Celle-ci se découpe en trois parties :

- Sélection des critères à utiliser et des valeurs associées.
- Option d'affichage des images de prévisualisation.
- Sélection et ordonnancement des critères de tri des résultats.

La sélection des critères se fait grâce au formulaire suivant :



Sélection instrument : • *sodism2*

Date d'observation (DATE_OBS): Du 1 : Janvier : 2011 : Au 1 : Janvier : 2011 :

- Longueur d'onde (WAVELNTH) : - Toutes les longueurs d'ondes sont sélectionnées -
- Qualité (QBAD): - Toutes les qualités sont sélectionnées -

Affichage des vignettes

Résultats ordonnés par :

- Date d'observation (DATE_OBS)
- Longueur d'onde (WAVELNTH)
- PHOTOMED

La création de la page s'est déroulée de la façon suivante : je l'ai tout d'abord créé avec une sélection simple des critères choisis par mon tuteur. Une fois cette partie fonctionnelle, j'ai ajouté un moyen de trier les critères par ordre croissant ou décroissant. Enfin, pour des raisons pratiques, la possibilité de changer l'ordre dans lequel les critères de tri sont appliqués a été ajoutée, chose qui était avant impossible sans changer le code. Cela permet d'améliorer la recherche.

Lorsqu'un critère n'est pas sélectionné, ses valeurs possibles sont masquées :

➔ Critère non sélectionné



Qualité (QBAD): ➔ - Toutes les qualités sont sélectionnées -

➔ Critère sélectionné



Qualité (QBAD): ➔

Une fois l'instrument et les critères sélectionnés et validés, les résultats sont affichés sur une nouvelle page.

➤ **Affichage des images trouvées : affichage.php**

Les informations des images trouvées incluent un champ pour les sélectionner et un aperçu (si l'option est activée) qui permet d'afficher l'image à taille réelle en cliquant dessus.

Pour voir l'en-tête d'un fichier, il faut cliquer sur le son nom qui est affiché sur un bouton. L'en-tête de l'image est alors affiché, sans les lignes de commentaires de l'en-tête qui n'ont pas été stockées dans la base de données. Ces lignes commencent dans la plupart des cas par le mot-clé '*COMMENT*'.

Sur cette page les résultats sont affichés dans un tableau pour les images, avec en plus des informations complémentaires qui peuvent être renseignés dans le fichier 'data.php' de l'instrument qui sert à la configuration (un exemple est disponible en annexe).


Tous les mots-clés affichés sur la page de résultats sont ceux renseignés dans le fichier de configuration de l'instrument. Une fonction de formatage des données est appelée à l'affichage pour chaque valeur. Si les valeurs d'un mot-clé doivent être affichés d'une façon particulière ou bénéficier d'un traitement (ex : formatage des données dans un format spécifique) un cas est défini dans la fonction pour le mot-clé. Toutes les valeurs du mot-clé seront alors affichées comme défini dans la fonction. Et si rien n'a été spécifié pour le mot-clé les valeurs sont affichées tel qu'elles sont dans la base de données au format textuel.

Résultats

Nombre max de résultats affichés : 1500

Informations :	
Nombre de fichiers trouvés :	339
Photomed moyen :	7854.151737463128
Photomed min :	152.618
Photomed max :	9885.422

(*)	Preview	FILENAME	DATE_OBS	WAVELNTH	QBAD	PHOTOMED
<input type="checkbox"/>		PIC_SOD2_RS_WL782DL_20140407_1206_v01_N1.fits	2014-04-07T12:06:00.000	782.37	F	3258.303
<input type="checkbox"/>		PIC_SOD2_RS_WL782DL_20140407_0726_v01_N1.fits	2014-04-07T07:26:00.000	782.37	F	7611.688
<input type="checkbox"/>		PIC_SOD2_RS_WL782DL_20140407_1151_v01_N1.fits	2014-04-07T11:51:00.000	782.37	F	2380.254
<input type="checkbox"/>		PIC_SOD2_RS_WL782DL_20140407_0727_v01_N1.fits	2014-04-07T07:27:00.000	782.37	F	7685.457


 Adrien Boucher
Observatoire de la Côte d'Azur

Depuis cette page, on a la possibilité via le bouton de téléchargement de récupérer l'url d'un dossier contenant des liens symboliques sur les images sélectionnées pour pouvoir les télécharger. En plus de l'url une commande pour les télécharger sous linux est donnée.

Information de téléchargement

Fichiers disponibles ici : [lien](#)

Commande Linux pour télécharger les images (pensez à changer le dossier de destination) :

```
wget -r -np -nH --cut-dir=2 -A.fits -P /Dossier/de/destination/ https://solar-data.oca.eu/liens/20140624_04h47m12s_53a98f7003d42/
```

Un mail vous a été envoyé avec les informations de téléchargement à cette adresse mail : 'adrien.boucher@oca.eu' .

Ces informations sont affichées sur une nouvelle page et également envoyées par mail à l'utilisateur et aux responsables des données pour effectuer un suivi de l'utilisation du site web.

Le mail est envoyé en copie aux personnes responsables des données.

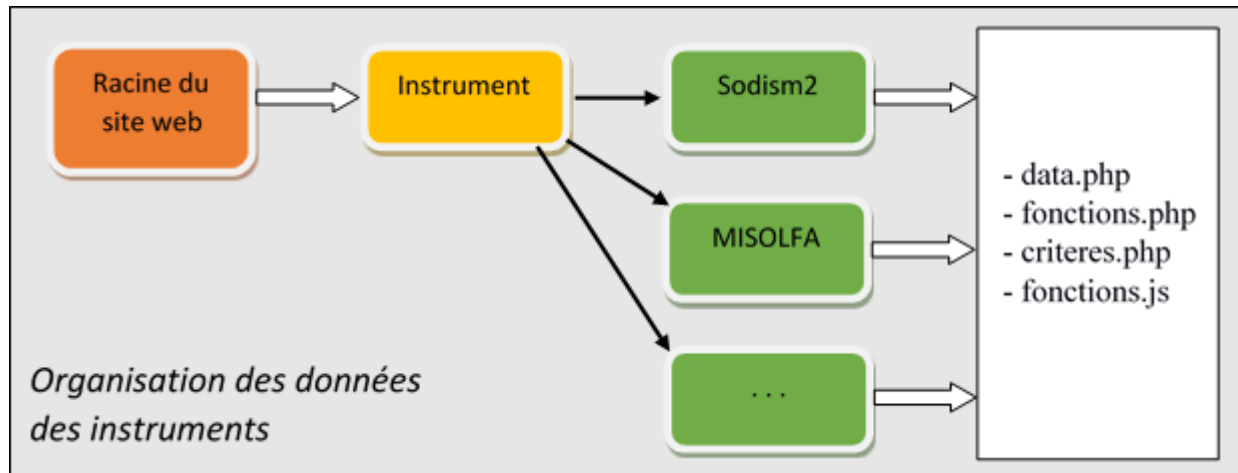
Les différentes pages du site ont été décorées avec un style homogène. Le fond est le même sur toutes les pages : c'est un dégradé bleu et le pied de page est identique, donnant une identité au site.

Après avoir développé ces fonctionnalités, j'ai aménagé le code du site web pour faciliter l'ajout de critères de sélection et d'instruments.

Les constantes et certaines informations utiles pour le site ont été regroupées dans un fichier de

configuration. Cela permet une meilleure visibilité et de facilement moduler le site.

Les fonctions spécifiques à un instrument ont été regroupées dans deux fichiers, un pour le PHP et un pour le JavaScript. Pour chaque instrument il y a un dossier contenant ces fichiers. Le contenu des fichiers est adapté à l'instrument.



L'ajout d'un mot-clé à afficher est devenu très simple grâce à l'automatisation de l'affichage des informations, pour chaque image trouvée dans la base de données les valeurs des mots-clés listés dans le fichier de configuration de l'instrument sont extrait un par un et affichés dans l'ordre spécifié. Il suffit d'ajouter un mot-clé dans cette liste pour qu'il soit affiché dans la page de résultats.

A chaque changement d'instrument, la page de sélection des critères est rechargée pour afficher les bons critères de sélection. Cette partie de la page est également stockée dans le répertoire de l'instrument.

Les pages d'affichage et de sélection chargent toujours les données spécifiques à l'instrument sélectionné qui est par défaut SODISM2.

Le regroupement de tous les éléments spécifiques à un instrument dans un dossier permet ainsi le chargement dynamique du contenu des pages et des méthodes de traitement des données.

4.4. Documentation

Afin de faciliter la compréhension, l'utilisation et la modification du site et des scripts, j'ai rédigé de la documentation qui vient compléter les commentaires des codes sources. Une partie de la documentation concerne l'utilisation des scripts. La manière de les utiliser y est décrite de façon détaillée. Les arguments attendus sont expliqués, le déroulement général des scripts est décrit.

L'autre partie de la documentation concerne le site. Elle décrit les procédures à suivre afin de modifier certaines parties du site qui ont justement été créées dans l'optique de pouvoir changer leur contenu afin d'adapter au mieux le site aux besoins des utilisateurs.

Cela vient compléter les commentaires écrits dans les codes source, qui servent à décrire le fonctionnement de ces derniers.

La documentation est disponible en annexes.

5. Tests

Parallèlement au développement, des séries de tests ont été effectués.

Les scripts en Bash ont été testés de plusieurs façons. Dans un premier temps, chaque fonctionnalité des scripts a été testée séparément avec un jeu de données grâce à des petits scripts pour effectuer des tests unitaires. Les valeurs obtenues ont été comparées avec celle attendues et affichées afin de vérifier la validité de la fonctionnalité testée.

Voici un exemple de test unitaire sur la fonction de traitement des lignes des en-têtes afin de récupérer les mots-clés et valeurs.

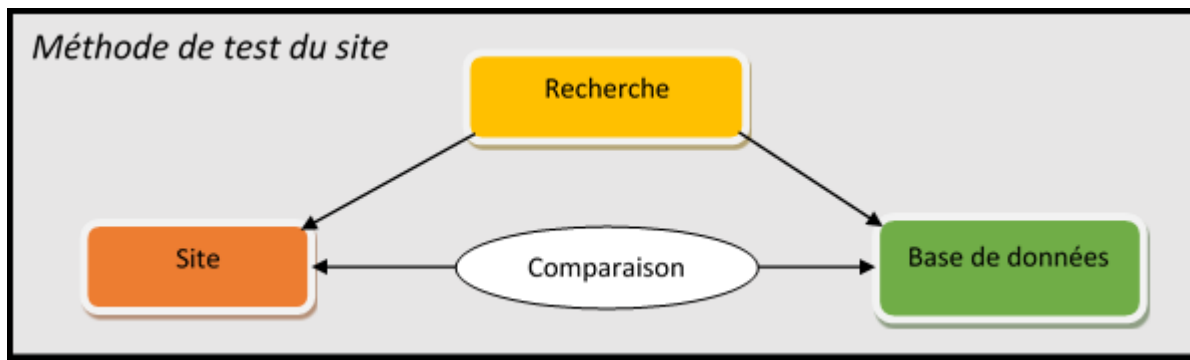
```
>./TU_parsing.sh jeu_test/tu_ligne_1
Chargement fichier - jeu_test/tu_ligne_1
Ligne : 'OBS_SITE= 'Calern / OCA'           / Observatory   '
--
Mot cle attendu  ='OBS_SITE'
Mot cle obtenu   ='OBS_SITE'
--
Valeur attendue ='Calern / OCA'
Valeur obtenue  ='Calern / OCA'
--
Test Unit - OK
```

Le traitement des en-têtes a notamment été testé avec une grande attention. C'est en effet un point critique dans la mise à jour des données. Si ceux-ci ne sont pas récupérés correctement, cela peut générer des erreurs par la suite et empêcher l'utilisation correcte des données sur le site.

Tous les différents formats de données qu'il est possible de rencontrer dans un en-tête ont été recensés et testés.

Dans un second temps, les scripts ont été testés avec un échantillon des données varié pour être le plus représentatif possible afin de vérifier le fonctionnement général, détecter les anomalies possibles et vérifier les résultats obtenus.

Le site web a aussi été testé à partir d'un échantillon des données afin de vérifier le bon fonctionnement du tri des fichiers avec différents critères, de l'affichage correct des données trouvées ainsi que le téléchargement, la sélection des images et l'affichage des en-têtes.



6. Résultats

6.1. État final du produit par rapport au cahier des charges

En reprenant le cahier des charges initial, les parties suivantes ont été réalisées et finalisées :

- Création de la base de données.
- Mise en place du système de mise à jour automatique de la base de données.
- Création du site web pour interroger la base de données.
- Mise en place d'un moyen de téléchargement des images.

L'archive a été remplacée par un lien de téléchargement afin de ne pas copier les données inutilement.

6.2. Difficultés rencontrées

Je n'ai pas rencontré de difficultés majeures dans le développement du site ou des scripts de mise à jour de la base de données.

Néanmoins l'optimisation des scripts s'est révélée être une tâche à laquelle j'ai consacré plusieurs heures de travail pour accélérer le traitement des nombreuses données. Cela m'a permis de passer de 3 secondes pour ajouter un fichier dans la base de données à seulement 0.2 seconde. Il reste tout de même la génération des images jpeg qui prend environ 3 secondes par fichier pour créer les deux images d'aperçu pour le site.

Il a fallu aussi prendre en compte le fait que certains fichiers diffèrent des autres. Idéalement la table de la base de données doit contenir tous les mots-clés présents dans les fichiers fits produits

par l'instrument. Mais il arrive que des fichiers possèdent des mots-clés supplémentaires, dans ce cas-là un mail est envoyé aux personnes pouvant effectuer les modifications nécessaires dans la base de données pour pouvoir intégrer le fichier. Les modifications à faire sont décrites dans le mail.

6.3. Quantifications

Le code fourni peut être divisé en deux grandes parties : les scripts Bash et le site web qui utilise les langages HTML, CSS, JavaScript et PHP.

	Lignes de code	Lignes de commentaires	Total
Scripts			
update_fits	200	110	310
fits_preview_generator	101	50	151
create_table	98	55	153
Total	399	215	614
Site web			
HTML/PHP/JavaScript	975	137	1112
CSS	85	0	85
Total	1060	137	1197
Total produit	1459	352	1811

Les scripts sont abondamment commentés afin de faciliter la compréhension et la prise en main du code.

6.4. Planning effectif

Tâche \ Semaine	Avril		Mai				Juin			
	17	18	19	20	21	22	23	24	25	26
Compréhension du sujet										
Recherche										
Cahier des charges										
Spécifications										
Développement et documentation										
Tests unitaires										
Tests de validation										
Rapport de stage / Soutenance										

Conclusion

Durant ce stage à l'Observatoire de la Côte d'Azur, j'ai réalisé un site web de consultation et de récupération des données scientifiques de l'instrumentation PicardSol, couplé à un système de mise à jour des données automatique afin de pouvoir accéder aux données des dernières observations.

J'ai tout d'abord travaillé à la gestion des données, pour les stocker et les mettre à jour afin qu'elles soient utilisables. Une fois cette partie réalisée, je me suis consacré à l'élaboration du site web, pour mettre en œuvre ces données

Pour cela, j'ai dû apprendre à gérer des grandes quantités de données. J'ai pu voir l'intérêt de l'optimisation des programmes avec des temps de traitement très longs au vu de la quantité d'informations à traiter. Les fichiers pouvant différer légèrement sur certains points en fonction de la manière dont ont été faites les observations, il a été nécessaire de faire en sorte que les données soient bien traitées dans tous les cas.

Il a fallu aussi faire en sorte que le site web soit facilement modulable et que le code soit compréhensible, pour permettre l'amélioration ou l'ajout de données provenant d'autres instruments que ceux existants. Le travail s'est fait de façon autonome.

J'ai pu aussi voir l'importance et l'intérêt des interactions entre le client et le développeur, afin de répondre au mieux à ses besoins, qui ne sont pas toujours faciles à comprendre. Le dialogue permet de rendre le produit beaucoup plus agréable à utiliser et plus pratique pour le client. Il faut alors savoir s'adapter et satisfaire au plus vite les nouvelles demandes.

Ce stage m'a permis de découvrir le monde du travail mais aussi de la recherche. J'ai eu la possibilité d'observer les fortes interactions qu'il y a tant au niveau humain que logiciel qui se font de l'acquisition des données à leur utilisation par les chercheurs. J'ai pu voir plusieurs logiciels coopérer avec des scripts fonctionnant en permanence qui les complètent et avec des services web pour afficher certaines des données traitées. Cela m'a permis de découvrir ce qu'est réellement un projet coopératif étant donné que plusieurs laboratoires sont impliqués, et même plusieurs pays. Je suis ravi d'avoir pu apporter ma contribution au projet PicardSol. Ce stage m'a motivé à améliorer

encore plus ma façon de développer afin de fournir un travail le plus propre possible et surtout bien documenté, pour que d'autres puissent le comprendre, le modifier ou le faire progresser. Il a aussi suscité en moi de l'intérêt pour le domaine de la science, qui fait énormément appel à l'informatique pour faire avancer la recherche.

Glossaire

OCA : Observatoire de la Côte d'Azur

CRAAG : Centre de Recherche en Astronomie, Astrophysique et Géophysique (Alger, Algérie)

FITS (Flexible Image Transport System) : Format de fichier le plus utilisé en astronomie

MySQL : Système de gestion de base de données

PHP (Hypertext Preprocessor) : Langage de programmation impératif orienté objet, il sert à produire des pages web dynamiquement via un serveur HTTP.

PICARD : Microsatellite utilisé pour mesurer différentes informations sur le soleil telles que la vitesse de rotation, la puissance rayonnée, les tâches à la surface, le diamètre. Il a servi à mieux comprendre l'influence du soleil sur la Terre.

SODISM (Solar Diameter Imager and Surface Mapper) : Télescope de 11cm de diamètre qui prend des images du soleil avec une résolution de 2048 par 2048 pixels.

Limbe : Bord du disque d'un astre.

ASCII (American Standard Code for Information Interchange): Norme de codage des caractères en informatique qui a été prévu pour écrire en anglais.

Observatoire virtuel : Ensemble de données, outils logiciels rassemblés et disponibles par internet pour conduire des recherches scientifiques. Un observatoire virtuel est composé de plusieurs centres de données.

Bibliographie

Documentation MySQL :

<http://dev.mysql.com/doc/>

Documentation PHP :

<http://www.php.net/manual/fr/>

Présentation de l'OCA :

<https://www.oca.eu/spip.php?rubrique406>

Présentation du laboratoire Lagrange :

https://www-n.oca.eu/lagrange/PDF/DS3_LAGRANGE_UMR7293.pdf

Introduction au projet PicardSol :

<https://solar-physics.oca.eu/spip.php?article229>

Logiciel FitHAS :

<http://www.china-vo.org/fithas.html>

Site web exemple – Solar Influences Data Analysis Center :

<http://sidc.oma.be/uset/searchForm.php>

Annexes