

# ADWUM

Classification automatique, méthodes neuronales et web usage mining  
Classificação automática, métodos neuronais e web usage mining



Yves Lechevallier AXIS/INRIA  
Francisco de A. T. de Carvalho CIn / UFPE  
[www.cin.ufpe.br/~fatc/CLADIS](http://www.cin.ufpe.br/~fatc/CLADIS)



## Descriptif

Grâce à la technologie informatique, de vastes ensembles de données sont recueillis et il est nécessaire de les résumer. Les objectifs du projet CLADIS, financé par la coopération CNPq-INRIA (2000-2002) et le projet européen ASSO (2001-2004) consistait à construire automatiquement des classes homogènes à partir de ces données et de définir de nouvelles unités statistiques décrivant ces classes via l'utilisation de mesures de proximités.

Actuellement nos thèmes de recherches communs sont proches de la problématique du "Data Mining" et tout particulièrement nous nous intéressons à :

- Construire des résumés à partir d'ensembles d'informations hétérogènes
- Développer de nouvelles méthodes de classification (méthodes neuronales, nouvelles distances ou mesures de proximité, approche conceptuelle,...) et étudier leur pertinence dans le cadre du WUM
- Concevoir de nouvelles représentations des données complexes

## Résultats obtenus

Dans le cadre ADWUM nous avons commencé à analyser les « logs » du Centre d'Informatique de l'université (UFPE) de Recife.



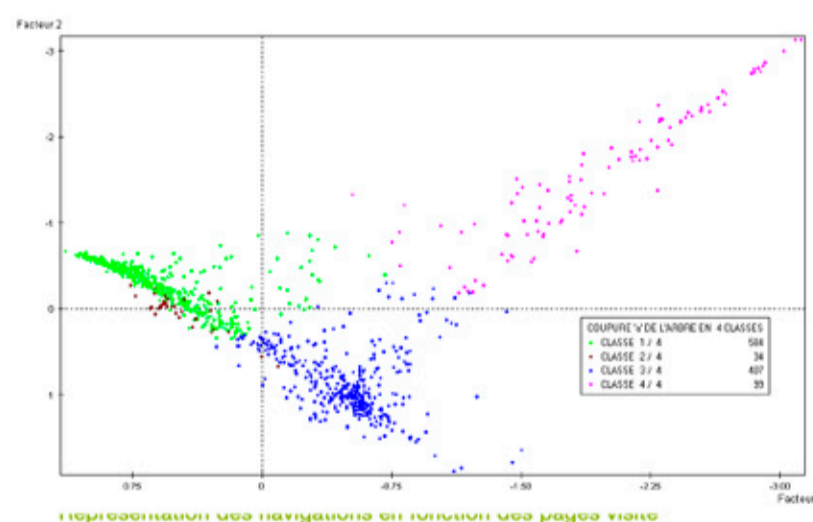
Web site du CIn-UFPE ([www.cin.ufpe.br](http://www.cin.ufpe.br))



Structure sémantique du Web site du CIn-UFPE

Deux sources d'informations ont été utilisées pour l'analyse de ce site

1. Les **traces d'usages** des internautes sur ce site (fichiers « logs »)
2. Les **informations sur le site** (sa structure, le contenu des ressources,...), informations de nature « Web sémantique ».



Représentation des navigations en fonction des pages visitées



Représentation des navigations en fonction des pages visitées

## Publications communes

F.A.T. De Carvalho, R.M.C.R. De Souza, M. Chavent et Y. Lechevallier.

**Adaptive Hausdorff distances and dynamic clustering of symbolic data.** Pattern Recognition Letters, 2005 (article in press)

F.A.T De Carvalho, Y. Lechevallier et R. M. C. R De Souza.

**A dynamic cluster algorithm based on adaptive Lr distances for quantitative data.** Proceedings of the 9th Conference of the International Federation of Classification Societies. Springer-Verlag, pp 33 – 42, 2004.

R.M.C.R De Souza, D F.A.T De Carvalho et Y. Lechevallier.

**Dynamic cluster methods for interval data based on Mahalanobis distances.** Proceedings of the 9th Conference of the International Federation of Classification Societies. Springer-Verlag, pp 351 – 360, 2004.

M. Csernel, F.A.T De Carvalho et Y. Lechevallier.

**Partitioning of constrained symbolic data,** IFCS'2004, Chicago, USA, 15-18 july, 2004.

M. Chavent, F.A.T De Carvalho, Y. Lechevallier et R. Verde.

**Trois nouvelles méthodes de classification automatique des données symboliques de type intervalle.** Revue de Statistique Appliquée. Paris (France) : v. LI, n. 4, pp 5-29, 2003.