

# **Repérage automatique d'informations temporelles touristiques dans des pages Web**

Stéphanie Weiser

Les expressions temporelles liées au domaine du tourisme ont des caractéristiques précises : elles ont une visée informative et sont assez spécifiques. Elles le sont d'autant plus lorsqu'elles figurent dans des pages Web. Notre travail consiste à extraire ces informations des pages Web et à les annoter afin d'alimenter la base de connaissance d'un portail touristique sur Internet. Après avoir caractérisé les informations que nous voulons détecter, nous verrons les difficultés qui peuvent se poser, au moment de l'extraction ou de l'annotation.

Sur un plan technique nous présenterons notre méthode d'extraction, et les transducteurs qui ont été réalisés à l'aide d'Unitex pour les besoins du projet.

Sur un plan théorique, nous ferons une étude comparative des expressions temporelles touristiques sur Internet et dans les guides touristiques papier. Nous étudierons également la façon dont le temps est représenté sur le Web : informations sans cesse actualisées, dates relatives, etc.