Etudiants en 3ème année de sciences informatiques Matthieu Delangre, Serhat Gören, Wacim Halit, Horac Horvath, Salim Latoundji, Zakari Sadoune, Yi Zhu, Ismaïl Oudahya. Machine ayant appris le contenu de la base de données de wikipedia auquel on peut poser des questions et qui nous renvoie une réponse correcte et intelligente

# Introduction

L’exploration s’est résumée pendant longtemps à la découverte d’endroits auparavant inconnus, cachés et potentiellement dangereux afin d’en apprendre plus sur le monde qui nous entoure. Aujourd’hui, au 21ème siècle, grâce aux nouvelles technologies et à la digitalisation du savoir humain, l’exploration prend une toute nouvelle dimension. Ce sont maintenant les machines qui explorent un réseau de serveurs contenant quasi toutes les informations connus par l’homme afin de servir son utilisateur de l’information que celui-ci veut obtenir. Dans notre travail cette exploration prend la forme de multiples recherches pour répondre à une question qu’un utilisateur pose à une machine dotée d’un système de question answering. Pour arriver à une réponse cohérente la machine effectue plusieurs opérations que nous allons vous expliquer en quelques mots.

## Le *machine learning*

Pour fournir à l’utilisateur une réponse correcte, la machine doit avoir accès à une grande base de données dans laquelle elle peut chercher la réponse à la question. Dans un premier temps, la machine transforme le langage naturel dans lequel l’utilisateur lui pose la question afin d’améliorer les résultats de la recherche. La transformation du langage revient à la reformulation de la phrase en priorisant les mots-clés importants et en effaçant potentiellement les mots-clés inutiles. La machine va ensuite chercher les mots-clés correspondants ou similaires dans la base de données qu’on lui a fourni au début. Si la réponse à notre question se trouve dans cette base de données, la machine va la trouver et l’extraire de cette dernière. Après cette extraction, afin de fournir la réponse la plus courte mais la plus compréhensible possible à l’utilisateur, la machine va transformer le contenu de ce qu’il a extrait en ce qu’il peut de plus proche au langage naturel. Une fois que la machine à le meilleur résultat possible, elle va afficher celui-ci à l’écran de l’utilisateur ou va le dire à voix haute selon ce qu’on lui demande.

Ce processus paraît simple et rapide, mais la machine effectue énormément d’instructions à l’arrière-plan, car souvent la base de données dans laquelle elle va chercher les informations contiennent des millions de pages dont la plupart n’ont rien à avoir avec celle dont l’utilisateur a besoin, mais que la machine doit quand même analyser. Une fois que la page contenant l’information est trouvée, la machine doit par lui-même sélectionner les paragraphes pertinents et les nettoyer de sorte que seules les informations utiles soient gardées. Une fois ces informations obtenues, elle doit formuler une phrase avec pour fournir une réponse intelligente à l’utilisateur.

De nos jours, il n’est pas difficile de créer une machine ainsi. Il suffit d’avoir un ordinateur, un accès internet et on peut trouver toutes sortes de base de données complètes et d’algorithmes effectuant les différentes tâches. La seule chose que l’utilisateur doit créer lui-même est une interface (site web, application mobile…) où il peut claquer la machine et c’est ciao.