

# Traitement et exploration du fichier Log du Serveur Web, pour l'extraction des connaissances : Web Usage Mining

Mostafa Hanoune\*, Faouzia Benabbou\*

\* Université Hassan II- Mohammedia, Faculté des sciences Ben M'Sik, Laboratoire TIM (Technologies de l'information et Modélisation), Casablanca, Maroc  
m\_hanoune@yahoo.fr, hgfbenabbou@menara.ma.

**Résumé.** Le but dans ce travail consiste à concevoir et réaliser un Outil Logiciel, en utilisant les concepts du Web Usage Mining pour offrir aux web masters l'ensemble des connaissances, y inclut les statistiques sur leurs sites, afin de prendre les décisions adéquates. Il s'agit en fait, d'extraire de l'information à partir du fichier log du serveur Web, hébergeant le site Web, et de prendre les décisions pour découvrir les habitudes des internautes, et de répondre à leurs besoins en adaptant le contenu, la forme et l'agencement des pages web.

## 1 Introduction

L'activité sur le Web et les données résultantes ont connu une croissance très rapide, vu la croissance exponentielle du nombre des documents mis en ligne.

D'après des statistiques sur des sites spécialisés, le nombre des utilisateurs d'Internet dans le monde a dépassé le milliard (1 022 863 307), au mois de mars 2006<sup>1</sup>, et le nombre de sites Web a atteint 74,4 millions au mois d'Octobre 2005<sup>2</sup>. Ces données, en particulier celles relatives à l'usage du Web, sont traitées dans le Web Usage Mining (WUM). Dans cet article, nous décrivons les fonctionnalités majeures du logiciel que nous avons conçu et réalisé, et qui permet l'analyse des fichiers Logs afin de comprendre le comportement des internautes sur un site Web (Site de l'université Hassan II- Mohammedia [www.univh2m.ac.ma](http://www.univh2m.ac.ma) Casablanca, Maroc).

## 2 Proposition

L'apport de ce travail réside principalement dans les points suivants :

1. Connaissances sur les visiteurs :
  - (a) Le pourcentage des visiteurs par semaine par mois et par an
  - (b) Avoir une visibilité internationale : d'où proviennent nos visiteurs ?
2. Connaissances sur les pages :
  - (a) Les pages les plus et les moins consultées (pages populaires et pages impopulaires)
  - (b) Les combinaisons des pages consultées
  - (c) Savoir quels sont les liens qui nous référencent le mieux
3. Connaissances sur les navigateurs et les OS
  - (a) Le pourcentage des navigateurs les plus utilisés

---

<sup>1</sup> <http://www.internetworldstats.com/stats.htm>

<sup>2</sup> <http://www.netcraft.com>