



Exploration des traces de navigation sur le Web

Djerbaoui Imad Eddine / Herrouz Hichem ; Université Kasdi Merbah Ouargla
 Emails: jdjerbaoui1992@gmail.com / Herrouz.hichem@gmail.com

Introduction

Aujourd'hui, plusieurs techniques sont utilisées pour collecter les données à partir des ressources du Web. Ces techniques sont les résultats des travaux du Web-Mining. Le processus du Web Mining se divise, selon le type de données à exploiter, en trois principales catégories : Web Content, Web Structure et Web Usage Mining. Chaque catégorie a son propre processus d'exécution et des différentes techniques d'application. Cependant, Plusieurs recherches ont été publiées dans ce cadre, et proposent des algorithmes, ou des technologies, qui permettent de collecter et ensuite de comparer ces mêmes données. Dans notre travail, nous nous intéressons en particulier aux outils d'extraction de la connaissance à partir des usages du Web (WUM – Web Usage Mining).
 Mots-clés : *Datamining, Web Mining, Web Usage Mining, fichiers logs, traces de navigation, prétraitement des logs.*

Problématique ?



Les données de navigation sauvegardées dans un fichier



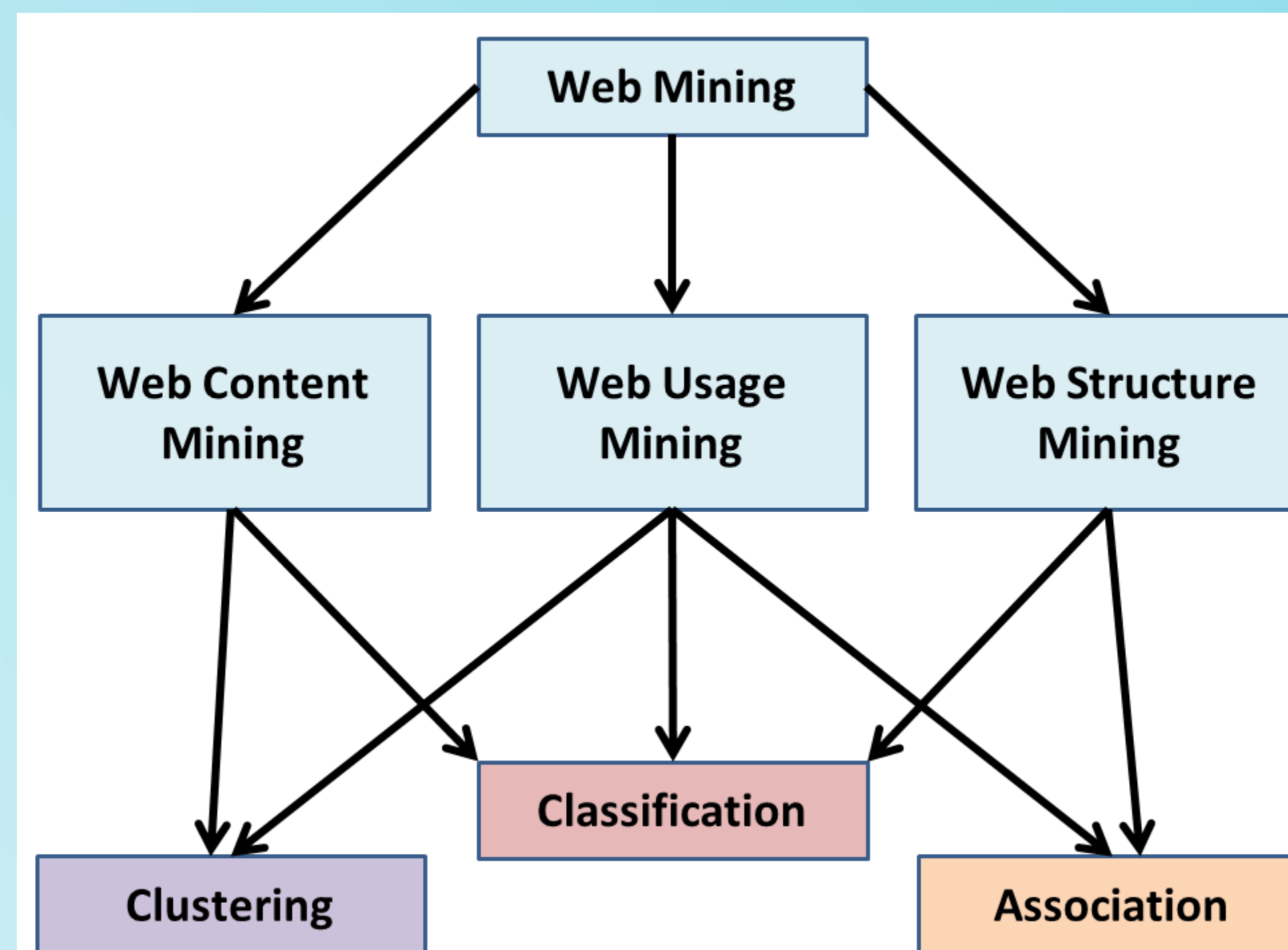
Problématiques ?



Trop données mais pas assez d'informations

C'est quoi Le Web-mining ?

La **fouille du Web (Web-Mining)** est l'application des techniques d'exploration de données en vue de découvrir des constantes, schémas ou modèles, dans les ressources d'Internet ou les données le concernant. Selon ses cibles, la fouille du Web peut être divisée en trois types : la fouille de l'usage du Web, la fouille du contenu du Web et la fouille de la structure du Web.

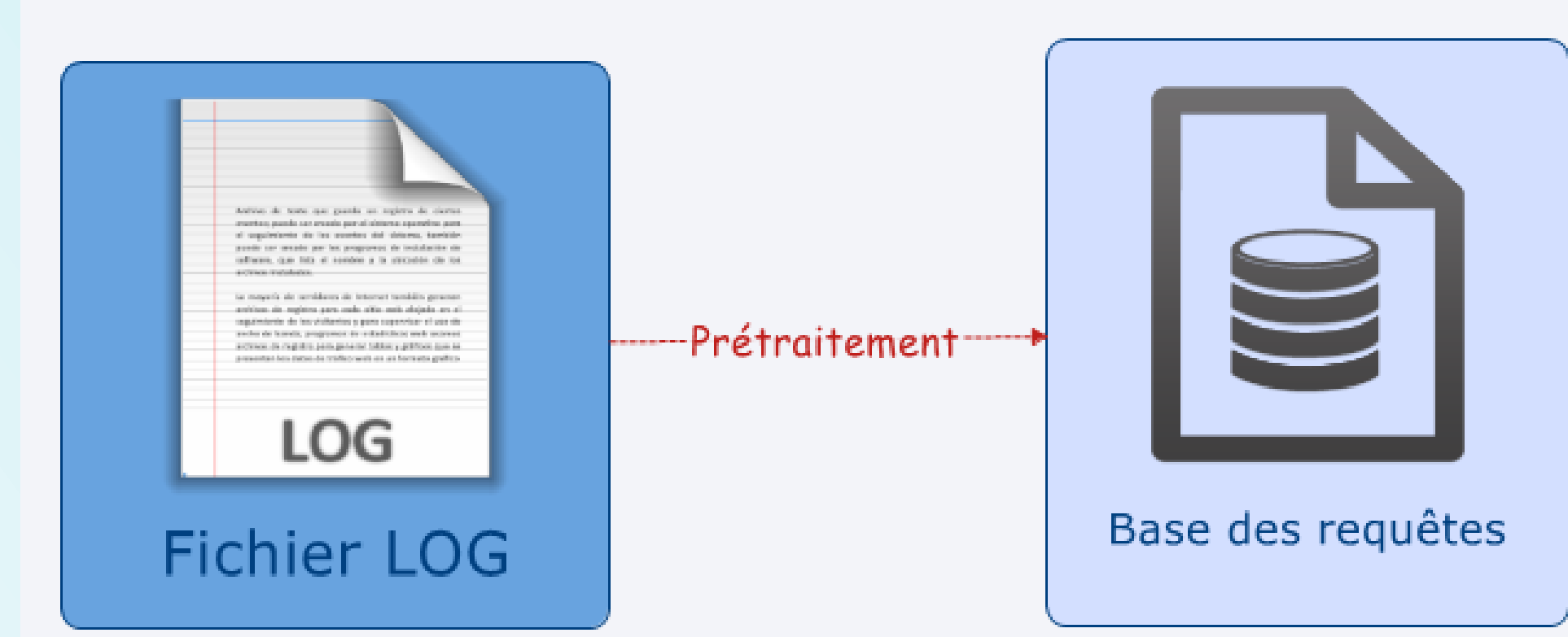


Web Usage Mining (WUM)

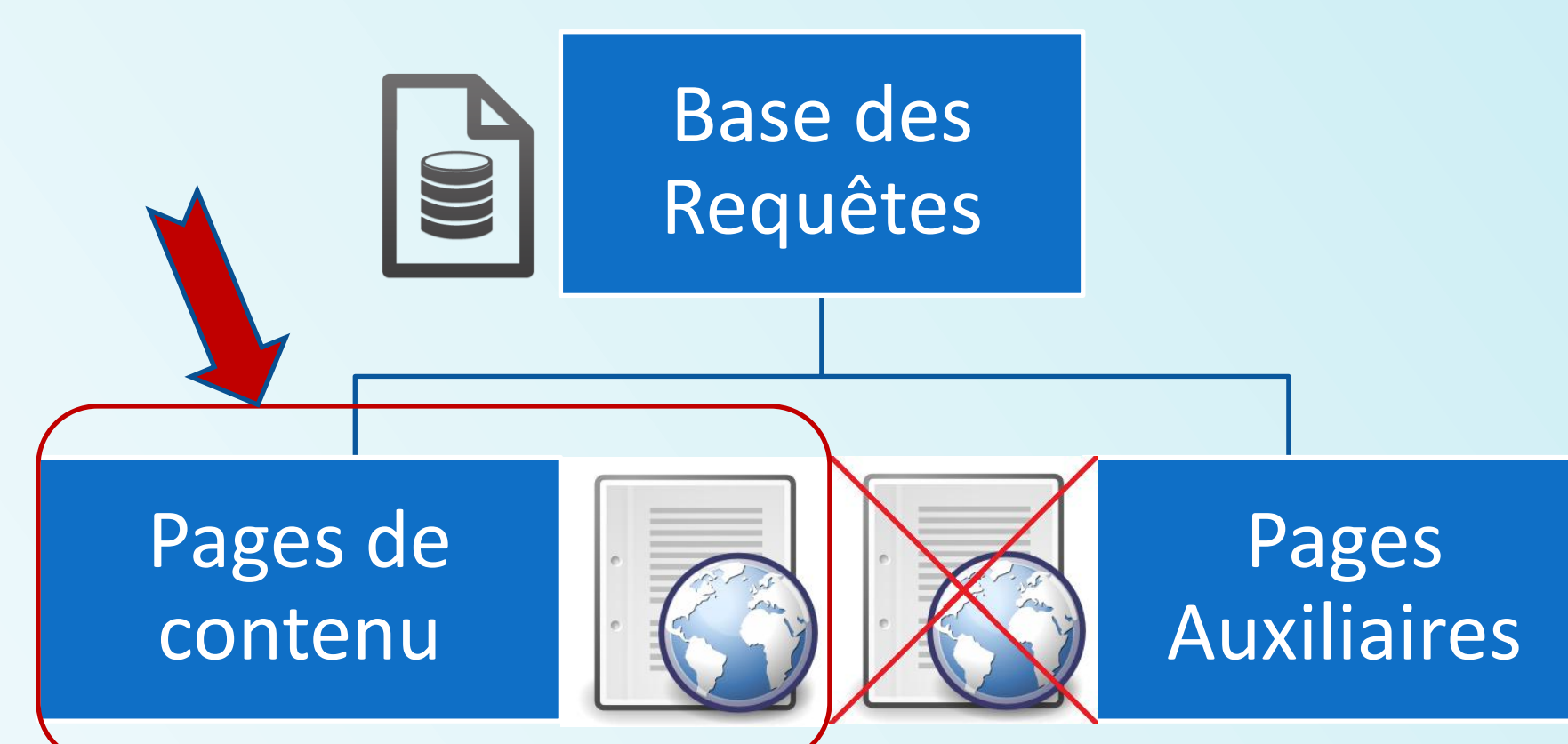
C'est l'application des techniques du Data Mining pour découvrir des modèles d'usage intéressants à partir des données du Web, et cela, pour mieux comprendre et bien servir les besoins des applications Web. Les données d'utilisation (Usage Data) capturent l'identité et l'origine des utilisateurs Web tout au long de leurs navigations dans un site Web.

Approche proposée

1. Prétraitement

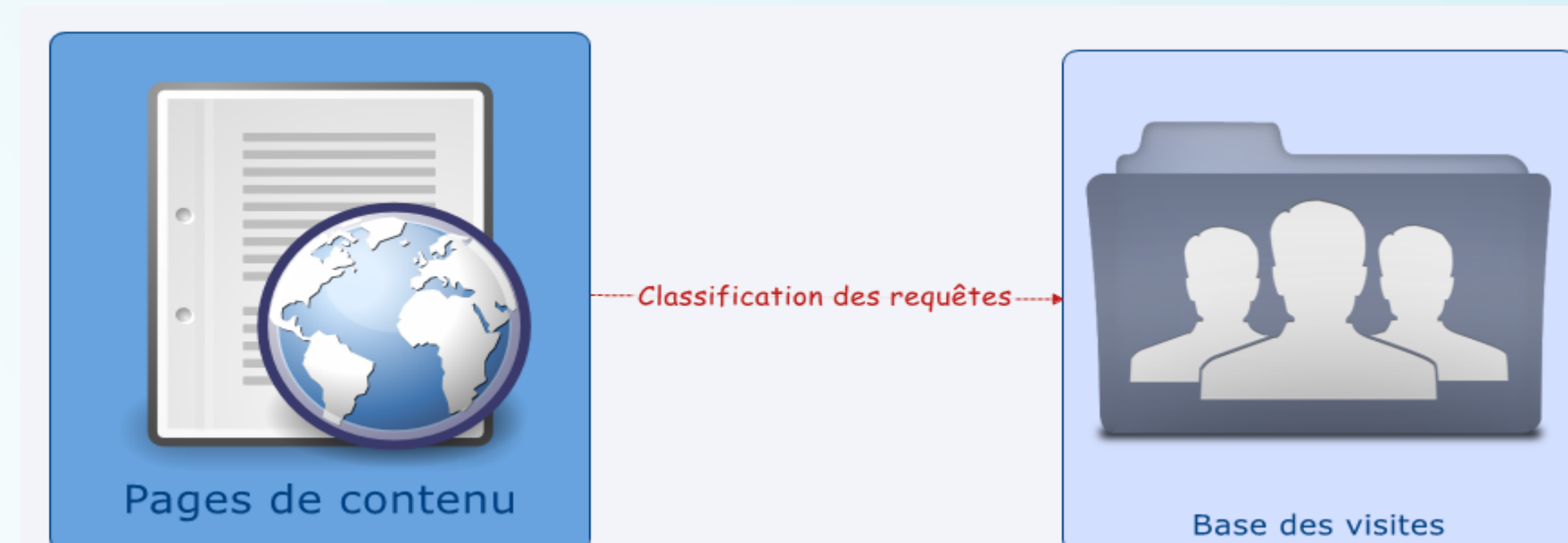


2. Classification des pages

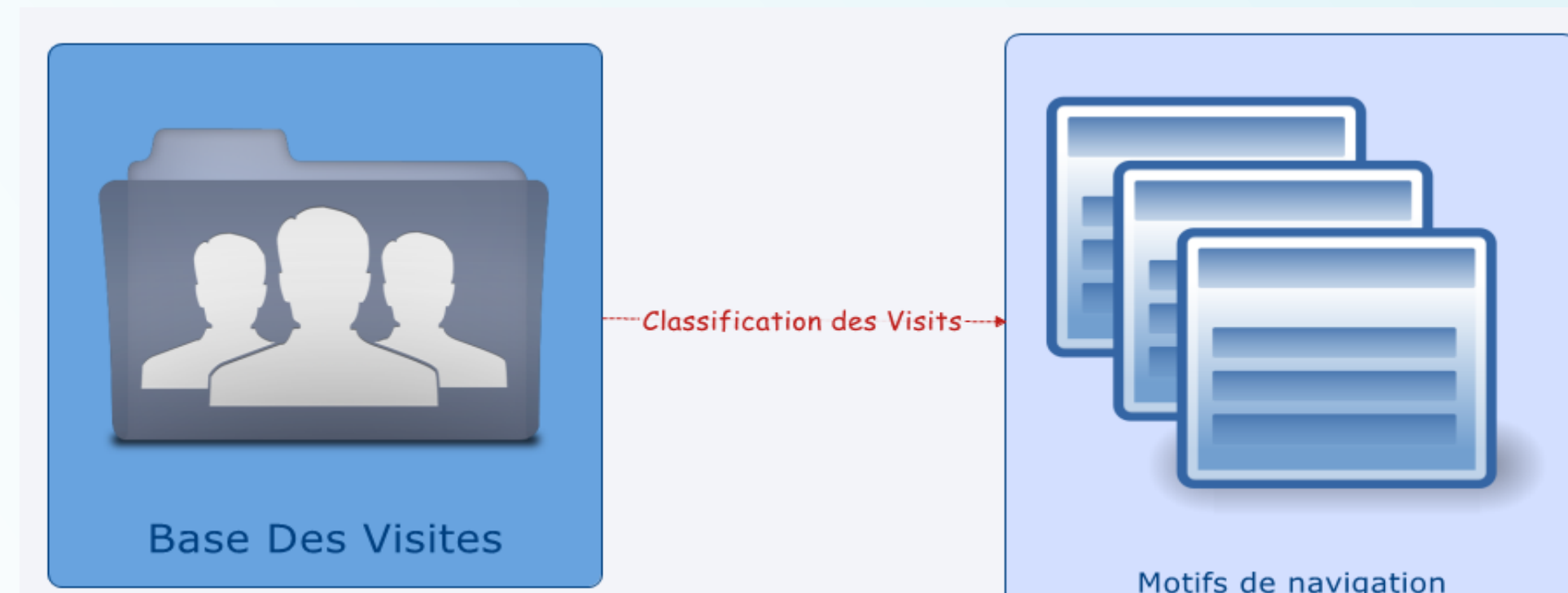


3. Classification des utilisateurs

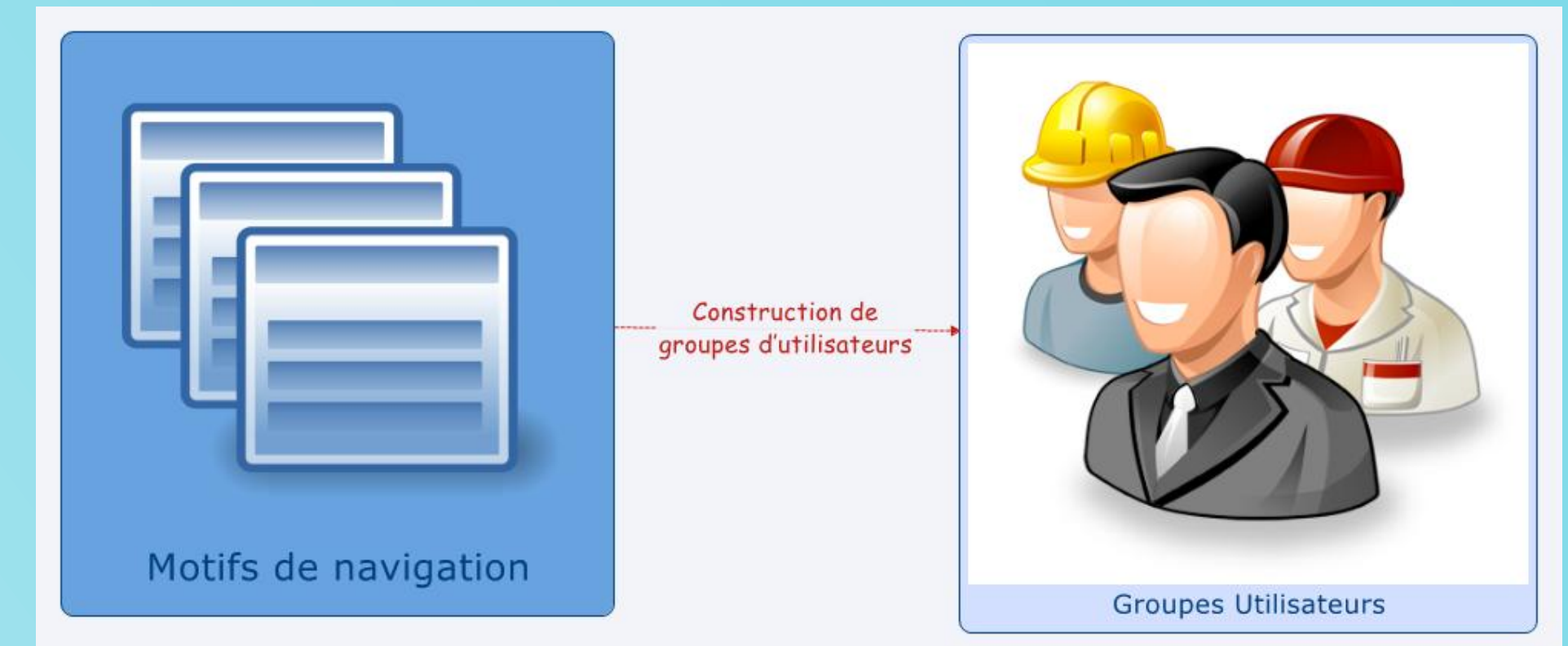
3.1 Classification des requêtes



3.2 Classification des visites



3.3 Construction de groupes d'utilisateurs



4. Analyse, interprétation et validation des résultats

Cette phase permet de sélectionner les motifs intéressants et de les valider. Néanmoins, peu de travaux ont abordé cette étape : citons tout même les travaux sur des outils de visualisation mais aussi ceux abordant la notion d'intérêt (ou d'utilité) des motifs extraits dans le contexte des systèmes Web d'aide personnalisée et dans le contexte de l'aide à la conception de sites Web.

5. Les outils utilisés dans notre approche

- ✓ **Les Algorithmes de classification :**
 - K-Means .
 - Fc-Means .
- ✓ **Analyse des Correspondances Multiples**
- ✓ **Analyse en Composantes Principales (ACP)**
- ✓ **Cartes de Kohonen**

Conclusion

Dans ce travail, nous avons présenté les notions de base sur le WUM (Web Usage Mining) ainsi que la méthodologie des étapes nécessaires. Notre travail, devra certainement permettre au webmaster de mieux comprendre le comportement de ses internautes afin de diagnostiquer et améliorer son site Web.

References

- [1] Charrad, M. (2005). *Techniques d'extraction des connaissances appliquées aux données du Web*. Mémoire de Mastère présenté en vue de l'obtention du diplôme de Mastère en Informatique, Ecole Nationale des Sciences de l'Informatique de Tunis, Laboratoire RIADI.
- [2] Cooley, R., B. Mobasher, et J. Srivastava (1999). *Data Preparation for Mining World Wide Web Browsing Patterns*. Journal of Knowledge and Information Systems.
- [3] Mostafa Hanoune, Faouzia Benabbou (2006). *Traitement et exploration du fichier Log du serveur web, pour l'extraction des connaissances : Web usage mining*. Afrique SCIENCE 02(3) (2006) 335 – 351.
- [4] Cheikh Ba, Philippe Reitz. *Implémentation d'une capacité d'apprentissage à un centre d'information sur le secteur des pêches en guinée*. Diplôme d'Etudes Approfondies « Informatique ». Université de Montpellier II