Observatoire Astronomique
de Strasbourg

Université
Louis Pasteur

THÈSE DE DOCTORAT

présentée par

Philippe Poinçot

le 15 décembre 1999

Classification et recherche d'information bibliographique par l'utilisation des cartes auto-organisatrices, applications en astronomie.

Présidente du jury :	A. Acker
Directeur de thèse :	F. Murtagh
Rapporteurs :	C. Chrisment
	F. Genova
	J. Lequeux

Table des matières

Chapitre 1

Introduction

L'astronomie tend à devenir la première science "tout-numérique" (Heck et Murtagh, 1993). À la source se trouve l'imagerie astronomique qui fait maintenant appel aux détecteurs numériques, que l'on retrouve aussi bien embarqués dans les véhicules spatiaux (sondes et satellites) que montés au foyer des télescopes au sol. Les données obtenues subissent alors des traitements numériques (élimination de bruits parasites, corrections d'artéfacts instrumentaux) que les astronomes effectuent avec les logiciels adaptés, puis sont archivées et mises à la disposition de la communauté astronomique. Depuis 1993, année de l'apparition du navigateur Mosaic, un nombre croissant de ces données est accessible en ligne, via Internet.

Depuis sa création en 1972, le Centre de Données astronomiques de Strasbourg (CDS) développe des services qui permettent de regrouper, organiser et diffuser cette information. La base de données SIMBAD[1] qui y est maintenue, contient actuellement des données observationnelles sur plus de 2,7 millions d'objets astronomiques, les références de plus de 108.000 publications ainsi que plus de 3 millions de citations d'objets dans ces publications. Le CDS s'interesse en particulier à l'information bibliographique : citation des objets dans les articles pour SIMBAD, mise en ligne de tables publiées dans le service catalogues et VIZIER[2].

Le volume des données est bibliographique semble toujours s'accroître exponentiellement (Abt, 1998), et il est important de chercher à mettre en œuvre des méthodes innovantes d'accès à cette information. C'est l'objectif de cette thèse. Nous proposons un système qui permet la visualisation globale d'un ensemble de données bibliographiques pour l'exploration de son contenu, tout en permettant la recherche de documents sur des sujets précis. Ce système, la *carte bibliographique*, fait appel à un réseau de neurones pour effectuer une classification des données. Une interface graphique, consultable via Internet[3], est construite autour de cette classification pour permettre différents types d'interrogation et de visualisation.

Dans la première partie de cet ouvrage, nous rappellerons les fondements de la recherche d'information. Ensuite, nous détaillerons le fonctionnement des réseaux de neurones en distinguant les deux types que constituent les réseaux à apprentissage supervisé

[1] Set of Identifications, Measurements and Bibliography for Astronomical Data.

[2]

[3] http ://simbad.u-strasbg.fr/A+A/map.pl, http ://simbad.u-strasbg.fr/ApJ/map.pl

et ceux à apprentissage non supervisé dont font partie les cartes auto-organisatrices que nous utilisons. Nous terminerons cette première partie par une revue des principaux types de systèmes de recherche d'information.

La seconde partie de cet ouvrage est consacrée à la carte bibliographique. Après une description des données dont nous disposons, nous nous intéresserons à l'apprentissage de notre réseau, ainsi qu'au contrôle du déroulement de celui-ci. Nous tenterons alors d'en déterminer les paramètres optimaux. Ensuite, nous ferons une description détaillée de l'architecture de notre système et de son interface de consultation, avec ses différentes fonctionnalités. Nous détaillerons ensuite la mise à jour de la carte bibliographique ainsi que les différentes techniques utilisables pour la prise en compte des nouveaux documents de la base. Quelques visualisations de résultats de classification seront ensuite commentés, de même que les résultats de deux apprentissages similaires pour illustrer les aspects stables et instables des classifications obtenues avec les cartes auto-organisatrices. Le dernier chapitre mettra en avant les apports de la carte bibliographique par rapport aux systèmes habituels de recherche d'information documentaire en astronomie, par une comparaison avec le système de recherche de l'Astrophysics Data System (ADS).