

Tanagra, un logiciel de Data Mining gratuit pour l'enseignement et la recherche

Ricco Rakotomalala

February 19, 2010

Résumé

Tanagra est un logiciel "open source", librement diffusé sur internet depuis janvier 2004. Il s'inscrit dans la lignée des outils librement accessibles de la communauté du Data Mining tels que Weka, Knime, RapidMiner, Orange, etc. Tous ont pour point commun de proposer, entres autres, un mode de fonctionnement sous forme diagramme de traitements, et d'offrir à l'utilisateur des techniques issues d'horizons divers. Ils possèdent une qualité essentielle pour un logiciel fait par et pour des chercheurs, le code est librement accessible, tout le monde peut vérifier ce qui est réellement implémenté.

Dans notre exposé, nous présenterons le logiciel Tanagra que nous avons développé. Dans un premier temps, nous essaierons de mettre en avant les spécifications qui ont conduit à son élaboration, son évolution au fil du temps, et les éléments périphériques qui accompagnent sa diffusion. Nous parlerons ainsi de l'élément principal de succès du logiciel qui est une production intense en matière de documentation, assurant une fréquentation relativement élevée de notre site (près de 650 visiteurs par jour sur le mois de janvier 2010).

Dans un deuxième temps, nous le comparerons aux principaux logiciels libres, largement reconnus dans notre communauté : Orange, Knime, RapidMiner, Weka ; mais aussi au logiciel R, plutôt d'obédience statistique, mais qui a fait une percée spectaculaire auprès des informaticiens ces dernières années. Nous nous baserons principalement sur les étapes clés du Data Mining pour évaluer les solutions qui ont été mises en place par les différents outils. Des petits exemples didactiques permettront de juger de leur comportement réel. Nous concluons en essayant de positionner ces outils par rapport aux besoins des différents profils d'utilisateurs.

Mots-clés : data mining, logiciel libre, Tanagra, applications