# Segmentation des fichiers logs

Hassan Saneifar*,**, Stéphane Bonniol **, Pascal Poncelet*, Mathieu Roche*

*LIRMM, CNRS, Université Montpellier 2; **Satin Technologies

## Résumé

Avec la méthode de segmentation appelée passages de discours, la reconnaissance des divisions logiques de documents est essentielle. Cela s'avère plus difficile dans les documents ayant des unités logiques différentes de celles trouvées dans les textes classiques comme les paragraphes ou les sections. Ainsi, nous proposons une méthode automatique pour caractériser les unités logiques complexes propres à ce type de document en fonction de certaines caractéristiques. Ensuite, un processus d'apprentissage supervisé est mis en place afin de pouvoir reconnaître les unités logiques. Les résultats obtenus en utilisant des données issues du monde industriel sont encourageants.

## Summary

Several application areas require methods for segmenting documents. Depending on the characteristics of our domain, we choiced the segmentation method called "discourse passages" which is based on the identification of "logical units of documents". Thus, we propose here a method to characterize complex logical units found in this type of documents according to their characteristics. Then, a supervised learning process is used to recognize these logical units. Experimental results on the recognition of complex logical units in the log files from the industrial world are encouraging.