

## الخلاصة

في مساهمة علمية سابقة تتعلق بنمذجة الخبرة في قواعد البيانات وتطبيقها على مدونات الأحاديث النبوية الشريفة، قمنا بدراسة وتطوير نموذج للتنقيب الآلي عن النصوص باستخدام الطرق الإحصائية لاستكشاف أوجه التشابه المعنوي بين نصوص هذه المدونات. الهدف من عملنا في هذا البحث هو دراسة الطرق الجديدة للتصنيف والتجزئة الموضوعية للمعارف في قواعد المعطيات النصية مثل تلك الخاصة بالأحاديث النبوية الشريفة. هذه الطرق تمثل عملية جديدة لتكوين مجموعة المعلومات والإجابات المقترحة من طرف نظام البحث عن المعلومات. التصنيف الموضوعي يسمح للمستخدم بتوجيه الاستكشاف بدلالة المواضيع العامة للأصناف وبالتالي الوصول بسرعة إلى الأهداف المرجوة من بحثه، أما التجزئة الموضوعية فإنها تسهل له استعراض الأجزاء الموضوعية التي يعتقد أنها على صلة بموضوع البحث وبالتالي إعادة ترتيب بعض الوثائق الطويلة والتي لا تمثل فيها المعلومات المبحوث عنها إلا واحد من المواضيع التي تحتويها هذه الوثيقة. هذا العمل يمكن تطبيقه في الحالات التي يكون لدينا فيها مجموعة كبيرة من البيانات مثل تلك المتعلقة بالحديث النبوي الشريف.

**الكلمات المفتاحية:** استكشاف المعارف، التنقيب عن البيانات، التنقيب عن النصوص، أنظمة البحث عن المعلومات، التصنيف الموضوعي، التجزئة الموضوعية.

## Résumé

Dans un travail antérieur, sur la modélisation d'expertise dans les bases de données avec application au corpus des Traditions Prophétiques (*Hadith*), nous avons étudié puis développé un modèle automatique de fouille de texte basé sur des méthodes statistiques afin de découvrir les similarités sémantiques entre les textes en question.

A présent, notre recherche a pour but d'étudier les nouvelles méthodes de classification et de segmentation des connaissances dans les bases de données textuelles telles que celles des Traditions Prophétiques (*Hadith*). Ces méthodes représentent un processus de structuration de l'ensemble des informations et des réponses fournies par un système de recherche. La classification thématique permet à l'utilisateur d'orienter son exploration en fonction des thématiques générales des classes et d'accéder ainsi plus rapidement au but de sa recherche. La segmentation permet de présenter à l'utilisateur les segments textuels jugés pertinents et de mieux positionner certains documents longs dans lesquels l'information recherchée ne constitue que l'une des thématiques abordée. Ce travail s'applique dans le cas où nous avons une quantité trop importante de données ce qui est le cas des corpus du *Hadith*.

**Mots clés:** Découverte de connaissances, Fouille de données, Fouille de textes, Système de recherche d'information, Classification thématique, Segmentation thématique.

## Abstract

In a previous work on modeling expertise in databases with application to the corpus of Prophetic Traditions (*Hadith*), we studied and developed a model of automatic text mining based on statistical methods to discover semantic similarities between texts.

At present, our research aims to study new methods of classification and segmentation of knowledge in textual databases such as those of Prophetic Traditions (*Hadith*). These methods provide a process for structuring all the information and answers provided by a Information Retrieval System. The topic classification allows the user to direct its exploration based on general themes of class and access them more quickly to the goal of his research. Topic segmentation allows the user to explore the relevant segments and to better position some long documents in which the searched information presents one of the topics discussed. This work is usable if we have a lot of data which is the case of *Hadith*.

**Keywords:** Knowledge discovery; Data mining; Text mining; Information retrieval systems; Topic classification; Topic segmentation.