# Types de parsers

L’automatisation du processus de recrutement connait une véritable montée en popularité ces derniers temps. La plupart des acteurs de l’industrie du recrutement ont embrassé avec enthousiasme les technologies d’automatisation quelles que soient les formes qu’elles prenaient, espérant ainsi augmenter leur efficacité et, finalement, de placer plus de gens, aux postes qui leurs conviennent.

Pour trouver les meilleurs candidats, le point de départ de la plupart des recruteurs est dans la recherche de CV.

L’analyse précise des cv est une grande partie du recrutement, puisque les recruteurs ne peuvent trouver les bons candidats uniquement si leurs informations ont été chargées correctement dans leurs bases de données.

## Qu’est-ce qu’un parser, et qu’est-ce que l’analyse ?

Le parsing s’applique généralement au texte, au langage naturel, et à l’action d’analyser et convertir du texte dans un langage compréhensible par l’ordinateur. C’est de la que vient le terme Natural Language Processing, qui fait partie des fonctionnalités « artificiellement intelligente » dans les analyseurs de CV.

Si on se fie à G2 Learning Hub, un analyseur de cv va extraire, stocker et analyser automatiquement les CVs ou les données des candidats d’une telle manière que l’information est ensuite prête à être catégorisée, triée et parcourue lors de recherches.

Le leader de l’industrie des logiciels de recrutement décrit les trois principaux bénéfices d’un analyseur de CV comme ceci :

* Plus rapide
* Une meilleure expérience pour les candidats
* De meilleures données dans le système logiciel de gestion des candidatures (ATS).

## Les trois approches :

En général, il y a trois approches pour analyser un CV ou un métier (une fiche de poste par exemple) pour l’utiliser dans un ATS ou dans un système de gestion des relations avec les candidats (CRM) :

1. Basée sur des mots clés
2. Basée sur la grammaire
3. Basée sur des statistiques

### Basée sur des mots clés :

Un analyseur de CV basé sur les mots clés fonctionne en identifiant des mots, des phrases et des schémas simples dans le texte d’un CV. Ensuite, un algorithme heuristique simple, mis en place pour solutionner le problème d’une manière plus rapide et plus efficace, est appliqué au texte autour de ces mots.

Cet algorithme va par exemple chercher quelque chose qui ressemblera à un code postal puis essaiera d’interpréter les mots qui l’entourent en tant qu’adresse. Il pourrait aussi chercher dans le texte des schémas qui ressemble à des plages de données et « convenir » que le texte aux alentours est un historique des emplois.

Il s’agit à la fois du plus simple des analyseurs, mais aussi du moins précis, avec un taux de précision d’environ 70%.

Ce taux est le moins élevé car un analyseur basé sur des mots clé ne peut extraire les informations qui ne sont pas comprises dans les mots clés enregistrés. Si un mot clé est trop vague, comme le mot « directeur », alors le parser va fréquemment réaliser de fausses déductions à partir de son interprétation, car la phrase ou le mot sera trop « large » pour pouvoir l’identifier avec précision.

### Basée sur la grammaire :

Un analyseur basé sur la grammaire utilise des règles grammaticales pour comprendre le contexte de chaque mot dans le CV.

Il va également combiner les mots et phrases ensembles pour capturer le sens de chaque phrase dans le CV. Ces règles de grammaires, ou ces restrictions de langage forment une hiérarchie de langages formels, connue sous le nom de hiérarchie de Chomsky (d’après Noam Chomsky). Dans les années 50’, ce dernier a exploré les règles de la grammaire dans ses théories linguistiques sur la structure des phrases et les contraintes syntaxiques.

Beaucoup plus complexe qu’un analyseur basé sur des mots clés, un analyseur basé sur la grammaire capture généralement plus de détails. Basés sur la sémantique computationnelle, les analyseurs grammaticaux sont aussi capables de distinguer deux sens différents pour le même mot, ou la même phrase, dans différents contextes.

Avec les analyseurs grammaticaux, il est possible d’obtenir un score de précision supérieur à 90 %.

Néanmoins, un analyseur grammatical requiert un grand nombre d’encodage manuel par des ingénieurs qualifiés. Une quantité substantielle de tests est également requise pour s’assurer que les améliorations dans une « zone » n’impactent pas négativement les performances dans une autre.

### Basée sur des statistiques.

Un analyseur basé sur des statistiques applique des règles de grammaire avec un taux de probabilité, ou trouve l’analyse la plus probable d’une phrase.

Tout comme un analyseur basé sur la grammaire, un analyseur statistique peut différencier différents contextes d’un même mot, ou d’une même phrase, et peut également capturer une large variété de structures comme les adresses ou les lignes temporelles. Il applique des modèles numériques de textes pour identifier les structures dans un CV.

Pour être plus précis, il demandera en entrée un vaste nombre de CV qui seront étiqueté manuellement avec les informations à extraire. Plus on aura de statistiques sur des CV, meilleures seront les données.

Tandis que ses performances sont meilleures qu’un analyseur par mots clés, un analyseur statistique n’est pas aussi précis qu’un analyseur grammatical pour des données sur lesquelles l’analyseur n’a pas été entraîné. Pour améliorer la précision, il devra donc être entrainé préalablement sur les données qu’il est censé traiter.

C’est là que le Machine Learning intervient.

Comme défini dans SAS Insights,

« Le Machine Learning est une méthode d’analyse de données qui automatise la construction de modèles analytiques. C’est une branche de l’Intelligence Artificielle basée sur l’idée que les systèmes peuvent apprendre à partir des données, identifier des schémas, et prendre des décisions avec le minimum d’intervention humaines. »

Inévitablement, plus il y a de données disponibles, plus les « sorties » seront précises.

### Hybrides

Certains analyseurs sont des combinaisons de ces approches, ou plutôt, des hybrides de ces différents types, prenant le meilleur de l’une ou l’autre technologie. Ils peuvent ainsi partir sur une base de précision « juste correcte », et s’améliorer avec le temps.