Etude: comment repérer le spam dans le texte

L'étude a porté sur la sémantique d'un millier de messages spammés différents, en anglais, mais dont l'analyse peut être facilement transposable à la linguistique francophone.

Elle est d'autant plus intéressante qu'Assurance Systems fournir le service *Message Checker* employé dans plus de 900 règles de filtrage du 'spam'. Assurance Systems a déterminé des convergences significatives dans les contenus des emails spammés, dont trois 'déclencheurs' qui qualifient la majorité d'entres eux: **« Click Below/Click Here » – *Cliquez ci-dessous/Cliquez ici*** La présence de ces expressions est une généralité sur les emails spammés. On les retrouve aussi dans des phrases du type « Click here for your great offer » (*Cliquez ici pour bénéficier d'une offre exceptionnelle*) ou « Click on the link below for details » (*Cliquez ci-dessous pour plus de détails*). Par contre, les règles ne détectent pas encore les phrases qui comportent l'expression « Click ». Quant à l'expression « Visit here » (*Visitez ici*), elle peut présenter une alternative intéressante. **Police HTML de couleur bleu** La couleur bleu est historiquement associée au 'spam'. Sur ces emails, la balise HTML de police de caractère *Font* est associée à la balise de couleur *Color*, avec l'attribut de la couleur bleu. Cette pratique est aujourd'hui connue. Son usage fluctue selon des critères encore mal connus, sans doute avec des effets de mode ou de recopie. Elle tend à disparaître. **Les phrases spammées** Certaines phrases ou expressions tendent à se répéter sur les messages spammés. « *free offer* » (*offre gratuite*), « *just for you* » (*que pour vous*) ou « *make $$$* » (*des dollars à gagner*) figurant parmi ces phrases les plus répandues. **D'un bon usage des règle de filtrage** Les règles de filtrage permettent d'identifier des généralités redondantes qui encombrent les messages spammés. C'est par leur intermédiaire que les messages sont scannés afin d'écarter ceux qui peuvent être assimilés à du spam. Bien évidemment, d'autres règles ou techniques viennent compléter ces algorithmes 'gendarmes' de nos messageries. Mais il est bon de connaître ces trois 'déclencheurs' génériques pour éviter de voir disparaître des messages anodins. *Tu n'as pas reçu mon message ?* ?