Résumé :

Le problème de la recherche d'un motif dans un texte (*pattern matching*) est le problème fondamental de l'algorithmique du texte. Une façon d'aborder ce problème consiste, dans un premier temps, à indexer le texte de façon à réduire le temps de recherche du mot. L'index d'un texte est généralement un automate d'états finis représentants tout les suffixes du texte. La recherche du mot s'effectue alors directement sur l'automate, plutôt que sur le texte.

Dans le cas où le motif est un mot pris dans un alphabet fini, on peut construire un automate des suffixes d'un texte t n'occupant qu'un espace $O(|t|)$. Le calcul de cet index ne prend aussi qu'un temps $O(|t|)$. Muni de l'automate la recherche d'un mot m ne prend qu'un temps proportionnel à la longueur du mot : $O(|m|)$, au lieu d'un temps proportionnel à la longueur du texte.

Dans un travail récent trois nouveaux problèmes de *pattern matching* ont été posés :

- étant donné un mot m et une position i dans un texte t : rechercher m dans le i-ème suffixe de t.
- étant donné un mot m et deux positions i et j ($i< j$) dans un texte t : rechercher m dans le segment $[i,j]$ de t.
- étant donné un motif à joker m (motif pouvant contenir des jokers * remplaçable par n'importe quelle suite de lettres) et un texte t : rechercher m dans t.

Dans ce travail il à été montré qu'on peut répondre aux trois problèmes en temps $O(|m|)$ en utilisant un automate représentant un automate des suffixes de tous les suffixes du texte. Les deux types d'automates proposés (MASDAWG et MASCDAWG) ont une complexité temporelle et spatiale en $O(|t|^2)$. Une telle complexité exclue leur utilisation pour accélérer la recherche dans des textes longs, en particulier dans les génomes dont la longueur se compte en millions et en milliards d'acides nucléiques.

L'objet de ce travail consiste à rechercher une représentation plus efficaces des suffixes de suffixes d'un texte de façon à pouvoir répondre aux trois problèmes énoncés ci-dessus en utilisant un minimum d'espace mémoire et de temps de calcul.