

De la recherche à la génération de mots clés

[Retour au sommaire de la lettre](#)

Que ce soit en termes de recherche ou de référencement, la plupart des stratégies d'approche sont basées sur la notion de mots clés. Comment trouver les meilleurs mots clés pour garantir la meilleure pertinence des réponses ? Comment identifier les requêtes sur lesquelles tenter un positionnement pour un meilleur référencement ? Une linguistique s'est penchée pour nous sur le sujet et nous donne ses réflexions en même temps qu'une méthodologie d'identification des termes les plus efficaces lorsque l'on désire travailler dans le monde des moteurs de recherche...

Introduction

La pratique quotidienne de l'Internet a rendu la formulation de requêtes sous forme de mots-clés quasi automatique et quotidienne. Il est intéressant de se questionner sur les processus qui sous-tendent cette formulation et ce pour deux raisons. La première concerne la recherche de mots-clés permettant de positionner un site sur un domaine donné. La seconde est la génération de mots-clés automatique qui relève davantage de l'intelligence artificielle tentant de reproduire les processus cognitifs humains.

Il existe en effet un amalgame entre "recherche" et "génération" de mots-clés. Les "générateurs de mots-clés" de Google Adwords ou de MIVA ne génèrent en rien des mots-clés, en cela qu'il ne les crée pas. Il recense simplement sur une période limitée les mots-clés tapés par les internautes selon un mot (concept ?) donné. La recherche de mots-clés que nous avons menée nous a amené à nous questionner sur ces processus cognitifs et linguistiques mis en jeu. Une fois identifiés, il serait intéressant de s'en servir pour générer automatiquement des mots-clés selon une thématique donnée à l'image de ce que peut faire un internaute.

Recherche de mots-clés

Nous avons donc procédé à la recherche de mots-clés selon des thématiques précises (sexe, voyage, immobilier, hôtel, actualité, stars, informatique). Pour effectuer cette recherche nous avons procédé par étapes en y mêlant une réflexion linguistique.

Quand on veut faire une recherche par sujet, la première étape a pour objectif de trouver les mots ou les concepts importants décrivant le sujet.

La première étape réside donc dans l'**identification de concepts proches** (synonymes, proxèmes) de la thématique. Ainsi pour "immobilier", nous avons identifié des mots comme immeuble, maison, appartement, studio, ... ainsi que des concepts plus indirects tels que assurance, architecte, étudiant, particulier, etc.

Une fois ces concepts identifiés, nous avons **divisé la thématique en catégories constituantes**. Pour une thématique comme "immobilier", nous avons mis au jour les paramètres pratiques, la transaction et la localisation géographique. A partir de cette classification, nous avons demandé Google Adwords pour trouver les mots-clés correspondant à chaque catégorie. Par exemple, "achat appartement", "vente appartement" pour la catégorie transaction, "appartement paris" pour la catégorie localisation géographique.

Pour arriver à la catégorisation d'une thématique, il faut combiner à la fois la recherche empirique sur la thématique et le recensement de mots-clés sur Google Adwords. Ainsi pour "immobilier", nous nous sommes rendu sur des sites immobiliers et BTP pour cerner les divers domaines d'intervention et avons croisé les données avec le recensement des mots-clés qui ressortaient sur Google Adwords par rapport à la requête "immobilier". Une fois ces données croisées, les catégories peuvent être établies. Pour chaque catégorie, on trouve des mots génériques mais aussi quelques mots spécifiques tels que "maison en bois", ou "appartement pierre de taille".

La troisième étape réside dans **la combinaison des mots génériques et spécifiques** ressortant dans chaque catégorie. A titre d'exemple, sur la thématique "immobilier", on trouvera :

Catégorie 1 : maison, appartement

Catégorie 2 : particulier, architecte

Catégorie 3 : location, assurance, prix, louer

Catégorie 4 : paris, grenoble

Ce qui donnera : "location maison", "location maison particulier", "assurance maison grenoble", "prix architecte", "prix architecte maison", "location de maison", "prix d'appartement paris", "appartement a louer grenoble particulier", etc.

La combinaison de ces mots génère des expressions de deux, trois, quatre, ... mots qui, par leur spécificité, assurent le principe de longue traîne. Par exemple "achat maison pierre de taille orléans", "construction maison en bois bioclimatique", etc.

Génération de mots-clés

Les requêtes, sous la forme de mots-clés, tapées par les internautes sont la réalisation de stratégies à la fois linguistiques et cognitives. Les mots-clés que nous avons répertoriés pour chacune des sept thématiques répondent à des processus mis en place par les utilisateurs afin d'obtenir en retour les résultats les plus pertinents compte-tenu de leur requête.

Google Adwords, le "générateur de mots-clés" recense une liste de mots-clés mais ne les génère pas comme il le laisse entendre. Il donne l'impression d'un sac de mots sans véritable cohésion ni hiérarchie comme si l'internaute tapait au hasard un mot-clé sans que sa démarche ne soit réfléchie au préalable. Or, le recensement de ces expressions nous a amené à mettre au jour différentes étapes de génération de mots-clés, génération humaine cette fois-ci et non automatique.

La première étape passe par la **conceptualisation de la thématique** et la recherche par les internautes de concepts proches de celle-ci. Il s'agit d'une sorte de mindmapping ou "carte de l'esprit" qui représente la carte sémantique liée au concept fédérateur. Le cerveau fonctionne par visualisation et association. Ainsi, pour une thématique telle que "actualité", les concepts proches seront "nouvelles", "informations", "presse", "news", etc, concepts à partir desquels l'internaute sera amené à nouveau à composer, combinant d'autres mots, lui permettant ainsi d'affiner ses recherches. Cela fera l'objet d'une autre étape. Un outil très simple d'utilisation est mis à disposition par le CNRTL permettant, pour un concept donné, d'afficher sa carte proximique (<http://www.cnrtl.fr/proxemie/actualit%C3%A9>). Un outil un peu plus développé "visuwords" en langue anglaise basé sur les ontologies propose en plus les relations entre les concepts (<http://www.visuwords.com/>).

Au-delà de l'étape de proxémie, pour certaines thématiques, l'utilisateur peut faire appel à des concepts "fantasmés" c'est-à-dire à une représentation stéréotypée du concept fédérateur. En clair, pour la thématique du "voyage", les concepts proches (proxémie) seront ceux de "vacances", "week-end", "séjour", tandis que les concepts stéréotypés ou "fantasmés" seront "soleil", "mer", "plage", "sud". Ainsi l'internaute sera amené à taper des expressions-clés telles que "voyage soleil", "voyage mer", "week-end soleil", etc.

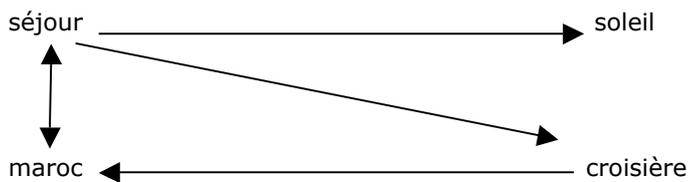
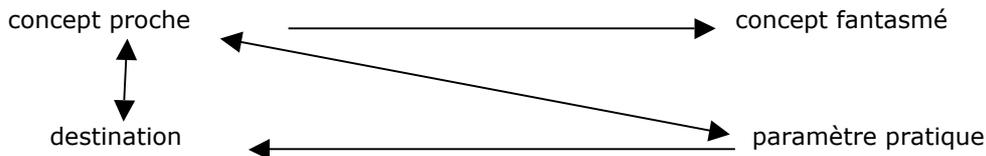
Ces étapes correspondent à une recherche large sans idée précise au préalable de l'internaute. Par ce type de requête, l'internaute se laisse une ouverture. Mais, en terme de positionnement, ces mots-clés sont très concurrentiels.

En revanche, plus la combinaison des mots se complexifie, plus la recherche est ciblée. L'étude des diverses expressions-clés a révélé qu'il existait une récurrence dans les mots employés par les internautes. Nous trouvons par exemple pour la thématique du voyage ces quelques notions "pas cher" (et tous les synonymes), "location" (et sa forme verbale), "réservation" (et sa forme verbale), des destinations récurrentes notamment "maroc", "tunisie", les transports tels que "avion" et tout ce qui s'y rapporte "billet", "vol", "horaire", etc. Ces quelques mots reliés entre eux offrent déjà la possibilité de multiples combinaisons. D'une manière générale, les mots tapés sont des noms plutôt que des verbes ou des adjectifs puisqu'ils sont, d'une part, davantage porteurs de signification et, d'autre part, ils ont davantage de pouvoir de représentation.

Une classification peut être réalisée des différents mots employés par les internautes reflétant les catégories constitutives de la thématique. Cette classification faite, la génération de mots-clés, automatique cette fois-ci, est tout à fait réalisable. Gardons l'exemple de la thématique "voyage", nous avons identifié deux catégories opérantes : destination et paramètres pratiques, les deux

autres catégories correspondant aux concepts proches et fantasmés. Dans les destinations, nous l'avons dit, elles sont récurrentes, soit une vingtaine au total. Il faudra distinguer les pays des villes qui ne s'emploient pas avec le même concept. Ainsi, on aura "voyage maroc" mais "week end paris". En ce qui concerne les paramètres pratiques, on retrouvera tout ce qui est lié à la notion de préparation du voyage tel que "réservation hotel", "billet avion pas cher", "vol low cost", "vacances dernière minute", etc.

Pour résumer, la thématique du "voyage" fait appel à des concepts proches, des concepts fantasmés, les uns étant reliés aux autres dans des expressions-clés ou bien peuvent avoir une existence à part entière. Elle fait également appel à deux catégories qui constituent pour ainsi dire "le corps de la thématique" et qui sont la destination et les paramètres pratiques. On trouvera alors la combinaison des concepts proches et/ou des paramètres pratiques et/ou de la destination telle que "week end pas chers thalasso", "voyage organise au maroc", "croisiere voyage de noce", "croisiere voyage de noce seychelles", etc. Ces quelques règles nous permettent de prédire les mots-clés tapés par les utilisateurs.



A ces étapes s'ajoute la déclinaison des termes de l'expression-clé. A savoir, l'alternance singulier/pluriel, forme verbale/forme nominale avec une prévalence, comme on l'a vu, pour la forme nominale.

La génération de mots-clés passe par quatre étapes :

A partir d'une thématique donnée,

- Identification des concepts proches (on peut étendre cette identification en procédant de même pour les concepts nouvellement trouvés, par exemple "immobilier" a comme concept proche "maison" qui, à son tour, a comme concept proche "pavillon".)
- Eventuellement, identification des concepts stéréotypés ou "fantasmés"
- Identification de catégories opérantes faisant partie intégrante de la thématique (par exemple, "transaction", "paramètres pratiques", ... pour la thématique "immobilier") et des mots génériques "location", "achat", "crédit" et spécifiques "appartement pierre de taille", "appartement duplex meublé» qui les constituent.
- Combinaison des trois étapes.

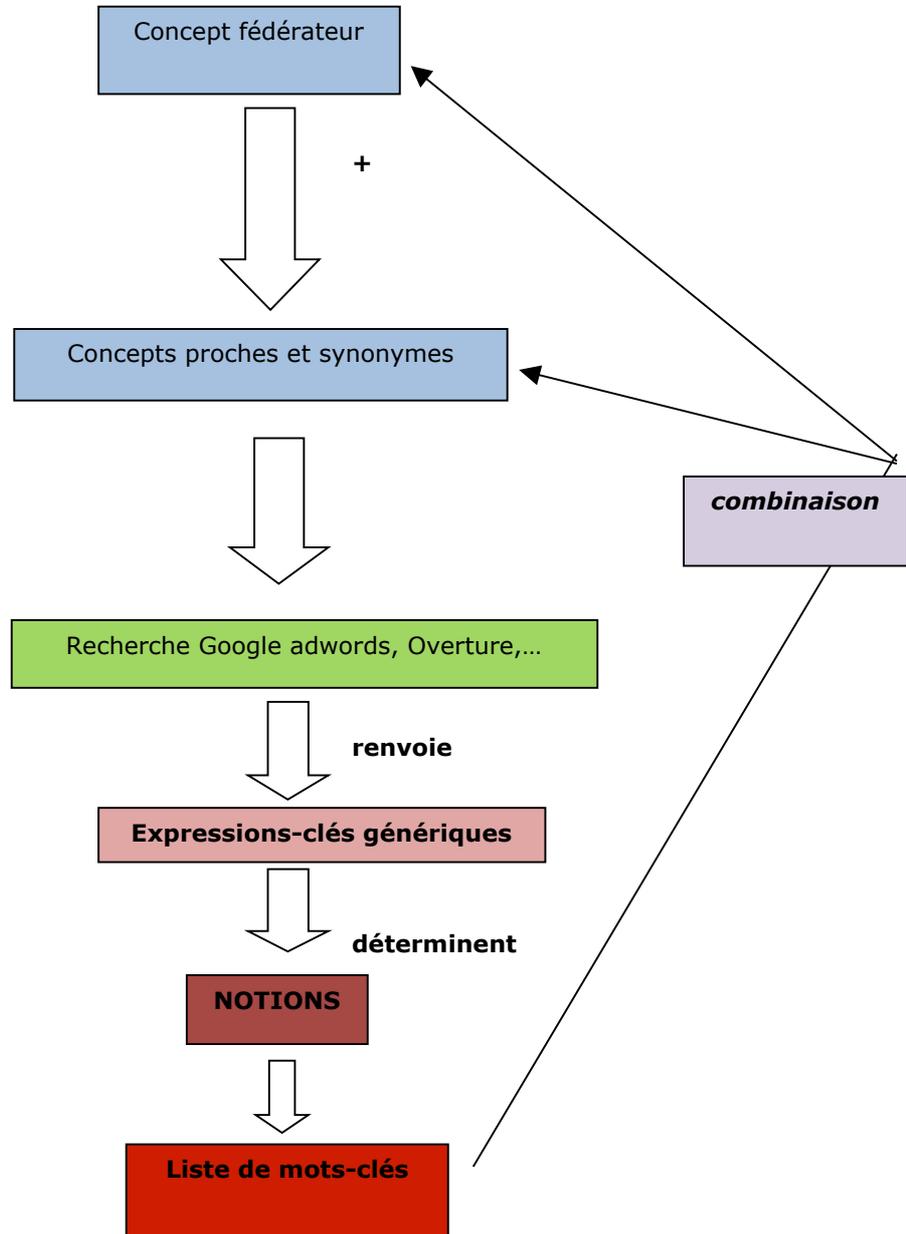
Parallèlement à l'identification et la combinaison des concepts, il y a la déclinaison en genre, nombre et forme morphologique (verbe, adjectif, nom) ainsi que la présence de préposition (à, de, d',...).

Une fois ces classifications réalisées, elles sont opérantes sur le long terme et n'ont pas à être réactualisées dans la mesure où il s'agit d'une thématique générale. Si les trois étapes d'identification sont faites, il n'y a plus qu'à les réutiliser afin de générer des mots-clés par leur simple combinaison.

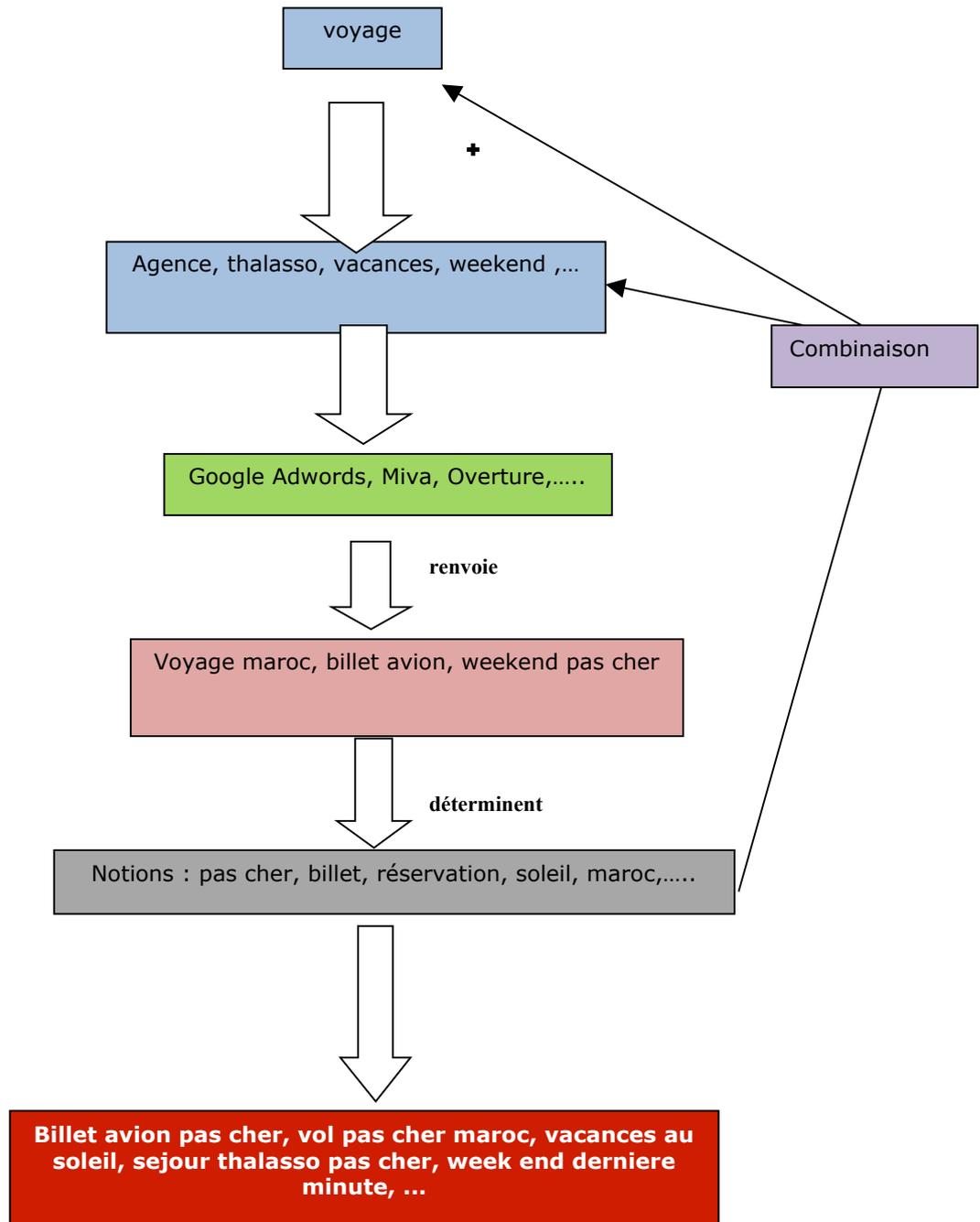
Plus la combinaison est complexe (utilise plus de trois mots) plus elle est ciblée et plus elle assure le principe de longue traîne (exemple : "achat appartement pierre de taille meublé particulier à particulier paris"). De plus, d'une thématique à une autre des catégories opérantes se retrouvent (voyage, hôtel, immobilier, ...) donc les mots appartenant à ces catégories sont limités et réutilisables. Il faut savoir que l'internaute utilise un vocabulaire de faible taille comparativement aux

possibilités qu'offre la langue et que, par là même, les mots utilisés sont récurrents, facilement identifiables et stockables. C'est la combinaison de ces mots qui génère des expressions-clés assurant le principe de la longue traîne.

Nous illustrerons la démarche de recherche exposée ci-dessus par le schéma suivant:



Dont voici un exemple sous forme de schéma:



Myriam MALORON, linguiste informaticienne

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://abonnes.abondance.com/blogpro/2008/06/de-la-recherche-la-gnration-de-mots-cls.html>