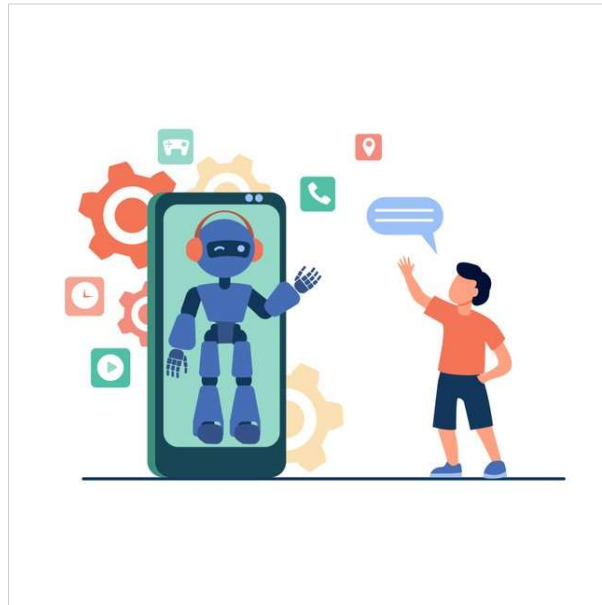


SYSTÈME DE RÉSERVATION INTERACTIF VIA CHAT-BOT



Projet de bachelor présenté par

Fabio RODRIGUES DOS SANTOS

**Informatique et systèmes de communication avec orientation en
Informatique Logicielle**

Mars, 2024

Professeur-e HES responsable

Niklaus EGGENBERG

Légende et source de l'illustration de couverture :

Illustration d'un utilisateur conversant avec un Chatbot depuis son téléphone.

Source : <https://shopifieddna.com/wp-content/uploads/2021/07/chatBot.jpeg>

TABLE DES MATIÈRES

Remerciements	1
Résumé.....	2
Liste des acronymes	3
Liste des illustrations.....	4
Liste des tableaux	6
Introduction.....	7
Chapitre 1 : Définition de Chatbot	10
1.1. Un Chatbot, qu'est-ce ?.....	10
1.2. Trois composants d'un Chatbot	11
a) L'interface.....	11
b) Le moteur de traitement de données	11
c) Le moteur de réponses	11
1.3. trois types de Chatbots.....	12
a) Les Chatbots à Menus	12
b) Les Chatbots par Tokens.....	15
c) Les Chatbots "intelligents"	17
Chapitre 2 : Étude des solutions existantes.....	22
2.1. Agenda.ch	22
2.2. Klara.ch.....	23
2.3. Meetme.io	24
2.4. Comparaison à ce projet	25
Chapitre 3 : Détail technologique.....	26
3.1. Natural Language Processing	26
a) Natural Language Understanding.....	31
b) Natural Language Generation	32
3.2. TAPAS.....	35
a) Transformers	36
b) BERT	37
c) TAPAS : Fonctionnement	40
Chapitre 4 : Étalage des technologies existantes	45
4.1. Chatbots par NLU	45
a) Les intents	45
b) Les Entités	46
c) Les Stories	46
d) Les Actions	47
e) Les Règles	47
f) Les Slots.....	47
g) Les Lookup Tables.....	48

h)	Les Forms	48
i)	Les Pipelines	48
4.2.	Rasa	49
a)	Rasa X.....	49
b)	Rasa Open Source	52
c)	Rasa Pro	54
d)	Duckling	56
4.3.	Chatterbot	58
4.4.	Microsoft Bot Framework.....	58
4.5.	NLTK	59
4.6.	OpenNLP	59
Chapitre 5 : Présentation des prototypes.....		61
5.1.	Prototypes NLP	61
a)	Java	61
b)	Python.....	62
5.2.	Prototype TAPAS	64
a)	Données de test.....	65
b)	Résultats	66
c)	Conclusion	66
5.3.	Prototype Chatbot	68
a)	Données d'entraînement	69
b)	Exemple de conversation.....	71
c)	Conclusion	71
5.4.	Réflexion sur l'ensemble des Prototypes	72
Chapitre 6 : Conception du projet.....		74
6.1.	Technologies présentes	74
a)	Docker	74
b)	FastAPI	75
c)	Express JS.....	75
d)	Socket.IO	76
e)	Vite + React.....	76
f)	Rasa	77
g)	Duckling	77
h)	Postgresql	77
6.2.	Architecture Globale	78
6.3.	Déroulement de réalisation du projet	79
6.4.	Chatbot Rasa	81
6.5.	API	83
6.6.	Base de données	88
6.7.	Interfaces à disposition	90
a)	Telegram	90
b)	Application Web	91
Chapitre 7 : Réalisation du projet.....		93

7.1.	Implémentation Chatbot	93
a)	Stories	93
b)	Forms	96
c)	Rules.....	96
d)	Action Server	97
7.2.	Implémentation API	99
7.3.	Implémentation Interfaces	99
Conclusion		100
Références documentaires		102

Dédié à mon père et ma mère qui ont tout fait pour m’aider à en arriver là et de continuer
à persévérer malgré tout.

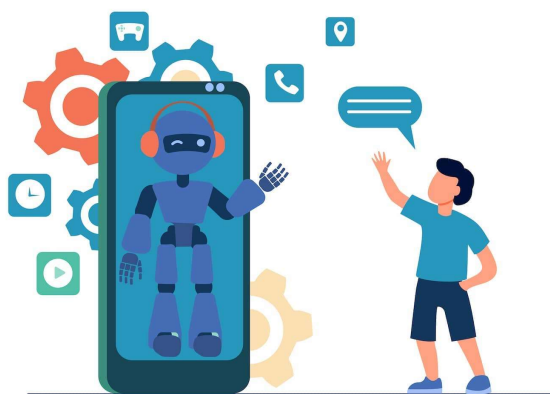
REMERCIEMENTS

Je remercie avant toute chose mon professeur attiré, M. Niklaus Eggenberg, m'ayant suivi tout au long de la réalisation de ce projet et mes camarades avec qui j'ai fait face à maintes épreuves tout au long de ces trois années.

Finalement, je tiens à remercier mes proches pour tout le soutien apporté avant et pendant la rédaction de ce mémoire.

RÉSUMÉ

La majorité des moyens de réservation de nos jours sont soit au travers de formulaires sur une page web, d'appels téléphoniques ou de prendre rendez-vous en personne. Il existe parfois la possibilité de réserver par téléphone mais ces cas sont rares et nécessitent l'intervention d'un opérateur humain. C'est donc dans ce cadre-ci que le projet sur la réalisation d'un Chatbot se place afin d'éventuellement être disponible sur divers canaux de communications tels que Telegram ou What's app. Le but premier étant de fournir un nouvel outil qui permettrait de donner un rendez-vous aux clients ne souhaitant pas passer par des pages web mais ne voulant pas non plus passer des appels téléphoniques. Or, afin de mener à bien ce projet, il a été dans un premier temps nécessaire de comprendre quels composants sont nécessaires au fonctionnement d'un Chatbot et de par ce modèle, quels types de Chatbots existent et surtout comment sont-ils implémentés. Un concept important étant le langage naturel et comment il est traité dans un contexte d'automatisation d'un opérateur conversationnel afin que ce dernier soit capable de comprendre ce qu'il lui est dit et ainsi de pouvoir répondre correctement à sa requête. Afin de trouver la technologie la plus pertinente à la réalisation de ce Chatbot, une multitude de prototypes concernant divers aspects du bot ont été réalisés et étudiés afin de finalement pouvoir, de par l'ensemble des éléments étudiés dans ce rapport, pouvoir prochainement conceptualiser et planifier la réalisation du Chatbot qui sera créé dans le cadre du projet de Bachelor associé.



Candidat-e :

RODRIGUES DOS SANTOS FABIO

Filière d'études : ISC

Professeur-e(s) responsable(s) :

EGGENBERG NIKLAUS

Travail de semestre soumis à une convention de stage en entreprise : non

Travail soumis à un contrat de confidentialité : non

LISTE DES ACRONYMES

NLP Natural Language Processing

NLU Natural Language Understanding

NLG Natural Language Generation

I.A. Intelligence Artificielle

UX User Experience

POS Part-of-Speech

NER Named Entity Recognition

LLM Large Language Model

TAPAS TAbLe PArSing

SDK Software Development Kit

CALM Conversational AI with Language

API Application Programming Interface

LUIS Language Understanding Intelligent Service

RNN Recurrent Neural Network

LISTE DES ILLUSTRATIONS

Illustration 1: Graphique de la relation entre les trois composants de base.....	10
Illustration 2: Graphique d'un Chatbot à menus	12
Illustration 3: Arbre de décision d'un Chatbot à menus	14
Illustration 4: Graphique d'un Chatbot de règles	15
Illustration 5: Exemple de sélection de règle pour Chatbot par tokens	16
Illustration 6: Graphique d'un Chatbot Intelligent	18
Illustration 7: Pyramide de la difficulté technologique et UX selon le type de Chatbot	21
Illustration 8: Séquence de traitement NLP	27
Illustration 9: Représentation de la relation NLP, NLU et NLG.....	31
Illustration 10: Cas exemples du moteur google utilisant BERT.....	37
Illustration 11: Représentation de l'entrée du modèle BERT	39
Illustration 12: Entrée du modèle TAPAS.....	40
Illustration 13: Schéma de fonctionnement du modèle TAPAS.....	42
Illustration 14: Interface contenant le menu de Rasa X.....	49
Illustration 15: Page de gestion des modèles avec Rasa X	51
Illustration 16: Exemple de conversation parmi l'historique des conversations	51
Illustration 17: Architecture Rasa	53
Illustration 18 : Illustration de comparaison entre une Machine Virtuelle et des conteneurs Docker.....	75
Illustration 19: Diagramme d'architecture général du projet.....	78
Illustration 20 : Diagramme en arbre de la structure Docker Compose du projet	79
Illustration 21 : Liste de labels sur le GIT du projet	80
Illustration 22 : Capture d'écran d'un issue board sur Gitlab	80
Illustration 23 : Diagramme représentant une interaction quelconque entre l'utilisateur et le bot	81
Illustration 24 : Diagramme d'un échange portant sur la demande de ressource	82

Références des URL

URL01	https://www.ovationcxm.com/blog/3-kinds-chatbots-youll-meet
URL02	https://www.javatpoint.com/nlp
URL03	https://datasolut.com/natural-language-processing-vs-nlu-vs-nlg-unterschiede-funktionen-und-beispiele/
URL04	https://www.youtube.com/watch?v=ioGry-89gqE
URL05	https://humboldt-wi.github.io/blog/research/information_systems_1920/bert_blog_post

URL06	http://arxiv.org/abs/2004.02349
URL07	https://www.datocms-assets.com/30881/1608730961-screenshot-2019-05-21-at-12-46-37.png
URL08	https://www.datocms-assets.com/30881/1608730956-screenshot-2019-05-21-at-12-44-11.png
URL09	https://www.datocms-assets.com/30881/1608730966-screenshot-2019-05-21-at-12-48-34.png
URL10	https://rasa.com/docs/rasa/arch-overview
URL11	https://www.netapp.com/blog/containers-vs-vms/

LISTE DES TABLEAUX

Tableau 1: Représentation du tableau fourni en entrée de TAPAS	41
Tableau 2: Tableau de dimensions de Duckling.....	58
Tableau 3: Extrait de Terrains.csv	65
Tableau 4: Contenu de Horaires.csv.....	65
Tableau 5 : Tableau des routes disponibles sur l'API	87

Références des URL

URL01	https://github.com/facebook/duckling
-------	---

INTRODUCTION

De nos jours et comme il a été le cas par le passé, que ce soit pour aller au médecin ou au restaurant ; il est commun de placer des réservations. Or pour ce faire il est d'usage de passer un coup de fil ou dans d'autres cas de placer la réservation en se rendant au dit lieu du rendez-vous. Aujourd'hui, il est possible de placer ses réservations directement depuis de nombreux sites web. Qu'ils soient possédés par le particulier chez qui on souhaite placer rendez-vous ou par une autre entité chargée de gérer l'aspect de placement de réservations et/ou rendez-vous. Ces sites web permettant de prendre rendez-vous se présentent très régulièrement sous la forme de boutons cliquables et de quelques champs à remplir telle que la date du rendez-vous, le nom/prénom et autres informations. Cependant, il est bien plus rare qu'une alternative à cela existe et ne nécessitant pas de devoir se rendre sur une page internet quelconque. Le projet qui sera réalisé a pour optique d'apporter un autre moyen de placer une réservation se trouvant à l'intersection d'un appel téléphonique et un simple formulaire cliquable : Un Chatbot servant à prendre des rendez-vous depuis des applications de messageries telles que What's app ou Telegram ou alors pouvant être aisément inclus dans un site web existant et qui permettra de rendre plus fluide le processus de prise de rendez-vous.

Ce Chatbot s'apparente plus au premier abord à un composant applicatif auquel il est possible de fournir du texte et en recevoir en sortie. Bien que le but premier soit de fournir un Chatbot avec lequel il sera possible de converser sur une application de messagerie, il est tout à fait possible de mettre en place une section Chatbot sur un site quelconque et d'y lier le Chatbot, le rendant de ce fait facile à mettre en place sans avoir à se soucier de l'interface graphique de l'unique nécessité du Chatbot étant un champ textuel mis à disposition à l'utilisateur.

Une des raisons primaires de ma décision de réaliser ce projet étant qu'à présent, en 2024, de nombreux progrès ont été réalisés dans le secteur du Machine Learning et de

l’I.A. et bien que de nombreuses compagnies aient commencées à doucement mais sûrement à implémenter des technologies diverses en termes d’I.A. ou Machine Learning. Il y a encore un grand nombre de secteurs dans lesquels la présence de ces dernières se fait encore discrète ou simplement car peu ont considéré faire usage de ces technologies dans ces derniers. Le secteur en question présentement est celui du Service Client et plus spécifiquement l’assistance à réservation et prise de rendez-vous. Je trouve tout particulièrement pertinent l’exploration de cette voie qu’est l’usage de ces technologies pour la réalisation d’un Chatbot pour prendre rendez-vous car moi-même devant de temps à autre prendre rendez-vous, il est bien pénible pour bon nombre de gens de passer des coups de fil, moi inclut. C’est pour cela que fournir une solution qui soit modulaire et ne nécessitant pas d’architecture particulière que simplement un moyen d’incorporer une boîte de dialogue ou champ textuel quel que soit la plateforme choisie faciliterait grandement son utilisation par les particuliers de tout secteurs et qui fournirait une alternative viable aux appels et sites web.

Ce rapport précède le travail de Bachelor en Informatique et Systèmes de Communications à L’Haute Ecole du Paysage, d’Ingénierie et d’Architecture de Genève et a pour objectif d’explicitier les divers concepts prééminents dans le projet, l’étude des divers technologies existantes pouvant servir à la réalisation de ce projet, l’étude de divers produits existant similaire et de finalement montrer des exemples concrets de leur utilisation à l’aide de quelques prototypes.

Actuellement, la réalisation de ce rapport s’est effectuée de manière régulière chaque semaine lorsque le temps me le permettait et un état des lieu fut régulièrement effectué à l’aide d’entretiens hebdomadaires avec M. Niklaus Eggenberg. Afin de réaliser ce rapport, l’outil Git a servi à versionner et stocker les divers prototypes réalisés pour tester les technologies ainsi que le journal de bord tenu à jour régulièrement pour faire compte du travail effectué à chaque session de travail. D’ailleurs une copie du rapport a été régulièrement mise à jour sur Google Drive. Le stockage s’est effectué sur la plateforme

Gitlab mise à disposition par la HES et sur un projet GitEdu¹ possédé par M. Niklaus Eggenberg. L'outil Google Docs fut d'une grande aide lors de l'écriture d'ébauches, de prises de notes, d'explications préemptives des divers technologies recherchées et pour la formulation de mes idées concernant le travail à réaliser par la suite.

La principale méthode de recherche de sources fut grâce au moteur de recherche Google et occasionnellement Google Scholar. Une grande partie de mes sources sont divers sites de blogs et articles tels que certains venant de Medium.com, IBM ou encore Analytics Vidhya expliquant certains concepts détaillés tout au long de ce rapport. Néanmoins une certaine quantité d'articles de recherche notamment celui du modèle TAPAS par des ingénieurs de Google qui ont servi à comprendre certaines technologies.

Ce rapport va par conséquent servir de fondation à la réalisation du Chatbot lors du projet de Bachelor à venir. Concernant la structure du présent rapport, elle se compose de cinq chapitres. Le premier définit de manière concrète et conceptuelle ce qu'est un Chatbot. Le deuxième porte sur l'étude des solutions existantes de réservation et prises de rendez-vous se trouvant sur le marché et comment elles comparent à celle proposée ici. Le troisième est une énonciation et explicitation des divers concepts piliers composant les divers structures de Chatbot énoncées au chapitre précédant. Le quatrième a pour but de faire un état des nombreuses technologies existantes et utiles pour la réalisation de ce projet et finalement, le cinquième chapitre me sert à présenter les quelques prototypes que j'ai pu réaliser de certaines des technologies énoncées précédemment ainsi que mes réflexions les concernant.

¹ [HTTPS://GITEDU.HESGE.CH/NIKLAUS.EGGENBER/2324_RODRIGUESDOSANTOS](https://gitedu.hesge.ch/niklaus.eggenber/2324_RodriguesDosSantos)

Chapitre 1 : DÉFINITION DE CHATBOT

Avant même de commencer à parler technologies, il est important d'explicitement ce qu'est un Chatbot.

1.1. UN CHATBOT, QU'EST-CE ?

Les Chatbots sont des applications disponibles sous une multitude de formes et plateformes et ayant pour but premier d'engager dans une conversation avec un utilisateur afin de tenter au mieux de répondre à quelconque requête que ce soit. Les Chatbots font affaire à un grand nombre de requêtes variées comme fournir des renseignements, fournir un service ne nécessitant pas d'opérateur humain, faire office de service client afin de tenter de résoudre tout éventuel problème dans la limite des capacités du Chatbot (le cas échéant, le bot peut rediriger l'utilisateur vers un opérateur humain), récolter du feedback, et bien d'autres usages encore. Bien que des Chatbots existent dans un nombre incalculable de formes et structures, il y a néanmoins une manière de décomposer n'importe quel bot en trois éléments distincts

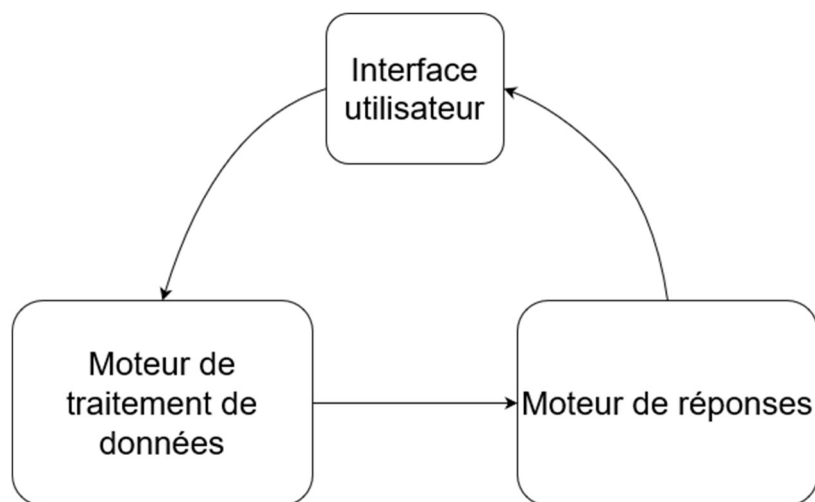


Illustration 1: Graphique de la relation entre les trois composants de base

Source : Rodrigues dos Santos Fabio

1.2. TROIS COMPOSANTS D’UN CHATBOT

a) L’INTERFACE

Aucun Chatbot ne pourrait avoir la dénomination de *Chatbot* s’il ne possédait pas d’interface avec laquelle l’utilisateur peut interagir. Au plus simple, une interface pourrait être une simple ligne de commande de terminal et au plus poussé, une fenêtre de dialogue ayant l’apparence d’un logiciel de messagerie comme sur nos smartphones. L’essentiel ici est que l’interface mise à disposition comporte au moins une zone de saisie textuelle ou autre support de communication tel que des boutons cliquables par exemple. C’est au travers de ce support que l’utilisateur fera parvenir ses diverses requêtes au bot pour traitement ultérieur.

b) LE MOTEUR DE TRAITEMENT DE DONNÉES

Une fois que l’utilisateur a exprimé sa demande au travers d’une des interfaces disponibles, cette demande sous forme de texte va être immédiatement acheminée au moteur de traitement de donnée. Les requêtes utilisateurs dans le cas où le canal de communication est un champs de texte se présentent sous la forme de suites de caractères. Ce qui est tout à fait compréhensible pour un humain mais pas pour une machine, c’est donc pourquoi un traitement se doit d’être appliqué au texte utilisateur. Le texte se présentant sous formes de phrases sera verra le plus souvent mis sous la forme de mots-clés ou instructions pouvant eux être compris par la machine.

c) LE MOTEUR DE RÉPONSES

Une des dernières étapes du cycle d’échange entre l’utilisateur et le bot est le passage des données précédemment traitées dans le moteur de réponse. Son rôle est de faire sens des divers mots-clés et/ou instructions qu’il a reçues et d’envoyer à l’utilisateur la réponse correspondant le plus possible à sa requête.

Ces trois composants, une fois mis ensemble forment un Chatbot fonctionnel capable de recevoir des requêtes d'utilisateurs, de les rendre compréhensibles par le système et d'en sortir la réponse adéquate à l'utilisateur et ce ainsi de suite jusqu'à ce que l'échange prenne fin.

1.3. TROIS TYPES DE CHATBOTS

Le simple cycle d'échange de données expliqué ci-dessus permet à présent de diviser cette fois les Chatbots en trois catégories distinctes définissant les variantes de Chatbots les plus communes.

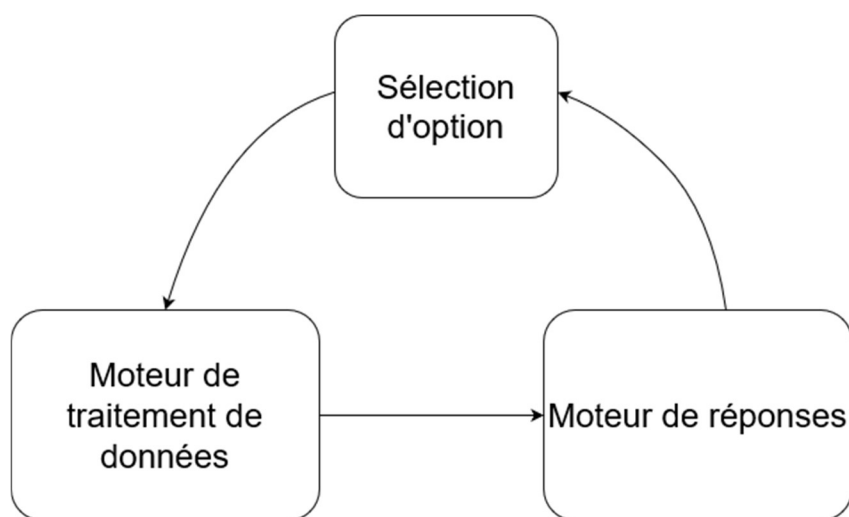


Illustration 2: Graphique d'un Chatbot à menus

Source : Rodrigues dos Santos Fabio

a) LES CHATBOTS À MENUS

Les Chatbots les plus basiques sont ceux fonctionnant à l'aide d'un menu. L'utilisateur discute avec le bot en sélectionnant, soit par un clic ou par une entrée texte, une option.

Les Chatbots à Menus, bien qu'étant factuellement très simplistes ne sont pour autant pas mis de côté du fait de leur aspect simpliste et parfois préféré par rapport à d'autres types de Chatbots. En effet, un Chatbot à menus permet d'éviter tout soucis qui

pourraient advenir lors de l'implémentation des autres types de Chatbots plus bas. Grâce à leur canal de communication on ne peut plus facile à utiliser pour l'utilisateur : Des boutons. L'intégralité de l'échange est réalisée grâce aux pressions successives des multiples boutons s'affichant à l'écran de l'utilisateur et ne nécessitant aucune autre forme d'interaction de ce dernier.

Ce type de bots est notamment très prisé dans des scénarios où les requêtes pouvant être effectuées par l'utilisateur sont prédéfinies. Par exemple, pour un système de commande de plats en ligne ou un système de support téléphonique qui lui, va imiter un Chatbot en demandant à l'utilisateur de presser des boutons sur leur téléphone pour spécifier quel type d'aide ce dernier nécessite. Dans ces deux exemples, l'usage d'un menu permet de s'affranchir d'un opérateur humain à cette étape de l'interaction et donc de gagner du temps. Le résultat de l'interaction avec le client va communiquer des informations à l'opérateur humain avant même qu'il n'ait à demander quel type de soucis le client rencontre dans le cadre d'un support téléphonique ou encore lors de la prise de commande de ce dernier. Actuellement, ce type de Chatbot est l'un des plus répandus dans la majorité des systèmes de réservations.

La séquence de pression de boutons peut être représentée sous la forme d'un arbre de décision. Comme le moyen d'interaction usuel de ce bot se représente sous la forme de boutons cliquables, le système impose un cadre prédéfini ainsi qu'un nombre limité d'interactions possibles et prohibe toute sortie de ce cadre. En effet, à chaque pression de bouton, l'utilisateur prends un des nombreux chemins existants dans cet arbre de décision jusqu'à arriver (ou non) à une réponse convenable et renvoyé à l'utilisateur au travers du moteur de réponses.

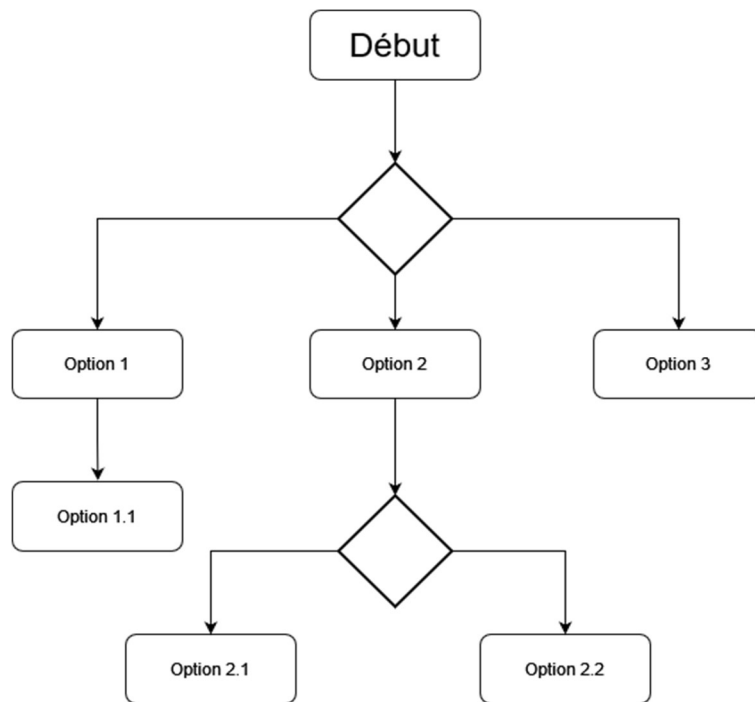


Illustration 3: Arbre de décision d'un Chatbot à menus

Source : Rodrigues dos Santos Fabio

Un des inconvénients possibles de ce Bot est en termes de scalabilité. Plus l'on souhaite étendre le panel de décisions possibles plus il faudra adapter l'affichage en conséquence et entre autres, un panel trop large pourrait être un désavantage plus qu'un avantage car bien que donner le plus de choix possible à l'utilisateur semble être optimal, en avoir trop et donc obliger l'utilisateur à cliquer sur une trop grande quantité de boutons peut ruiner l'expérience utilisateur.

Un autre inconvénient possible est que bien que dans certains cas, il est davantage souhaité d’avoir un cadre fixe, il y en a d’autant plus d’autres où l’utilisateur souhaite effectuer une requête qui ne correspond à aucune des options disponibles. Ce qui obligerait à créer une très grande quantité de variations dans les possibles choix que l’utilisateur peut prendre et qui n’arrangerait pas plus le problème. Au contraire, ça l’aggraverait.

b) LES CHATBOTS PAR TOKENS

Se trouvant un cran au-dessus niveau complexité, se trouvent les Chatbots par tokens. Ces bots ajoutent un degré de liberté qu’il n’est pas aussi aisément trouvable avec un Chatbot à menus grâce à l’utilisation de règles.

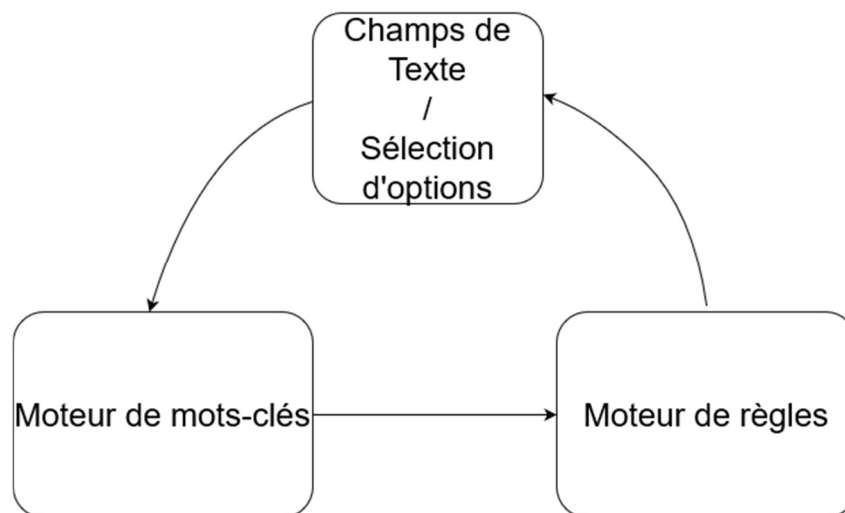


Illustration 4: Graphique d'un Chatbot de règles

Source : Rodrigues dos Santos Fabio

Ce type de bot est presque aussi populaire que ceux à base de menus car tout comme eux, ils répondent à des demandes similaires. Les interfaces souvent rencontrées avec ce genre de Chatbots sont soit des boutons ou cette fois-ci des champs de texte. Il est important de noter la distinction entre les Chatbots à Menus et par Tokens, car l’un possède un chemin prédéfini avec des options se succédant, alors que l’autre possède un ensemble de règles qui selon les attributs ressortis de la requête utilisateur, la règle la plus adéquate sera sélectionnée.

Un exemple pour le cas avec boutons serait qu’au lieu qu’un bouton ne fasse simplement avancer dans l’arbre de décision, que chaque bouton soit pondéré et que même avec une succession de pressions de boutons différentes on puisse arriver à la même réponse.

Dans le cas où l’entrée utilisateur est un texte, la méthode la plus facile et répandue est le simple passage de mots-clés dits « Tokens ». Selon les tokens ayant été ressortis de la requête, une règle comprenant un certain nombre de tokens correspondant ou tous se verra sélectionnée. Ci-dessous, on observe bien le fait que la règle n°3 a été choisie car comportant une succession de mots correspondant à cette règle.

Ceci est une phrase d'exemple, bien qu'elle paraisse importante



Illustration 5: Exemple de sélection de règle pour Chatbot par tokens

Source : Rodrigues dos Santos Fabio

Il existe deux approches possibles à la tokenisation du texte

La première consistant à faire usage d’un dictionnaire de mots-clés établi au préalable. Lorsqu’une entrée utilisateur se voit traitée par le moteur de mots-clés, ce dernier ira simplement vérifier si le token est présent dans le dictionnaire. Si c’est le cas, alors il est ajouté à la suite de tokens qui sera envoyée au moteur de réponse.

La deuxième, elle, consiste à faire usage du NLP (Natural Language Processing) qui permet de faciliter davantage l’implémentation du dictionnaire. Car dans la première approche, une limitation qui pourrait rapidement se faire ressentir est pour commencer le fait que pour chaque token ajoutés dans le dictionnaire, il faudra les associer à des règles et donc plus il y a de token, davantage il faudra créer d’associations règles-tokens. De plus, pour chaque token il peut exister une infinité de variations à cause de fautes d’orthographe

ou tokens similaires en écriture et/ou sens ce qu’une simple comparaison mot à mot ne pourrait pas détecter correctement.

En d’autres termes, le problème sous-jacent à une implémentation si simpliste est celui de la scalabilité de l’arbre de décisions. Pour un petit projet, ce n’est pas un grand soucis mais avec le temps cela peut vite devenir ingérable. C’est donc pour cette raison qu’il est judicieux de faire usage d’un moteur NLP si le projet commence à prendre de l’ampleur pour que la première approche soit viable car ce dernier permet d’éviter de devoir par exemple y inscrire toutes les variations d’un mot grâce à l’usage de diverses techniques de NLP voire de regrouper des familles de mots sous un seul type de token. De ce fait, la taille du dictionnaire final se verra grandement réduite et donc même avec un nombre de règles, l’application reste bien plus maintenable et scalable.

Les inconvénients à prendre en compte venant avec l’utilisation d’un Chatbot de règles est que bien qu’il permette un plus grand degré de liberté avec l’usage de champs textes et par conséquent de tokens qui élargissent le panel de possibilités à disposition de l’utilisateur en termes de requête, on reste tout de même dans un cadre restreint car tous les comportements sortant du moteur de règles doivent être implémentés au préalable et continuellement mis à jour au besoin. Ce qui donne l’illusion à l’utilisateur qu’il peut poser quelconque requête que ce soit mais il se verra vite confronté à une quantité de réponses limitées s’il sort trop du cadre initial par inadvertance.

c) LES CHATBOTS “INTELLIGENTS”

Le dernier type de Chatbot étant à la fois le plus complexe et pouvant apporter le plus de qualité en termes d’expérience utilisateur sont les Chatbots dit « Intelligents ».

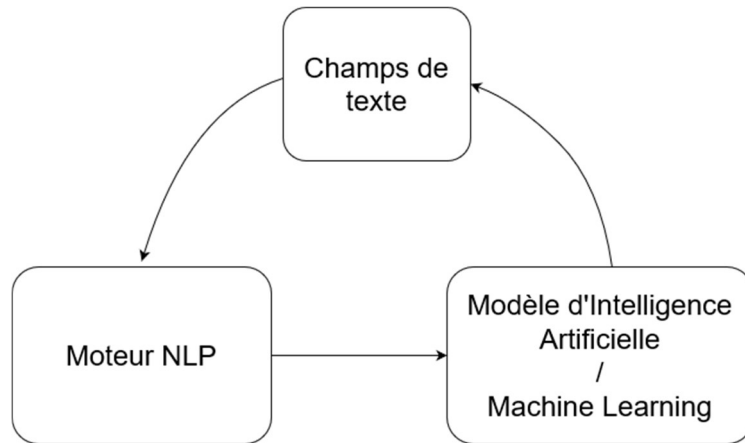


Illustration 6: Graphique d'un Chatbot Intelligent

Source : Rodrigues dos Santos Fabio

Les Chatbots intelligents permettent de pallier au problème commun aux deux types de Chatbots précédents étant le contexte restreint. Dans le cas du Chatbot de règles, il a bien été soulevé qu'ajouter de nouveaux tokens et règles rends le problème exponentiellement complexe dû à l'infinité des tournures que peut prendre une phrase. Or, un Chatbot intelligent est capable d'outrepasser ces limitations par sa capacité à comprendre le sens des mots, le contexte dans lequel ils sont employés et d'engager dans une conversation avec un utilisateur en se rappelant des informations que ce dernier a pu fournir tout au long de la discussion. Étant donné la grande variété de modèles d'intelligence artificielle, il est difficile de donner un exemple d'implémentation spécifique. Le point commun entre tous cependant est qu'ils ont tous pour but de simuler une discussion fluide et dynamique avec l'utilisateur, comme s'ils discutaient avec un agent humain.

Ce type de Chatbot est souvent retrouvé actuellement dans des assistants virtuels tels qu'Alexa, Google Assistant, Siri, Amazon Echo et bien d'autres. Il en existe aussi sous la forme de Chatbots web comme le très populaire ChatGPT, Google Bard, Bing AI qui lui est une intelligence artificielle de type Générative qui a pour but de générer du nouveau contenu à partir d'un contenu existant. Ce qui, bien que cela est intéressant, n'est pas le focus du projet ici.

Cependant, un point commun à toutes ces I.A. est le traitement des requêtes par le biais du NLU (Natural Language Understanding). Selon le graphique présenté plus haut, une fois que l'application reçoit la requête utilisateur sous forme de texte, elle subit un prétraitement grâce à l'utilisation d'un moteur NLP. Ce prétraitement peut être plus ou moins utile selon le modèle d'I.A. ou de Machine Learning utilisé car certains modèles comme des I.A. conversationnelles sont programmées pour se charger de toute la partie de Tokenisation et autres traitements NLP avant d'appliquer un traitement NLU afin de comprendre le contexte, intention et sens des phrases et mots présents dans la requête. Le tout étant finalement traité par le modèle correspondant et sa réponse renvoyée à l'utilisateur.

Il est évident qu'un Chatbot de ce type apporte une grande plus-value à l'expérience utilisateur car ce dernier donne l'impression de comprendre quelque requête que ce soit et de rendre l'expérience bien plus personnelle qu'un Chatbot avec règles ou menus pourrait fournir. Hélas, cela ne s'obtient pas aussi aisément car tout comme le moteur NLP qui nécessite une certaine quantité de données pour être entraîné et donc, efficace; il faut une vaste quantité de données d'entraînement si l'on souhaite avoir un Chatbot capable de gérer toute situation se présentant ou alors une quantité assez large et variée pour au moins gérer la plupart des situations dans le cadre d'un projet de moins grande envergure. Cela implique qu'il n'est pas impossible de mettre en place un Chatbot intelligent pour un petit projet mais qu'il faut nécessairement assez de données à disposition. C'est pourquoi un prétraitement par un moteur NLP contribue à réduire le besoin de données le plus possible si l'entrée utilisateur peut être transformée en quelque chose de plus général et moins propice à être des cas uniques non pris en compte. Car l'usage de ce moteur permet de réduire le nombre de variations possibles pour, par exemple, un mot donné et à grande échelle réduire considérablement la taille des jeux de données.

L'essentiel à retirer de ces types de Chatbots se présente ainsi :

- Les Chatbots à Menus sont simplistes tant niveau interface que dans la manière de les utiliser. Ils permettent facilement d'arriver à une réponse en peu d'entrées utilisateur, la complexité d'implémentation étant basse les rend attractif selon le besoin, surtout si le cadre dans lequel il est utilisé est déjà un cadre restreint. Cependant, l'expérience n'est pas très personnelle et ne permet pas de sortir du cadre imposé par le menu lorsque la réponse à la requête utilisateur pourrait ne pas se trouver dans le cadre imposé par l'application. Très utilisé quand l'on sait d'avance que l'utilisateur ne peut pas sortir du cadre imposé et que l'on souhaite avoir un Chatbot nécessitant le moins de maintenance possible.
- Les Chatbots par tokens permettent d'obtenir des résultats similaires à ceux d'un Chatbot à menus mais en permettant une plus grande liberté coté utilisateur avec la possibilité de mettre un champs de texte à sa disposition. En plus d'une liberté accrue, l'expérience fournie sera plus personnelle car l'utilisateur peut formuler par ses propres mots sa requête. Or, un soucis de cadre existant persiste car ce qui définit le cadre est la quantité de règles et mots-clés gérés par le Chatbot. S'il n'y a pas assez de règles ou mots-clés pour couvrir tous les cas d'usage, cela empièterait sur l'expérience utilisateur et faire en sorte qu'aucune solution viable ne soit trouvée.
- Les Chatbots intelligents sont très prisés en ce moment de par leur avantages attractifs. Ces derniers étant la personnalisation de chaque discussion, de la liberté donnée à l'utilisateur en matière de requêtes tout en gardant un cadre et que le Chatbot tentera d'y ramener l'utilisateur s'il s'en écarte trop lors d'un échange, la facilité à implémenter un système dynamique sans avoir à se préoccuper de tout l'aspect de compréhension par NLU et traitement du texte par NLP. Cela permet après coup de n'avoir qu'à donner la sortie de ces traitements au modèle pré-entraîné qui trouvera la réponse adéquate à la requête

utilisateur. Mais pour y parvenir le plus grand obstacle reste toujours les données utilisées lors de l'entraînement des divers composants du Chatbot qui impactent grandement sa qualité. S'il n'y en a pas assez ou qu'elles ne soient pas assez variées pour coller aux spécificités souhaitées, le bot pourrait peiner en premier lieu à comprendre l'entrée utilisateur et par la suite à trouver la solution adéquate. De plus, l'usage de modèles d'I.A. ou de Machine Learning nécessitent un plus grand niveau de compréhension de ces derniers afin de déterminer quels modèles et implémentations sont les plus adéquates aux cas d'usages du projet.

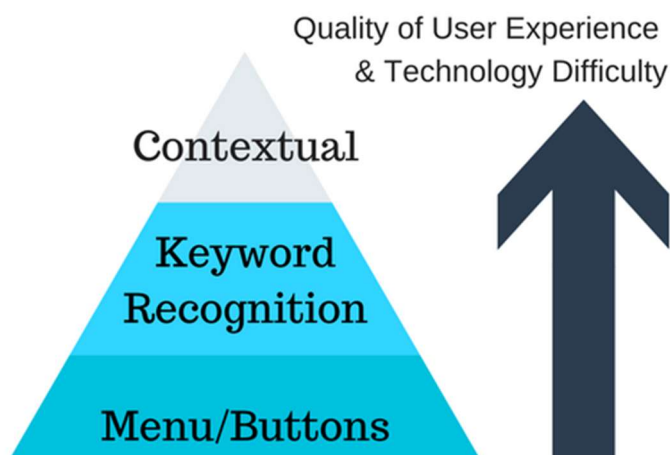


Illustration 7: Pyramide de la difficulté technologique et UX selon le type de Chatbot

Source : Ovationcxm, réf. : URL01

L'illustration ci-dessus démontre de manière visuelle la relation entre la difficulté de la technologie elle-même dans la création de Chatbots par Menus, par Règles ou Intelligents. Il est important de noter que le choix final du type de Chatbot souhaité ne dépend pas uniquement du niveau de qualité de l'expérience utilisateur fournie car sinon seul les Chatbots intelligents seraient utilisés mais la complexité d'implémentation, l'étendue des possibles interactions (tant larges que plus restreintes), ressources à disposition et autres paramètres sont à prendre en compte afin de choisir le Chatbot répondant au mieux aux attentes du projet.

Chapitre 2 : ÉTUDE DES SOLUTIONS EXISTANTES

2.1. AGENDA.CH

Agenda.ch est une application de réservation en ligne Suisse basée à Genève et lancée en 2011. Agenda.ch présente actuellement 5 services dédiés et 1 personnalisable à la demande. Les services à disposition sont :

1. Application dédiée aux physiothérapeutes
2. Application dédiée aux thérapeutes et autres services médicaux
3. Application dédiée aux centres de beauté et Bien-être
4. Application dédiée aux centres sportifs et loisirs
5. Application à visée d'administration publique

Fondamentalement, Agenda.ch fournit une application qui est un très clair exemple de Chatbot à menus. Peu importe le service, elle permet de réserver des ressources étant une salle de sport, une heure pour un rendez-vous médical, une séance pour une coupe de cheveux, etc., de choisir son créneau horaire et finalement tout autre information complémentaire.

L'aspect Chatbot n'est pas le seul utilitaire fourni par cette entreprise :

- L'application n'est pas fournie sous forme d'un module à intégrer directement dans le site web ou d'une API à appeler mais elle est disponible chez les serveurs loués par Agenda.ch et tout l'aspect de réservation est mis en place chez Agenda.ch et non chez le client.
- Selon le type de service fourni, il est possible de stocker une multitude de fichiers pertinents à la réservation des clients sur les serveurs d'Agenda.ch

- Selon le type de service fourni, un grand panel de fonctionnalités sont disponibles.
- La possibilité de configurer des messages envoyés automatiquement aux clients pour des rappels de réservations, informations, newsletter, etc.

Ce qui ressort de l'analyse de l'application d'Agenda.ch est qu'elle ne demande aucune implémentation de l'application par le client lui-même et fourni l'ensemble de ses services depuis leur site web que ce soit la configuration ou l'utilisation même du système de réservations. L'application est donc détachée du site web du client ce qui peut être vu comme un avantage étant donné que cela facilite son usage car il ne suffit que d'intégrer un simple bouton « réserver » sur le site.

Si l'on se penche sur les tarifs, on remarque qu'il y en a trois. Un à 35 CHF/mois qui n'offre pratiquement que le système de calendrier et de rappel aux clients d'une réservation. Ce n'est qu'à partir de 60 CHF/mois que l'on obtient les fonctionnalités telles qu'un mini-site permettant aux clients d'effectuer leurs réservations dessus ainsi que d'avantages de fonctionnalités en ligne pour faciliter l'aspect réservation. Ainsi, le système de réservation en ligne n'est pas forcément fourni avec tous les tarifs, rendant cette dernière comme un produit premium. Le dernier tarif étant à 80 CHF/mois propose principalement des systèmes de gestion de documents et ressources diverses directement effectuées par Agenda.ch, pratique si l'on ne possède pas de grande infrastructure ou que l'on ne souhaite pas avoir à se préoccuper d'en gérer une soi-même. Il existe évidemment des modules supplémentaires ou agendas supplémentaires selon des coûts additionnels variables.

2.2. KLARA.CH

Klara.ch est une entreprise Suisse fondée en 2016 basée à Berne et a pour but de fournir un panel varié d'outils d'administration. Contrairement à la solution précédente, Klara.ch

ne se spécialise pas uniquement dans la mise en place d'un système de réservations mais d'une panoplie d'outils d'administration comme : des gestionnaires de client, gestionnaires de budget, gestionnaires d'inventaire, création de shop en ligne, mise en place d'un système de réservations et bien d'autres. L'application de réservation en ligne ainsi que les autres fonctionnent sous forme de widgets soit des modules applicatifs que le client intègre directement à son site web. La configuration de ce dernier se fait directement au travers du panneau de configuration Klara. Le système de réservation se présente sous la forme d'un Chatbot à menus, similaire à celui d'Agenda.ch. Le tarif pour le widget de réservation en ligne est de 39 CHF/mois.

La solution de Klara.ch est intéressante de par son coût réduit par rapport à Agenda.ch et permet d'avoir l'ensemble des fonctionnalités disponibles sur le site client et de ce fait, ne requiert pas de devoir se rendre sur une autre page internet. Cependant le degré de personnalisation et de fonctionnalités supplémentaires spécifiques à certains services ne semble pas comparable à celui d'Agenda.ch, expliquant ainsi son coût plus élevé.

2.3. MEETME.IO

Meetme.io est une application de réservation de rendez-vous Suisse basée à Lausanne et ayant pour objectif de fournir un service de réservation en ligne et de gestion d'agenda grâce à une application ou depuis un panneau d'informations. Meetme.io possède un catalogue de fonctionnalités bien plus réduit comparé aux solutions précédentes en ne fournissant qu'un système de formulaire similaire à celui d'Agenda.ch. Tout comme ce dernier, il est uniquement accessible par un lien qui redirige le client sur le site de Meetme.io à l'url du formulaire de réservation du service souhaité. Meetme.io fourni par ailleurs un système de rappel par sms. Similairement aux solutions précédentes, le formulaire de Meetme.io est comparable à un Chatbot à menus. Le tarif quant à lui s'élève à 69 CHF/mois.

2.4. COMPARAISON À CE PROJET

Après analyse de quelques solutions Suisses disponibles sur le marché actuellement, il est clair qu'il y a une tendance à réaliser un système de réservation par internet. Tous fonctionnent sous la forme d'un formulaire étant ici l'équivalent d'un Chatbot à menus. Bien que certaines solutions offrent parfois plus de fonctionnalités que simplement un système de réservation ou proposent des systèmes de réservation avec des fonctionnalités étendues et spécifiques à certains secteurs, aucun ne propose de solution par téléphone uniquement. C'est donc ici que va se placer la solution qui sera proposée lors du travail de Bachelor étant donné qu'elle permettra de placer des réservations depuis un simple échange par texte avec un Chatbot par téléphone et d'ensuite éventuellement de récupérer facilement l'ensemble des réservations sous format Caldav ou d'exporter le tout en tant que calendrier google. Cela permettrait de compléter les systèmes de réservations existant d'un particulier en offrant un outil supplémentaire pour placer des rendez-vous. Voir d'entièrement remplacer leur système par un Chatbot de qualité supérieure tel qu'un Chatbot par Tokens ou I.A.

Chapitre 3 : DÉTAIL TECHNOLOGIQUE

Au cours de mes recherches concernant les diverses technologies associées à la réalisation de mon Chatbot, je me suis retrouvé à étudier quelques concepts jusque-là nouveaux pour moi et qui ont par la suite servi à affirmer mon choix technologique pour ce projet.

3.1. NATURAL LANGUAGE PROCESSING

La première technologie et une des plus importantes dans le cadre de ce projet est celle du **Natural Language Processing** (Traitement du langage naturel) (NLP). Cette technologie est une pierre angulaire au sein du domaine du machine learning et de l'intelligence artificielle. En effet, le principal intérêt de ce dernier est de permettre de prendre quelconque texte et d'appliquer divers traitements dessus en identifiant des patterns dans les phrases et mots et qui donne en sortie une suite de données structurée que les divers modèles seront capables d'exploiter afin qu'eux aussi appliquent des traitements variés sur ces données pour finalement obtenir une réponse. Ce qui sans le NLP serait bien plus ardu.

L'être humain en quelque sorte applique une multitude de techniques de NLP tous les jours pour faire sens des paroles et textes que nous lisons. De manière générale, tout traitement NLP suit globalement cette séquence :

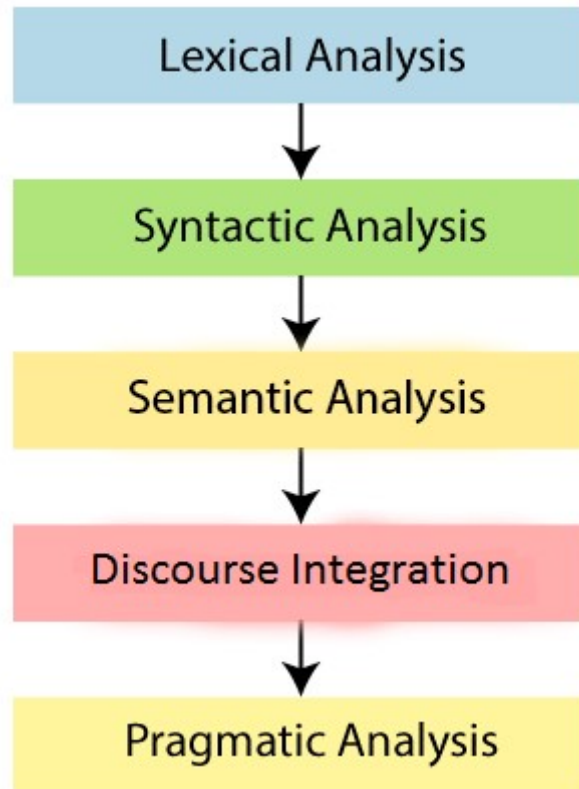


Illustration 8: Séquence de traitement NLP

Source : javapoint.com, réf. : URL02

Pour chaque étape il existe une grande quantité de traitements divers, cependant je ne vais ici qu'en présenter une liste non exhaustive :

1. Analyse Lexicale

a. Tokenisation

La Tokenisation est une des premières techniques NLP à être appliquée à un texte donné. Cette dernière a pour but de séparer tout ce qui s'apparente à des mots sous forme de Tokens.

Un exemple de Tokenisation serait « Je vais marcher dehors » qui deviendrait « 'Je', 'vais', 'marcher', 'dehors' ».

b. Lemmatisation et Stemming

La Lemmatisation et le Stemming sont présentés ensemble car ils ont un but très semblables étant de prendre un mot et de le réduire à une forme commune.

Pour se faire, la Lemmatisation va faire usage d'un dictionnaire de mots comprenant des groupes de mots similaires et étant tous lié par une même racine commune.

Par exemple : « Jouer, Joueur, Jeu, joué, ... » vont avoir pour mot racine « Jeu »

Ensuite, le Stemming lui, de par la signification du mot en anglais Stem voulant dire tige d'une plante et étant la base de cette dernière, consiste à prendre chaque mot et de les séparer en deux parties. Comme des mots de la même famille possèdent une base commune, retirer l'excès de lettres permet de dégager une racine commune.

Par exemple : « Concourût, Concourir, Concourant, Concours » vont se voir réduits à « Concour ».

2. Analyse Syntaxique

a. Part-of-speech Tagging (POS)

Le POS Tagging permet de donner un tag/label à chaque mot dans un texte afin de lui assigner sa classe correspondante grammaticalement parlant. L'intérêt principal de ce traitement est de pouvoir par la suite interpréter la structure de la phrase d'une manière grammaticale. Ensuite, donner un sens clair à certains mots qui parfois selon la structure grammaticale de la phrase où ils se trouvent peuvent avoir un tout autre sens. En d'autres mots, cela permet de contextualiser le sens d'un mot.

Par exemple : « Je me rends au pied de la montagne », les labels de POS tagging correspondants seraient « (Je, Pronom), (me, Pronom), (rends, Verbe), (au, Préposition), (pied, Nom), (de, Préposition), (la, Déterminant), (Montagne, Nom) ».

3. Analyse Sémantique

a. *Analyse de sentiments*

L'analyse de sentiments est une sorte de boîte à outils composée de nombreux outils permettant de dégager les divers émotions présentes dans un texte. Nous mentionnons cette partie par souci d'exhaustivité sans l'approfondir, car nous estimons cette approche peu adaptée à un système de réservation.

Le principe général de l'analyse de sentiments repose sur des dictionnaires de mots étant associés à des sentiments et ensuite passés au moteur NLP. Si leur présence est détectée, selon la quantité présente de sentiments ou même le manque de mots évoquant un sentiment particulier, un ou plusieurs sentiments seront assignés au texte analysé. Il est possible dans certains cas d'associer des sentiments à des principes ou objets afin d'affiner l'analyse.

b. *Named Entity Recognition*

Le NER ou Reconnaissance d'entités nommées est un des nombreux outils NLP permettant de donner davantage de sens aux tokens présents. Selon un dictionnaire associant des mots à une catégorie prédéfinie ou même une association de classes grammaticale commune avec une catégorie d'entité, l'ensemble de tokens sélectionnés est attribué à un type d'entité ou dans le cas contraire, aucun.

Par exemple : « La semaine dernière, le cours de la bourse chez Twitter a chuté de 4.8% ». Les entités ressorties de cette phrase sont : « (Semaine dernière, Date), (Twitter, Organisation), (4.8%, Pourcent) »

4. Intégration de données

L'intégration des données est le simple fait de prendre les textes précédemment analysés tant dans la même session ou dans un jeu de donnée existant. Ensuite, les diverses analyses précédemment effectués sont comparées entre elles après cela à celles réalisées sur le texte courant. Selon le niveau de ressemblance un sens/contexte sera dégagé et donné à ce dernier.

Un exemple serait que si l'on analyse une multitude de phrases ayant mentionné initialement un personnage homme et un personnage femme dans un livre et que par la suite le nom du personnage homme ne soit plus mentionné mais qu'à la place le pronom « il » est utilisé, par analyse des phrases précédentes il sera défini que les « il » font référence au personnage masculin.

5. Analyse Pragmatique

L'analyse Pragmatique n'indique pas spécialement une technique en particulier mais consiste davantage en la combinaison des techniques précédemment utilisées afin de déterminer le contexte global du texte. L'ensemble des tokens et labels ressortis précédemment vont, une fois comparés entre eux à l'aide de dictionnaires et autre jeu de données permettant d'aider à la définition du lien entre eux, donner un contexte et sens au texte courant.

Le NLP en tant que tel sert principalement à traiter du texte de manière générale pour être utilisé par la suite par d'autre algorithmes. Le NLP peut être vu comme la catégorie principale en matière de traitement de texte, cependant il existe deux sous catégories au NLP étant le Natural Language Understanding ou Compréhension de langage Naturel (NLU) et le Natural Language Generation ou Génération de langage Naturel (NLG).

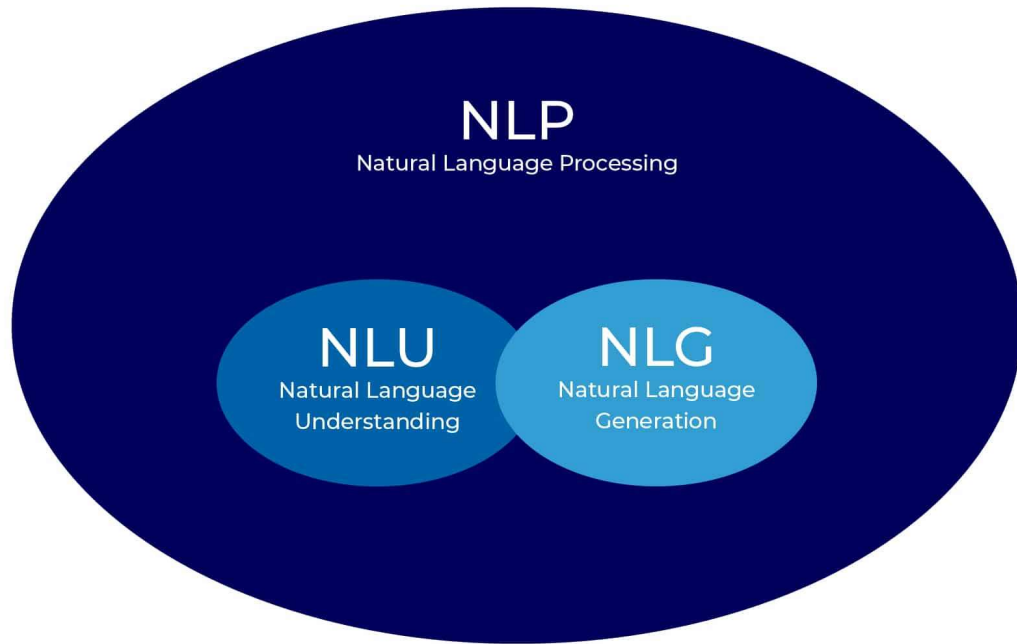


Illustration 9: Représentation de la relation NLP, NLU et NLG

Source : Datasolut, réf. : URL03

a) NATURAL LANGUAGE UNDERSTANDING

Selon Datasolut, “Le domaine du NLU s’occupe de la compréhension du langage naturel. Diverses méthodes de compréhension de texte sont utilisées à cet effet. Spécifiquement, la grammaire et le contexte de paires de mots ou mots uniques sont analysés afin d’en ressortir la signification de ces derniers et de la phrase. En outre, la sémantique, syntaxe, intention et émotion dégagée par un texte sont examinés.” ²

² WUTTKE, LAURENZ, 2023. NLP vs. NLU vs. NLG: UNTERSCHIEDE, FUNKTIONEN UND BEISPIELE. DATASOLUT GMBH [EN LIGNE]. 24 MAI 2023. DISPONIBLE À L’ADRESSE : [HTTPS://DATASOLUT.COM/NATURAL-LANGUAGE-PROCESSING-VS-NLU-VS-NLG-UNTERSCHIEDE-FUNKTIONEN-UND-BEISPIELE/](https://datasolut.com/natural-language-processing-vs-nlu-vs-nlg-unterschiede-funktionen-und-beispiele/) [CONSULTÉ LE 1 MARS 2024].

Deux concepts principaux pour parvenir à la compréhension de textes par NLU sont :

1. Détection de l'intention : Le processus d'identification des sentiments d'un utilisateur et de déterminer son objectif. Un des buts premiers du NLU étant l'établissement du sens du texte donné.
2. Détection d'entités : Un processus plus spécifique qui identifie des entités spécifiques et extrait les informations les plus importantes de ces dernières. Il existe deux types d'entités : Celles dites nommées ou numériques.

Les éléments nommés sont groupés en catégories telles que des noms d'individus, de business et de lieux.

Les éléments numériques sont reconnus en tant que quantités, dates, monnaies et pourcentages.³

Le NLU est une technologie prévalente dans le développement de Chatbots et autres assistants intelligents ayant pour but de converser avec un utilisateur et celle-ci leur permet de comprendre ce que ce dernier leur répond.

b) NATURAL LANGUAGE GENERATION

Le NLG est la deuxième sous-catégorie de NLP ici servant à générer du texte en langage naturel. Cette catégorie de NLP est une partie vitale de nombre d'I.A. conversationnelles car il ne suffit pas simplement de comprendre ce que l'utilisateur leur dit mais d'aussi pouvoir générer une réponse cohérente et intelligible par un humain. Pour se faire, on emploie de la génération de langage naturel.

³ WHAT IS NATURAL LANGUAGE UNDERSTANDING (NLU)? | DEFINITION FROM TECHTARGET, ENTERPRISE AI [EN LIGNE]. DISPONIBLE À L'ADRESSE : [HTTPS://WWW.TECHTARGET.COM/SEARCHENTERPRISEAI/DEFINITION/NATURAL-LANGUAGE-UNDERSTANDING-NLU](https://www.techtarget.com/searchenterpriseai/definition/natural-language-understanding-nlu) [CONSULTÉ LE 1 MARS 2024].

Le fonctionnement du NLG varie selon son application, cependant un modèle général est détaillé par R. Dale et E. Reiter⁴ dans une de leurs publications comme suit :

1. Détermination de Contenu et planification du discours : La détermination de contenu est une des premières étapes du NLG qui se charge de trouver quelles informations seront communiquées sous la forme d'un texte à l'utilisateur. Afin d'y parvenir, un filtre est appliqué sur l'ensemble des données afin de ne récupérer que ce qui semble pertinent et de résumer chaque donnée afin de ne garder que l'essentiel.

La planification du discours, lui, est l'action de réordonnancement des informations précédemment récupérées dans un ordre spécifique afin de ne pas exprimer ces dernières de manière disparate.

2. Aggrégation de phrases : Cette étape a pour but de prendre plusieurs messages et de les regrouper en une seule phrase. Par exemple si de multiples messages mentionnent des informations concernant un cours comme suit :

- “Le cours de Français est donné par Mr. Bournon”
- “Le cours de Français aura lieu en salle A201”
- “Le cours de Français commence à 15h”

Sera possiblement aggrégé de la manière suivante : “Le cours de Français qui aura lieu en salle A201, donné par Mr. Bournon, commence à 15h”.

Il est important de mentionner qu'il existe diverses techniques d'aggrégation et que la phrase présentée ci-dessus n'est qu'un exemple.

⁴ REITER, EHUD ET DALE, ROBERT, 1997. BUILDING APPLIED NATURAL LANGUAGE GENERATION SYSTEMS. NATURAL LANGUAGE ENGINEERING. VOL. 3, NO 1, PP. 57-87.
DOI 10.1017/S1351324997001502.

3. **Lexicalisation** : Cette étape du processus de NLG n’est pas obligatoire mais elle sert à remplacer des données présentes sous forme de concepts ou entités et de les transformer en un ou plusieurs mots à afficher.

Un exemple de lexicalisation est : Dans un cadre d’une réservation de tables dans un restaurant (“RESERVATION” -> “Une réservation”, “MARINIÈRE” “à la Marinière”, “4 TABLES” -> “de quatre tables”), on obtiendrait donc après lexicalisation : “Une réservation de quatre tables à la Marinière”

4. **Génération d’expressions référentielles** : Cette étape va se concentrer sur la désambiguïsation des diverses entités trouvées dans les informations à disposition en les remplaçant par des noms, pronoms, descriptions détaillées et autre type d’expressions référentielles.

Par exemple, si les informations contiennent des références à Microsoft, il est possible qu’après un traitement par Génération d’expressions référentielles qu’initialement Microsoft soit mentionné comme tel dans une phrase puis que par la suite ce dernier soit mentionné sous la forme de “Ce dernier”, “La compagnie ayant crée le système d’exploitation Windows”, “Ce géant de la Silicon Valley” et bien d’autres encore.

5. **Réalisation** : Cette dernière étape se charge surtout d’appliquer des correctifs orthographiques et de forme aux mots déjà présents dans la phrase mais ne se charge plus de la syntaxe, seulement de l’orthographe, conjugaison, accords et tout ce qui ne s’apparente pas à de la syntaxe.

Ce qui est intéressant à dénoter si l’on se réfère à nouveau à *l’illustration 9*, le fait que la bulle NLU se retrouve partiellement couverte par le NLG n’est pas une coïncidence car comme on le comprends avec le modèle précédemment établi, le NLU est une partie intégrante du processus de NLG. Afin de pouvoir générer une quelconque phrase, il faut

tout d’abord comprendre les données à la disposition du modèle afin de les assembler en une phrase éligible et faisant sens.

La technologie du NLP ainsi que ses deux sous-catégories enfants NLU et NLG sont des éléments fondateurs des modèles d’I.A. conversationnelles comme ChatGPT qui utilise ces deux principes pour interpréter les entrées des utilisateurs et leur renvoie une réponse.

3.2. TAPAS

La seconde technologie qui a été étudiée est TAPAS ou Parseur de Tables. C’est un modèle d’I.A. à faible supervision pré-entraîné développé par des ingénieurs de chez Google en 2020, basé sur le modèle BERT, comme détaillé dans un article de recherche publié par l’ACL (Association for Computational Linguistics)⁵.

Avant sa conception, le passage de textes sous forme de langage naturel se faisait principalement à l’aide de moules logiques décrivant en simples termes la forme que prends l’information que l’on cherche sémantiquement parlant. Ce qui demandait une quantité considérable de pré-calculs en amont du passage du texte.

C’est en réponse à ce problème qu’une approche populaire au passage sémantique a été choisie : l’utilisation d’une supervision faible. Dans le contexte du machine learning, cette dernière consiste à chercher de manière exacte le label d’une donnée. Les labels utilisés lors de l’entraînement du modèle seront volontairement incomplets car ayant du bruit sous forme de lettres en trop par exemple et peu précis comparé au label cible. Dans notre cas, cela veut dire que lors d’une requête de passage de table de donnée, au lieu de chercher le label qui correspond exactement à celui fourni par l’utilisateur dans sa requête, on va chercher les données correspondant à un label s’approchant le plus de ce dernier sans pour

⁵ HERZIG, JONATHAN ET AL., 2020. TAPAS: WEAKLY SUPERVISED TABLE PARSING VIA PRE-TRAINING. IN : PROCEEDINGS OF THE 58TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, PP. 4320-4333. 2020. DOI 10.18653/v1/2020.ACL-MAIN.398. ARXIV:2004.02349 [CS]

autant chercher une exactitude forte. Or, cette approche reste tout de même couteuse car nécessitant la génération de moules logiques même pour ces labels.

L'approche que prends TAPAS est par conséquent de ne pas avoir à générer ces moules logiques pour trouver les bons labels mais d'effectuer une sélection de cellules ainsi qu'optionnellement d'appliquer divers opérateurs à cette dernière. Comme TAPAS est basé sur BERT, un framework NLP, il est donc possible d'effectuer de simples requêtes sous forme de langage naturel et qu'ensuite TAPAS comprenne cette requête et retourne l'ensemble de données souhaité similairement à une requête SQL mais sans avoir à entrer une requête en langage SQL.

Afin de mieux comprendre le fonctionnement de TAPAS, il faut se pencher sur sa fondation étant BERT et pour se faire, demandons-nous ce qu'est un Transformer, un des composants principaux de BERT.

a) **TRANSFORMERS**

Un transformer est “un réseau neuronal qui apprend le principe de contexte et de compréhension de mots au travers d'analyse séquentielle de données”⁶. Afin d'y parvenir, une suite de N encodeurs vont analyser de manière parallèle des séquences de tokens et y appliquer le mécanisme d'attention.

Le mécanisme d'attention propre aux transformers est le simple fait de donner un poids à chaque paire de tokens et ce poids est influencé par le lien entre les deux tokens. De ce fait, après avoir passé notre séquence de tokens dans les divers couches d'encodeurs, le résultat de ces encodeurs seront passés au décodeur. Il va ensuite permettre de générer la séquence résultat selon le contexte dérivé de la valeur d'attention propre à chaque token.

⁶ THE ULTIMATE GUIDE TO TRANSFORMER DEEP LEARNING, [EN LIGNE]. DISPONIBLE À L'ADRESSE : [HTTPS://WWW.TURING.COM/KB/BRIEF-INTRODUCTION-TO-TRANSFORMERS-AND-
THEIR-POWER](https://www.turing.com/kb/brief-introduction-to-transformers-and-their-power) [CONSULTÉ LE 2 MARS 2024].

Par exemple, si l'on a une phrase étant “As-tu bien dormis ?”, grâce à un set d'entraînement qui définit le lien entre les mots présents dans cette phrase et d'autres mots commun par une valeur d'attention élevée et selon un contexte donné comme ici étant une question, une sortie possible serait “Oui, j'ai bien dormis”.

En d'autres mots, les transformers sont fondamentaux à la compréhension de textes et prédiction de phrases et suites logiques possibles à un texte donné pour le modèle BERT grâce à ce mécanisme d'attention.

b) BERT

Bidirectional Encoder Representations from Transformers (Bert) (Représentation d'encodeurs bidirectionnels de Transformers) a été créée en Octobre 2018 par Google. L'utilité principale de BERT est la compréhension de langage naturel à l'aide de transformers et étant utilisé dans des produits très mondialement connus comme pour ne citer que lui, le moteur de recherche Google :

1. Récupération de similarité d'encodage de texte : La requête de l'utilisateur est traitée par des techniques NLP et ressortie sous forme de tags. Ce qui permet à

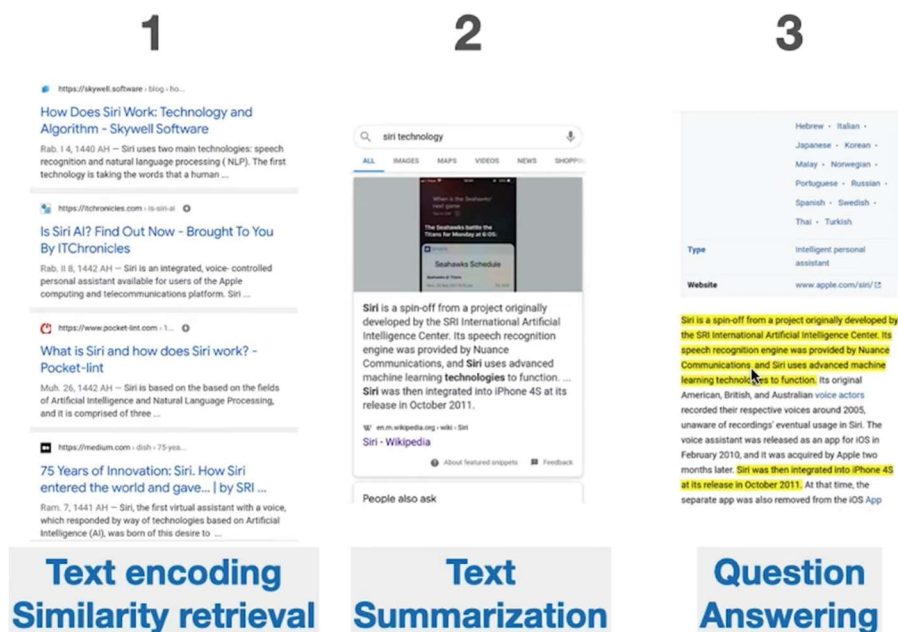


Illustration 10: Cas exemples du moteur google utilisant BERT

Source : Language processing with BERT by Jay Alammam, réf. :URL04

BERT de trouver des pages web qui possèdent des caractéristiques similaires à celles de la requête et lui retourner les pages les plus pertinentes.

2. Résumé de texte : Lors d’une recherche sur un sujet comme par exemple “Siri technology”, l’aide de diverses techniques NLP présentes dans le modèle BERT vont permettre de générer un résumé cohérent et concis des informations que l’on cherche à savoir grâce à la compréhension de contexte et sens des mots, ce à quoi BERT excelle de part l’usage de transformers.
3. Question-réponse : Une forme plus simple est en utilisant la requête utilisateur qui une fois passée dans le modèle BERT, ce dernier cherche quel passage du texte wikipedia correspond le plus à une réponse valable à la requête demandée et dans le cas présent le texte sera surligné sur ledit site.

Un exemple plus concret de BERT en application est donné par Techtarger⁷ :

La phrase que BERT reçoit ici est « The animal didn’t cross the street because it was too wide ». À un moment donné le mot “it” sera sélectionné lors du passage du texte dans les couches de l’encodeur. Il est ensuite mis face à tous les autres mots présents dans la phrase et selon l’algorithme NLP mis en place dans le modèle BERT courant, il est déterminé que IT est un pronom qui désigne une entité. De ce fait, BERT tente ensuite d’identifier à quelle entité « IT » pourrait faire référence. Les trois mots possibles ici sont

- « The »
- « animal »
- « street »

⁷ WHAT IS BERT (LANGUAGE MODEL) AND HOW DOES IT WORK?, ENTERPRISE AI [EN LIGNE].
DISPONIBLE À L’ADRESSE :
[HTTPS://WWW.TECHTARGET.COM/SEARCHENTERPRISEAI/DEFINITION/BERT-LANGUAGE-MODEL](https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model)
[CONSULTÉ LE 4 JANVIER 2024].

Après usage du mécanisme d'attention qui va calculer dynamiquement la connexion entre chaque paire de mots dans la phrase et leur attribuer un poids, le mot ayant la plus forte connexion ici est « Street ». Ayant un poids plus important, il est donc déterminé comme le mot auquel « IT » fait référence, donnant ainsi davantage de contexte.

Grâce à la lecture bidirectionnelle utilisée par BERT, le modèle est donc capable de donner un sens à des mots, phrases ou textes entiers selon la façon dont les mots sont employés et de quelle manière chaque phrase est formée. Sens qui est donné au travers de Tokens qui permettent à BERT d'être utilisé dans le cadre de :

- Question-Réponse à l'aide de divers textes donnés au modèle
- Capacité à résumer efficacement des textes
- Prédiction de phrases
- Génération de réponses lors d'une conversation, similaire à ChatGPT
- Analyse approfondie d'un texte afin de mieux cerner l'intention de ce dernier

L'entrée du modèle BERT est une suite de vecteurs nommés tokens et qui dans le cas de l'illustration qui suit représentent une phrase.

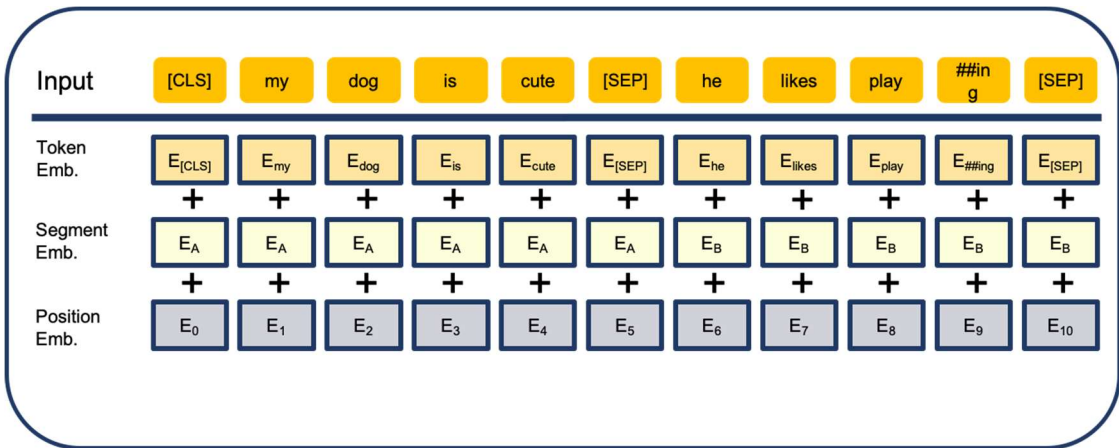


Illustration 11: Représentation de l'entrée du modèle BERT

Source : Humboldt Universität, réf. : URL05

Dans un input, le premier token est toujours un token de Classification [**CLS**], suivi ensuite ici de la première moitié du texte, d’un séparateur [**SEP**] pour marquer la fin de la première partie et le début de la deuxième et marqué pour finir d’un nouveau séparateur et ainsi de suite. Cette suite de tokens est le Token Embedding.

Ensuite une autre couche est ajoutée, le Segment Embedding. Ce dernier consiste à définir quels tokens appartiennent, dans le cas présent, à la première ou deuxième moitié du texte.

La dernière couche, le Position Embedding, permet d’indexer chaque token.

c) TAPAS : FONCTIONNEMENT

En contraste avec l’entrée du modèle BERT, l’entrée pour le modèle TAPAS se présente sous la forme suivante :

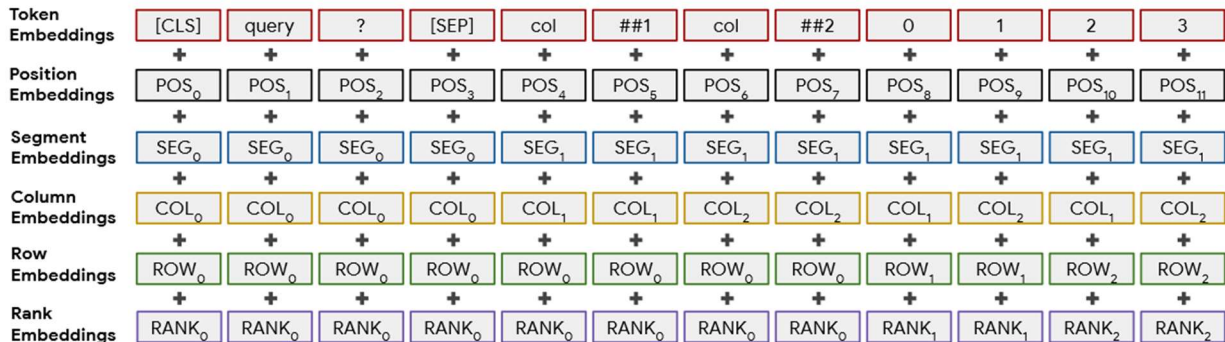


Illustration 12: Entrée du modèle TAPAS

Source : TAPAS: Weakly Supervised Table Parsing via Pre-training, réf. : URL06

Comparativement à l’entrée BERT, celle de TAPAS inclut avant la première séparation la requête de l’utilisateur et après le [**SEP**] le tableau à analyser mis à plat. Le “col” et “##1” ici représentent le nom de la première colonne et celui qui suit celui de la 2ème colonne. Il peut y en avoir autant qu’il y a de colonnes dans le tableau. Suivent ensuite les données qui sont assemblées de manière séquentielle de sorte à traverser le tableau de gauche à droite selon le nombre de colonnes et de haut en bas.

Si l'on devait créer le tableau que l'on a dans l'exemple il ressemblerait à cela :

Col1	Col2
0	1
2	3

Tableau 1: Représentation du tableau fourni en entrée de TAPAS

Source : Rodrigues dos Santos Fabio

La couche Position Embeddings est identique à celle de BERT, servant à donner des indexes à chaque token.

La couche Segment Embeddings permet de définir si un token appartient à la requête ou au tableau.

La couche Column et Row Embeddings ont pour but de spécifier dans l'entrée même si une donnée provient d'une colonne et ligne donnée afin de ne pas perdre la structure initiale du tableau.

La dernière couche, Rank Embeddings, permet de donner un ordre aux divers tokens présents. Un type d'ordre peut simplement être un ordre numérique croissant comme ici : Si les tokens présents sont "3", "4", "2" et "7", le rang correspondant à chaque token serait de l'ordre de "2", "3", "1" et "4".

La procédure qui survient après avoir fourni la requête et le tableau duquel l'on souhaite extraire des données, peut être représentée de la forme suivante :

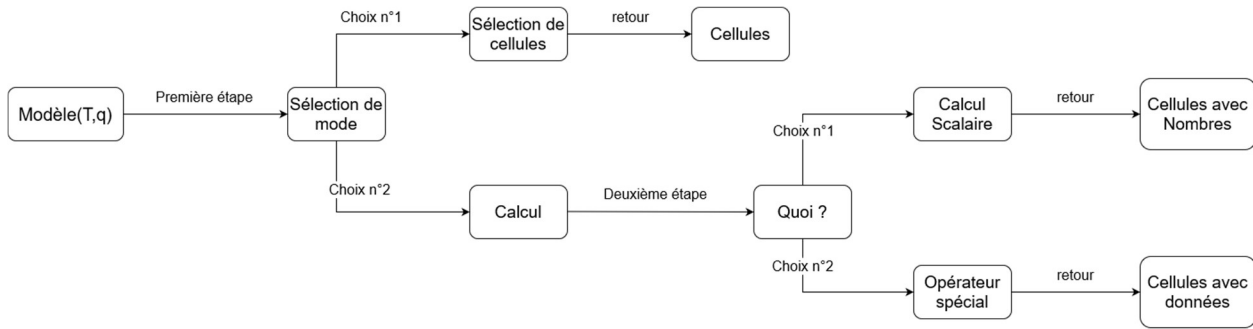


Illustration 13: Schéma de fonctionnement du modèle TAPAS

Source : Rodrigues dos Santos Fabio

La procédure définie ci-dessus démontre les divers choix qu’effectue le modèle afin de déterminer la bonne suite d’opérations à appliquer pour ressortir un résultat.

Premièrement, l’utilisateur fournit une requête et un tableau au modèle. Ensuite vient la première étape étant la sélection de mode. Les modes disponibles sont :

- Sélection de cellules
- Calcul

Si le mode choisi est une sélection de cellules, le résultat en sortie est une simple récupération de la donnée se trouvant dans la/les cellule/s permettant de répondre à la requête en entrée.

Sinon, le modèle n’aura d’autres choix que de choisir le mode calcul qui lui fait basculer le modèle dans une deuxième phase qui a pour but de choisir quelles cellules doivent être sélectionnées pour ensuite appliquer un opérateur parmi les 4 suivants :

- Rien, on renvoie simplement les cellules sélectionnées indiquant un cas de simple sélection de cellules.
- Compter, on compte le nombre de cellules sélectionnées. C’est une réponse ambiguë.
- Somme, on calcule la somme de toutes les cellules sélectionnées. C’est une réponse scalaire.

- Moyenne, on calcule la moyenne de toutes les cellules sélectionnées. C’est une réponse scalaire.

Afin de choisir les bonnes cases et opérations, un softmax (chaque choix possible reçoit un pourcentage basé sur la cohérence avec la requête) est appliqué sur chaque case et ensuite de même pour les opérateurs afin de trouver ce qui répond au mieux à la requête initiale.

En l’état, les opérateurs précédemment énoncés sont les seuls opérateurs disponibles présentement avec le modèle. Il est toutefois possible d’en ajouter soi-même, selon le papier de recherche des ingénieurs de Google⁷.

Le fonctionnement des opérations effectuées par le modèle en son état actuel dans le projet de base de Google⁸ fait qu’un tuple contenant le type d’opération et les cellules sélectionnées sont retournées, demandant donc un post-traitement des données afin d’appliquer l’opérateur choisi par le modèle.

La manière dont est entraîné le modèle est que chaque opérateur, mode et sélection de cellules possèdent, de manière inhérente à tout réseau neuronal, des poids probabilistes et leur valeur correspondante et en faisant la fonction de coût de l’erreur carrée basée sur le résultat attendu et le modèle va effectuer une propagation arrière (mise à jour des poids de chaque neurone dans le réseau selon le coût nouvellement calculé, partant) et ainsi de suite jusqu’à arriver de manière régulière à tomber sur les bons paramètres.

Une particularité est quand l’on peut avoir une solution ambiguë qui nécessite un traitement particulier car le résultat ne se trouve pas explicitement dans le tableau comme dans le cas d’un comptage de cellules. Pour se faire, un softmax sur les deux possibilités

⁸ GOOGLE-RESEARCH/TAPAS [LOGICIEL] [EN LIGNE]. 29 FÉVRIER 2024. GOOGLE RESEARCH. [CONSULTÉ LE 3 MARS 2024]. DISPONIBLE À L’ADRESSE : [HTTPS://GITHUB.COM/GOOGLE-RESEARCH/TAPAS](https://github.com/google-research/tapas) [CONSULTÉ LE 3 MARS 2024].

étant une sélection de cellules ou une agrégation sera appliquée pour déterminer la solution optimale.

Une dernière fonctionnalité intéressante de TAPAS est que l'on peut faire usage d'un système de checkpoints ou points de sauvegarde afin de pouvoir poser des questions subséquentes. Or, dans le cadre de ce projet, ce n'est pas nécessairement pertinent.

Chapitre 4 : ÉTALAGE DES TECHNOLOGIES EXISTANTES

Après avoir énoncé et approfondi quelques technologies, je m'apprête à présent à présenter quelques prototypes implémentés afin d'évaluer les modèles étudiés les plus prometteurs.

Afin de pouvoir correctement assimiler le fonctionnement des bots qui vont suivre, il est important d'explicitier le fonctionnement de Chatbots par NLU.

4.1. CHATBOTS PAR NLU

À noter que les terminologies explicités ci-dessous sont des termes trouvés principalement dans des Chatbots créés avec Rasa Open Source.

a) LES INTENTS

Le fonctionnement des Chatbots par NLU commence premièrement par la détection d'**intents**.

Un intent, soit une intention en français, est généralement un type d'action prédéfini dans le modèle d'I.A. utilisé par le Chatbot et servant à démarquer toutes les possibles actions que pourrait entreprendre l'utilisateur.

Un exemple d'intent peut être « **DETAILS_RESERVATION** » et « **RESTAURANT_RESERVATION** ». Dans un premier lieu, l'utilisateur démarre la conversation avec le Chatbot et à un certain point il pourrait dire quelque chose de la sorte : « J'aimerais réserver une table chez <Nom restaurant> ». L'intent qui sera dégagé ici sera très probablement **RESTAURANT_RESERVATION** car dans ce dernier on pourrait définir que le fait de trouver les mots suivants par l'usage de techniques de NLP : « <nom_restaurant> » et « réserver » consisterait à indiquer l'intention de l'utilisateur à réserver une table dans un restaurant. Ensuite il est possible de faire en sorte que le bot

demande à l'utilisateur des précisions sur sa réservation et si l'utilisateur fournit les paramètres de la réservation attendus par le bot, l'intention **DETAILS__RESERVATION** sera détectée et les informations fournies par l'utilisateur traitées en conséquence.

b) LES ENTITÉS

Un élément important composant bien souvent des intents sont les Entités ou Entities. Les entités peuvent être définies de multiples manières dans le contexte de développement de Chatbot par NLU mais elles sont généralement représentées par une liste de mots étant associés à une entité en particulier.

Par exemple, une entité définie comme “légumes” pourrait être trouvée lors de l'analyse de textes si les mots “Oignon”, “Carrote”, “Choux blancs”, “Poivron” et bien d'autres, sont détectés.

c) LES STORIES

Afin de faire sens d'une multitude d'intents, l'usage de stories est primordiale pour parvenir à mettre en place un flux logique de conversation et principalement pour l'entraînement du Chatbot qui sans stories, aurait de la peine à associer correctement les divers intents détectés à une action donnée.

Par exemple, si l'on a deux stories définies comme:

- Story 1
 - Intent: intent_1
 - Intent: intent_2
 - Action: bonjour

Et

- Story 2
 - Intent: intent__1
 - Action: bienvenue

Et qu'ensuite, quand l'utilisateur envoie au Chatbot une requête contenant l'intent__1 uniquement, celui-ci sélectionne la story 2. Si la requête contient également l'intent__2, alors la story 1 sera sélectionnée en fin de compte.

d) **LES ACTIONS**

Un élément apparaissant dans l'explication des stories ci-dessus sont les actions. Une action comme son nom l'indique est une suite d'instructions auxquelles la story courante fait appel. Son fonctionnement est techniquement identique à celui d'une fonction en programmation mais incorporée dans le fonctionnement de Chatbots par NLU.

e) **LES RÈGLES**

Les règles sont des suites d'actions qui sont exécutées impérativement selon l'intent ou condition spécifiée au préalable. Les règles sont similaires fonctionnellement parlant aux stories mais avec la différence qu'elles s'activent même si l'on est déjà à l'intérieur d'une story.

Par exemple, si une règle définit que lorsqu'**intent__1** est détecté, un message doit être affiché. Si l'on se trouve dans quelconque story et qu'**intent__1** est donc trouvé, le message est immédiatement envoyé en plus des autres messages éventuels du Chatbot.

f) **LES SLOTS**

Lorsque la conversation avec l'utilisateur progresse, il est possible que des informations fournies par l'utilisateur doivent être stockées afin d'influencer le reste de la conversation. C'est dans ce cas présent que les slots entrent en jeu. Les slots sont une

façon pour le Chatbot d’enregistrer en mémoire le contenu d’entités ou de tout autre type d’information le temps d’une conversation avec un utilisateur donné.

g) LES LOOKUP TABLES

Il existe une multitude de cas où nous souhaiterions détecter une entité mais qu’il existe une variété de mots correspondant à cette dernière. C’est pourquoi on fait usage d’une lookup table. Les lookup tables fonctionnent d’une manière similaire à une expression régulière dans le sens où si un mot se trouve dans une liste de mots associés à une entité, l’entité en question sera définie comme détectée dans le traitement du message utilisateur courant.

Par exemple, si l’on possède une entité “meuble” et que l’on ne souhaite pas à avoir à définir une entité pour chaque meuble, on peut alors définir une lookup table contenant “Table”, “Chaise”, “Armoire”, “Canapé” et bien d’autres. Si un seul de ces mots figure dans le message de l’utilisateur, alors l’entité meuble sera détectée.

h) LES FORMS

Les forms ou formulaires sont des moules pouvant être utilisés dans de multiples scénarios où l’on veut s’assurer de récupérer un certain nombre de slots. Pour se faire, on définit un formulaire et chaque slot associé et une fois cela fait, on utilise le formulaire nouvellement créé.

Par exemple, si l’on possède un formulaire identité avec comme slots requis “nom” et “prénom”, on peut paramétrer le Chatbot de sorte à ce qu’il persiste à demander dans la même boucle de conversation tant qu’il n’a pas reçu le **nom** et **prénom** de l’utilisateur.

i) LES PIPELINES

Dans le domaine du Machine Learning, une pipeline est la description d’une suite de traitements par lesquels passent des données allant de l’entrée à la sortie du modèle.

Dans le cadre du NLU, les éléments d’une pipeline typique sont des traitements NLP mentionnés précédemment qui, mis ensemble, permettent de créer un modèle fonctionnel.

4.2. RASA

La première technologie pouvant servir dans le cadre de la réalisation d’un Chatbot est Rasa. Rasa est un framework de création d’assistants conversationnels intelligents à l’aide de diverses techniques de machine learning créée en 2016.

a) RASA X

Avant toute chose, il est pertinent de mentionner qu’il existe un moyen d’entraîner, paramétrer et tester des chatbots créés avec une interface graphique mise à disposition par Rasa étant Rasa X.

Voici à présent ci-dessous un bref aperçu de son utilisation grâce à un tutoriel disponible sur la section blog de Rasa⁹:

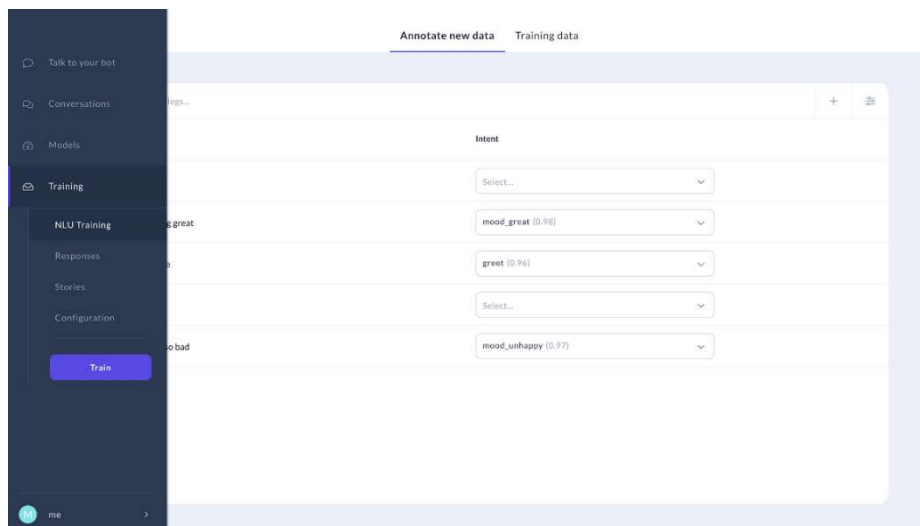


Illustration 14: Interface contenant le menu de Rasa X

Source : Rasa.com, réf. : URL07

⁹ GETTING STARTED WITH RASA X AS AN EXISTING RASA USER | RASA BLOG, RASA [EN LIGNE]. DISPONIBLE À L’ADRESSE : [HTTPS://RASA.COM/BLOG/RASA-X-GETTING-STARTED-AS-A-CURRENT-RASA-USER/](https://rasa.com/blog/rasa-x-getting-started-as-a-current-rasa-user/) [CONSULTÉ LE 3 MARS 2024].

Lorsque l'on démarre Rasa X, le menu que l'on voit ci-dessus présente les diverses pages auxquelles on peut accéder afin de :

- Gérer les modèles existant
- Entraîner un modèle
- Modifier les stories
- Modifier la configuration de la pipeline d'un modèle
- Visualiser l'historique des conversations
- Ouvrir une fenêtre de chat

Si l'on ouvre Rasa X dans un répertoire contenant un projet existant et possédant déjà des modèles pré-entraînés, ces derniers ainsi que leurs configurations apparaîtrons dans les diverses pages mentionnées ci-dessus.

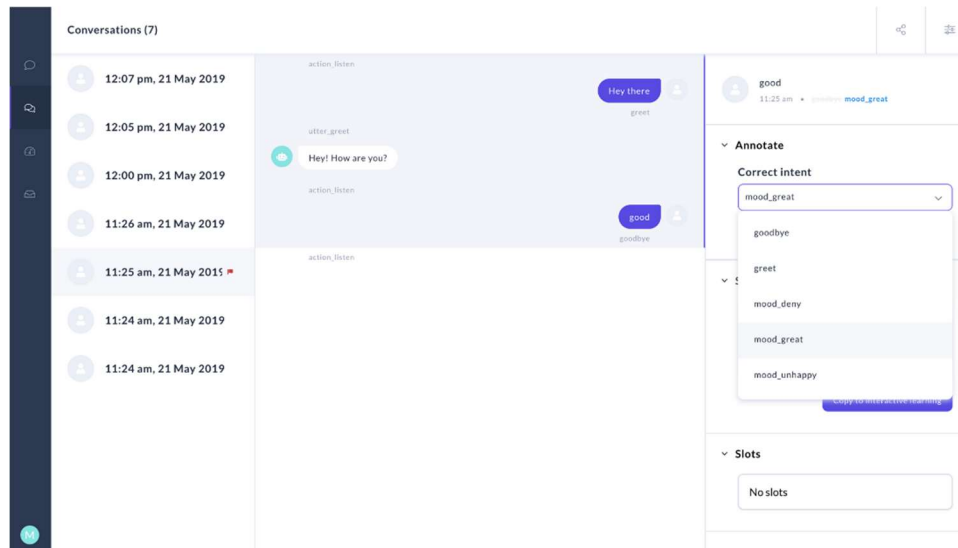


Illustration 16: Exemple de conversation parmi l'historique des conversations

Source : Rasa.com, réf. : URL08

Ci-dessus se trouve un exemple de conversation entre l'utilisateur et le chatbot et on peut apercevoir qu'il est possible de corriger presque en temps réel les prédictions faites par pas le bot si l'intent trouvé est erroné et d'autres paramètres tels que les slots afin de vérifier si les informations stockées sont correctes.

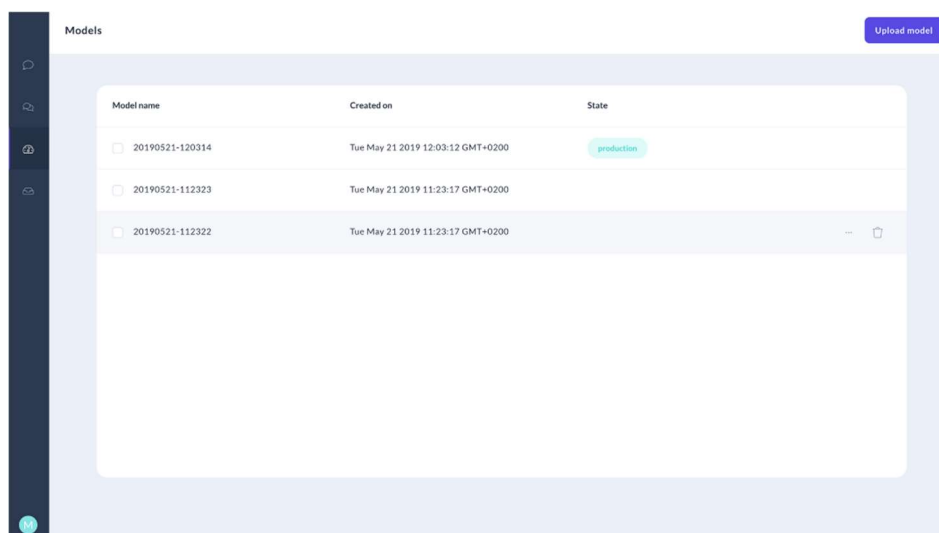


Illustration 15: Page de gestion des modèles avec Rasa X

Source : Rasa.com, réf. : URL09

Cette dernière image démontre la facilité de gestion des modèles existants, d'en supprimer ou ajouter de nouveaux au besoin et ce en quelques clics.

b) RASA OPEN SOURCE

Rasa Open Source est la solution Open Source de Rasa mettant à disposition tout un SDK disponible uniquement en python. Bien qu'il ne soit que disponible en Python il est toutefois possible d'utiliser quelconque autre langage pour traiter les données reçues du Chatbot Rasa étant donné qu'il fonctionne grâce à un serveur REST avec des points de sortie pouvant être appelé au besoin.

Rasa Open Source met à disposition un wiki très détaillé quand à l'usage du SDK. Le fonctionnement de Rasa Open Source peut être décomposé en trois parties principales.

La première, le Rasa NLU, correspond au moteur NLP défini plus tôt dans la définition d'un Chatbot. Ce dernier va se charger de détecter les divers entités et intents et de les extraire.

Le deuxième, le Rasa Core, correspond au moteur de réponses défini plus tôt dans la définition d'un Chatbot. Le Core gère tout l'aspect conversationnel en intégrant les entités et intents extraites par le Rasa NLU et de les intégrer au système de dialogues définis par les stories et règles configurées au préalable.

La troisième, L'agent Rasa, permet de fournir l'interface logicielle qui permet d'interagir avec les deux composants précédents. L'agent permet de démarrer un entraînement de modèle, gérer la réception de messages, le chargement de modèles de dialogue, la réception d'actions et la gestion de canaux externes afin de permettre au Chatbot Rasa de communiquer directement sur des applications de messageries connues comme Facebook Messenger, Slack, Mattermost et même de mettre en place une intégration sur son propre site personnel.

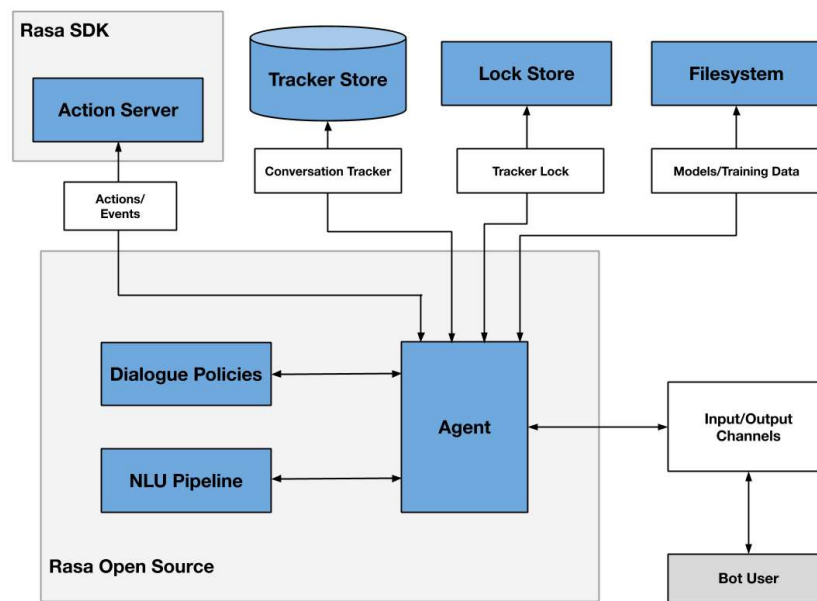


Illustration 17: Architecture Rasa

Source : Rasa.com, réf. : URL10

Rasa Core est représenté ici comme « Dialogue Policies »

Rasa NLU est représenté ici comme « NLU Pipeline »

En plus de ces trois composants, il est possible d'en ajouter quelques autres afin de rendre le Chatbot le plus modulable, efficace et maintenable :

Un quatrième composant dont on peut faire usage est un Rasa Action Server. Ce dernier utilise le Rasa SDK en python et permet fonctionnellement parlant d'utiliser les mêmes actions implicites comme explicites que les diverses commandes Rasa peuvent effectuer telles que :

- des Stories
- affectations de slots
- lancement d'évènement
- Définition de checkpoints

Le Rasa Action Server est utilisé dans le cas où l'on souhaite intégrer des actions personnalisées qui doivent, par exemple, interagir avec une API ou effectuer une multitude d'actions qui ne peuvent pas être simplement définies dans des fichiers YAML.

Un cinquième composant est Rasa Tracker Store. Un élément récurrent lors du développement avec le SDK de Rasa est le Tracker. Le Tracker est en pratique, la mémoire du Chatbot. Par défaut, le Tracker enregistre l'ensemble des conversations et événements passés. Comme le Tracker enregistre le tout en mémoire, si le bot est redémarré, l'ensemble des données seront perdues. C'est dans ce cas qu'il est possible d'utiliser un Tracker Store qui permet de lier le Tracker Rasa à une base de donnée pour stocker toutes ces informations.

Un dernier composant pouvant être utilisé est le Rasa NLG Server. Lors de l'entraînement du Chatbot, l'ensemble des réponses sont sauvegardées dans le modèle. Ainsi, si l'on souhaite modifier ne serait-ce qu'un caractère dans une réponse, il serait nécessaire de réentraîner l'ensemble du modèle. Afin d'éviter de perdre du temps à réentraîner le modèle constamment, on peut faire usage d'un NLG Server qui aura pour but de fournir les réponses à la requête utilisateur selon divers paramètres, séparant ainsi l'aspect de gestion de conversations et de gestion de réponses. Comme l'ensemble des réponses se trouve sur le serveur NLG, il ne suffit que de le redémarrer pour mettre à jour les réponses pouvant être retournées à l'utilisateur par le Chatbot.

Les divers composants présentés ci-dessus sont une liste non exhaustive des composants mis à disposition par l'outil Rasa et grâce à ceux-ci, on peut déjà si l'on le souhaite mettre en place un Chatbot bien assez robuste et efficace.

c) **RASA PRO**

Rasa Pro est un service fourni par Rasa qui contrairement à sa contrepartie Rasa Open Source, est elle Open Core. Open Core dans ce cas-ci voulant dire que toutes les

fonctionnalités de Rasa Open Source sont utilisables par les usagers de Rasa Pro. Sauf que les usagers de Rasa Pro possèdent des fonctionnalités supplémentaires et propriétaires.

Ce qui est ressorti au long de mes recherches sur Rasa Pro est que ce service fournit principalement un produit : CALM.

CALM soit **C**onversational **A**I with **L**anguage **M**odels (I.A. conversationnelle avec des modèles de langage), permet de presque entièrement se débarrasser du fonctionnement du Chatbot par NLU de Rasa et de faire usage de divers fournisseurs de LLM tels que OpenAI¹⁰, Cohere¹¹, Hugging Face Hub¹² ou Vertex AI¹³. Il est possible de faire usage d'autres fournisseurs mais cela nécessite une configuration manuelle.

CALM permet de se débarrasser du système d'entités et de stories et de faire à la place usage d'un couplage entre des Flows et un LLM au choix.

Un Flow est similaire à une story dans l'idée qu'il décrit la logique conversationnelle mais à la différence d'une story, un Flow n'a pas besoin de spécifier tous les branchements différents de conversation. À la place, il ne suffit que de définir une **description** qui sera utilisée lors de la communication du message de l'utilisateur au LLM afin de pouvoir déterminer l'action à exécuter ensuite. Il est aussi possible de faire usage de **Slots** et de **Conditions** qui seront ensuite appliqués sur ces slots selon l'état de ces derniers. Similairement à des Forms il est possible de faire usage de **Collects** qui ne passera pas à la suite de la conversation tant que l'utilisateur n'a pas donné les informations souhaitées. Il existe actuellement des limitations aux Flows qui sont dûs à la taille des contextes mis à disposition par les divers LLMs disponibles. La quantité de Flows pouvant être gérés par Rasa Pro dépend de :

- La longueur des descriptions des Flow

¹⁰ [HTTPS://OPENAI.COM/](https://openai.com/)

¹¹ [HTTPS://COHERE.COM/](https://cohere.com/)

¹² [HTTPS://HUGGINGFACE.CO](https://huggingface.co)

¹³ [HTTPS://CLOUD.GOOGLE.COM/VERTEX-AI](https://cloud.google.com/vertex-ai)

- Le nombre de slots dans chaque Flow
- La longueur de la description de chaque slot
- Les types de slots et nombres de valeurs autorisées

De plus, grâce à l’usage d’un LLM, il est possible de l’intégrer au processeur de NLG du Chatbot afin de pouvoir générer sur mesure et dynamiquement des réponses aux divers messages de l’utilisateur.

Une multitude d’autres fonctionnalités supplémentaires sont mis à disposition par Rasa Pro mais sont plus liés à de l’aide à la mise en place d’infrastructures autour du Bot et pas aussi conséquents que leur fonctionnalité CALM.

Il n’existe pas de prix fixe pour faire l’acquisition de Rasa Pro car le coût varie selon le client et ce qu’il cherche à réaliser avec ce dernier. Pour en faire l’acquisition il faut contacter directement le département des ventes de Rasa.

d) DUCKLING

Un composant découvert lors de la réalisation du prototype énoncé plus tard est Duckling. Duckling est une librairie écrite en Haskell créée par Facebook en 2017 et servant à parser du texte en une structure de donnée.

Les divers types de données étant actuellement supportées par Duckling sont appelés **domaines** et les domaines disponibles sont les suivants selon la documentation¹⁴:

Dimension	Exemple d’entrée	Exemple de sortie

¹⁴ FACEBOOK/DUCKLING [LOGICIEL] [EN LIGNE]. 21 FÉVRIER 2024. META. [CONSULTÉ LE 21 FÉVRIER 2024]. DISPONIBLE À L’ADRESSE : [HTTPS://GITHUB.COM/FACEBOOK/DUCKLING](https://github.com/facebook/duckling) [CONSULTÉ LE 21 FÉVRIER 2024].

Quantité d'argent	42€	{"value":42,"type":"value","unit":"EUR"}
Numéro de carte de crédit	4111-1111-1111-1111	{"value":"4111111111111111","issuer":"visa"}
Distance	6 miles	{"value":6,"type":"value","unit":"mile"}
Durée	3 mins	{"value":3,"minute":3,"unit":"minute","normalized":{"value":180,"unit":"second"}}
Email	<u>duckling-team@fb.com</u>	{"value":"duckling-team@fb.com"}
Nombre	Quarante-cinq	{"value":45,"type":"value"}
Nombre Ordinal	22ème	{"value":22,"type":"value"}
Numéro de téléphone	+1 (650) 123-4567	{"value":"+1 6501234567"}
Température	28°C	{"value":28,"type":"value","unit":"celcius"}
Adresse Web	<u>https://api.wit.ai/message?q=hi</u>	{"value":"https://api.wit.ai/message?q=hi","domain":"api.wit.ai"}
Volume	4 litres	{"value":4,"type":"value","unit":"litre"}
Quantité	3 cuillères à soupe de sucre	{"value":3,"type":"value","product":"sucre","unit":"tablespoon"}

Temps	Le 24 janvier 2024 à 9h00	<code>{"values":[{"value":"2024-01-24T09:00:00.000-08:00","grain":"hour","type":"value"}],"value":"2016-12-14T09:00:00.000-08:00","grain":"hour","type":"value"}</code>
-------	------------------------------	---

Tableau 2: Tableau de dimensions de Duckling

Il est d'ailleurs possible d'utiliser Duckling dans divers langages et même d'ajouter soi-même des dimensions personnalisées.

4.3. CHATTERBOT

Chatterbot est une librairie Python Open Source créée en 2014 qui permet l'élaboration de Chatbots capables d'engager dans des conversations avec un utilisateur. Chatterbot est capable de recevoir les messages utilisateur en entrée de par divers sources telles que par appels d'API, reconnaissance vocale, entrées console, etc.

À la différence de Rasa, Chatterbot n'utilise pas de techniques de compréhension de texte avancées telles que du NLU mais utilise plutôt de la détection de patterns et de similarités dans le texte. Pour parvenir à déterminer la réponse adéquate dans un cas donné, Chatterbot utilise des algorithmes de machine learning entraînés sur des conversations préparées à l'avance.

4.4. MICROSOFT BOT FRAMEWORK

Le Microsoft Bot Framework est un framework créé par Microsoft en Mars 2016 permettant de mettre en place un Chatbot conversationnel dans divers langages tels que C#, JavaScript, Python et Java (À noter que le SDK de Java n'est plus supporté depuis Novembre 2023).

Ce framework offre une panoplie de fonctionnalités similaires à Rasa comme :

- Un SDK de création de Bot
- Intégration possible de diverses applications de communication/messageries comme Microsoft Teams, Slack, Telegram et bien d'autres
- Possibilité d'utiliser du NLP et plus précisément du NLU avec LUIS (**L**anguage **U**nderstanding **I**ntelligent **S**ervice)
- Gestion de Dialogues et Conversations
- Outils d'analyses de performances et de surveillance
- Intégration facilitée avec d'autres services Microsoft comme Azure

Ce framework n'est hélas pas open source ou gratuit et nécessite un abonnement pour en faire usage pleinement. Il existe cependant un modèle gratuit mais avec des fonctionnalités réduites et un nombre de messages pouvant être envoyés aux divers canaux de messageries limité.

4.5. NLTK

NLTK est un package Python créé en 2001 et permettant de faire usage d'une multitude d'outils NLP. NLTK en tant que tel ne peut pas créer de Chatbots mais son usage peut être très intéressant lorsque l'on souhaite appliquer des pré-traitements à une entrée utilisateur avant de fournir ce dernier au chatbot pour accroître la précision des réponses et efficacité du bot.

4.6. OPENNLP

OpenNLP est un package développé par la fondation Apache depuis 2001 et ayant pour principale utilisation, comme NLTK, l'usage des divers outils NLP qu'il propose. Cependant, il est important de noter que contrairement à NLTK qui est en Python, NLTK ne peut être qu'utilisé en Java et langages supportant l'interopérabilité avec Java comme

Scala ou Kotlin. De plus, le langage Python est souvent favorisé dû à la grande quantité de modèles existant et le nombre de librairies à disposition pour divers processus de machine learning.

Chapitre 5 : PRÉSENTATION DES PROTOTYPES

Afin d'évaluer certaines technologies ainsi que la faisabilité du projet en faisant usage de ces dernières, quelques prototypes ont été réalisés.

5.1. PROTOTYPES NLP

Dans un premier lieu, j'ai pensé faire usage uniquement de bibliothèques fournissant des outils NLP afin de réaliser le moteur de traitement de données par NLU qui permettra de dégager les mots-clés nécessaires à l'établissement de règles pour mettre en place un Chatbot par règles.

Pour se faire, j'ai essayé deux bibliothèques. Une en Java : OpenNLP et l'autre en Python : NLTK.

a) JAVA

Concernant mon premier prototype en Java avec OpenNLP, j'ai dans un premier temps implémenté un Tokenizer afin de pouvoir séparer chaque mot dans mes phrases en tokens afin de faciliter la suite des opérations.

```
Phrase en entrée : Je veux réserver un terrain de tennis  
  
Je  
veux  
réserver  
un  
terrain  
de  
tennis
```

Comme on le remarque, la sortie de ce programme est bien la phrase d'origine mais correctement tokenisée. Cependant, la raison pour laquelle je n'ai pas poussé bien plus loin le développement de ce prototype est pour la simple et bonne raison que bien qu'il existe une certaine variété de librairies, modèles de machine learning pré-entraînés et jeux de données divers; ces derniers ne sont pas en nombre assez conséquent et reviendraient à demander une charge de travail bien plus grande que si je réalisais un prototype en Python. En effet, le manque niveau quantité des éléments précédemment cités implique que pour avoir un moteur de traitement NLP en Java, il faudrait créer un bon nombre de jeux de données qui n'existent simplement pas pour les librairies Java existantes.

b) PYTHON

Après la tentative précédente en Java, j'ai donc décidé de changer le langage pour passer à Python et faire usage de la librairie NLTK.

En l'état, le prototype permet de récupérer une entrée texte ou un fichier et de traiter ce dernier afin de recevoir en sortie une décomposition de chaque phrase avec chaque mot devenu des tokens et ayant leur classe grammaticale spécifiée. De plus, un traitement appliqué après le traitement par POS tagging est l'application d'un NER afin de détecter quand une date se trouve dans une phrase et lorsqu'un sport comme Tennis est détecté.

Texte de base : "Je souhaite réserver un terrain de tennis. Je veux réserver pour le 24 Novembre 2024."

NEXT SENTENCE:

```
(Je, PRON) (souhaite, VERB) (réserver, VERB) (un, DET) (terrain, NOUN) (de, ADP) (tennis, SPORT) (., PUNCT)
```

NEXT SENTENCE:

Special case(le 24 Novembre 2024, DATE)

```
(Je, PRON) (veux, VERB) (réserver, VERB) (pour, ADP) (('le', 'DET'), ('24', 'NUM')) (., PUNCT)
```

Afin de parvenir à ce résultat, divers jeux de données et librairies ont été utilisés :

- Punkt Tokenizer disponible avec NLTK : Punkt fournis des fonctions liées à la tokenisation de phrases et mots.
- Maxent_ne_chunker disponible avec NLTK : Maxent permet d'effectuer du « chunking ». Le chunking est le fait de prendre de multiples mots ayant été traités par du POS Tagging au préalable et selon des paternes donnés, ressortir un chunk. Un chunk est une nouvelle entité qui est le produit de la combinaison d'un certain nombre de tokens. L'exemple dans le cas du Prototype est que dans son implémentation actuelle, une date est détectée s'il y a au moins un déterminant, un nombre, un nom et à nouveau un nombre ce qui correspond à « Le 17 Janvier 2024 » par exemple. Le résultat sera donc l'ensemble de ces tokens agrégés en une entité de type **DATE**.
- Averaged_preceptron_Tagger disponible avec NLTK : Le but d'un tagger est de, selon un modèle fourni au préalable, de tagger les divers tokens selon leur POS correspondant ou classe grammaticale en Français. Le modèle de POS Tagging utilisé ici est celui de Stanford étant disponible en Français parmi d'autres langues¹⁵.
- Words Corpora disponible avec NLTK : Bien que ce package n'est pas directement utilisé dans le prototype, il serait utile dans le cadre d'identification d'erreurs orthographiques dans l'entrée utilisateur grâce à l'usage de la Distance de Levenshtein ou Edit Distance qui va permettre de mesurer la distance entre deux mots comme étant « Le nombre de suppressions, insertions ou substitutions requises pour transformer le mot initial à celui comparé »¹⁶.

¹⁵ THE STANFORD NATURAL LANGUAGE PROCESSING GROUP, [EN LIGNE]. DISPONIBLE À L'ADRESSE : [HTTPS://NLP.STANFORD.EDU/SOFTWARE/TAGGER.SHTML](https://nlp.stanford.edu/software/tagger.shtml) [CONSULTÉ LE 5 MARS 2024].

¹⁶ MEHTA, ANSH ET AL., 2021. SPELL CORRECTION AND SUGGESTION USING LEVENSHTein DISTANCE.. VOL. 08, NO 09.

Ce prototype fut intéressant car il m’a permis de prendre en main diverses techniques de NLP et de mieux assimiler leur fonctionnement afin de pouvoir davantage me projeter dans la bonne direction quant à l’usage de NLP pour la réalisation de mon Chatbot. Or, je l’ai vite compris que mon Chatbot ne serait pas réalisé de zéro avec seulement des librairies NLP pour réaliser un Chatbot de règles.

Cependant, je ne laisse pas complètement de côté ces librairies et jeux de données employés dans le prototype Python car ils pourraient être très utiles dans le cadre d’un pré-traitement de texte avant de le passer au Chatbot comme pour appliquer une correction de fautes d’orthographe pour réduire la possibilité que le Chatbot ne comprenne pas un message.

5.2. PROTOTYPE TAPAS

En attendant, une autre problématique que j’ai tenté de résoudre était celle de la recherche de ressources. Dans le cadre de ce projet, l’utilisateur sera éventuellement capable de demander une réservation pour une ressource donnée et que l’entrée utilisateur soit utilisée par le Chatbot pour qu’il puisse se renseigner sur ces mêmes ressources. C’est donc là que m’est venue l’idée de faire usage de TAPAS, qui comme expliqué plus tôt, permet de parcourir un tableau de données par langage naturel ou en d’autres termes, du texte.

Le langage préféré par la majorité de ceux qui ont employé TAPAS est Python de par le fait que le modèle réalisé par Google est aussi en Python.

De ce fait, j’ai réalisé un petit prototype avec l’aide d’une fondation de code venant de Christian Versloot sur divers exemples d’implémentations de modèles de Machine Learning, dont TAPAS.¹⁷

¹⁷ VERSLOOT, Christian. Machine Learning for Table Parsing: TAPAS. GitHub [en ligne]. Disponible à l’adresse : <https://github.com/christianversloot/machine-learning->

L'exemple que j'ai implémenté pour tester le modèle est le suivant :

a) DONNÉES DE TEST

Un certain nombre de Terrains de sport sont définis dans un fichier Terrains.csv qui comporte des données comme suit :

Id	Sport	Nom Terrain	Horaire
1	Tennis	1	1
4	Tennis	2	2
5	Badminton	1	1
8	Petanque	1	1

Tableau 3: Extrait de Terrains.csv

Source : Rodrigues dos Santos Fabio

Id	Nom	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
1	Horaire	9h00-	8h30-	8h30-	9h00-	11h30-	11h30-	None
	jour	16h00	15h50	15h50	17h00	14h00	14h00	
2	Horaire	18h00-	18h00-	18h00-	18h00-	18h00-	None	None
	nuit	22h00	22h00	22h00	22h00	22h00		

Tableau 4: Contenu de Horaires.csv

Source : Rodrigues dos Santos Fabio

articles/blob/main/easy-table-parsing-with-tapas-machine-learning-and-huggingface-transformers.md [consulté le 5 mars 2024].

b) RÉSULTATS

Avant toute chose, les requêtes décrites ci-dessous sont en anglais pour la simple et bonne raison qu'en l'état, le modèle utilisé pour la compréhension du Français a bien plus de peine à saisir l'intention des requêtes que celui en Anglais.

Le prototype va premièrement charger dans le modèle TAPAS **Terrains.csv** avec la requête suivante "What are all the available tennis courts by name column ?". Le résultat sortant du modèle TAPAS après traitement :

```
[['Tennis 1', 'Tennis 2', 'Tennis 3', 'Tennis 4', 'Tennis 1', 'Tennis 4']]
```

Une fois parsé et utilisé comme moyen de récupérer les données dans le fichier .csv on obtient :

Id_Dataframe	id	sport	nom terrain	horaire
0	1	1	Tennis 1	1
10	11	1	Tennis 1	2
1	2	1	Tennis 2	1
2	3	1	Tennis 3	1
3	4	1	Tennis 4	2
11	12	1	Tennis 4	1

Comme on le voit, nous avons bien reçu l'ensemble des terrains de Tennis comme demandé. Après cela, la requête qui suit est la suivante : « Which horaire by id allows for a reservation on Tuesday at 18h30 ? »

Le modèle nous réponds ensuite :

```
[['horaire nuit']]
```

c) CONCLUSION

Bien que dans le cadre de ces requêtes allant de simples à un niveau de complexité supérieur, une problématique principale posée par TAPAS est que le modèle n'est pas capable, comme une base de données traditionnelle, de faire usage de clés étrangères et

de références à d'autres tableaux par un identifiant unique. De ce fait, nous sommes contraints d'appliquer des traitements supplémentaires entre chaque requête.

De plus, la raison pour laquelle les requêtes ci-dessus sont en anglais est que bien que le modèle soit capable de comprendre les requêtes effectuées en Français, il lui arrive bien plus souvent de ne pas comprendre assez clairement la nature de la demande et de ne pas retourner le résultat attendu.

Pour finir, un des problèmes majeurs est le simple format de données car dans le cas où le Chatbot parle avec une multitude de clients, ce dernier devra modifier en continu le contenu .csv en mémoire ou le fichier lui-même ce qui ajoute un temps de traitement considérable et pourrait résulter en des soucis de cohérence et de concurrence car plus il y aurait de clients, plus le bot serait lent. En effet, pour assurer le fonctionnement du système de réservation, il est nécessaire d'effectuer des pré-réservations et comme chaque instance du bot essaierait d'interagir en même temps avec le fichier, des soucis de synchronicité surviendraient.

En conclusion, TAPAS est un modèle très intéressant lorsque l'on souhaite plus aisément chercher des informations dans un tableau sans avoir à effectuer des manipulations NLP diverses soi-même. Or, dans le cadre de ce projet, l'implémentation de la gestion des réservations et des ressources serait bien trop complexe et lente. Les raisons étant :

- La quantité de données étant dépendante d'autres à l'aide de références par clé, ce que TAPAS n'arrive pas à interpréter par lui-même.
- La synchronicité du système lorsque de multiples utilisateurs viendraient à utiliser le bot en même temps
- Un traitement presque nécessaire et complexe des requêtes utilisateurs en amont. Car dans le prototype actuel, l'utilisateur pourrait demander presque ce qu'il souhaite au bot et TAPAS n'étant qu'un outil de sélection de données n'est pas capable de permettre une véritable conversation avec choix à effectuer.

Un moyen de palier à ce problème serait de rédiger une certaine quantité de requêtes au préalable avec lesquelles on est sûrs que TAPAS nous retournera le résultat attendu. Mais cela ajoute un nouveau niveau de complexité car ces requêtes peuvent varier grandement selon le cas d’usage.

- Le taux de précision des requêtes en Français implique qu’il faut entraîner à chaque fois le modèle TAPAS avec des jeux de données en lien avec notre type de réservation et plus encore pour essayer d’atteindre un niveau acceptable de précision. Ce qui actuellement est loin d’être le cas comparé à l’alternative de faire des requêtes en Anglais.

5.3. PROTOTYPE CHATBOT

Après les essais précédents et quelques recherches, je me suis finalement penché sur Rasa Open Source en Python pour réaliser un prototype de Chatbot très simpliste.

Tout d’abord, afin de pouvoir réaliser le Chatbot il faut des données pour l’entraînement du gestionnaire de dialogues afin que le Chatbot soit capable de les reconnaître par la suite.

a) DONNÉES D'ENTRAÎNEMENT

```
- lookup: resource
examples: |
- Tennis
- Tenis
- TEnnis
- tennis
- Badminton
- badminton
- pétanque
- petanque
- Pétanque
- Petanque
```

```
- intent: book_date
examples: |
- 29 Juillet 2024
- 05/02/2024
- 04.11.2025
- le 19.01.2020
- le 02/14/2022
- le 21 Avril 2024
- le 12 mars 2024
```

```
- intent: book_resource
examples: |
- Je souhaite réserver un terrain de [tennis](resource)
- J'aimerais réserver un terrain de [Tennis](resource)
- Réserve un terrain de [Tenis](resource)
- Je souhaite réserver un terrain de [petanque](resource)
- J'aimerais réserver un terrain de [Pétanque](resource)
- Réserve un terrain de [Petanque](resource)
- Je souhaite réserver un terrain de [badminton](resource)
- J'aimerais réserver un terrain de [Badminton](resource)
- Réserve un terrain de [Badminton](resource)
```

```
- intent: book_resource+book_date
examples: |
- Je souhaite réserver un terrain de [petanque](resource) pour le 6 Septembre 2025
- J'aimerais réserver un terrain de [badminton](resource) le 28 Octobre 2024
- Je veux réserver un terrain de [tennis](resource) le 4 Janvier
```

Divers Intent utilisés pour la réalisation du prototype de réservation

```
- story: reservation 2
steps:
- intent: book_resource
- action:
action_reserve_resource
- intent: book_date
- action: action reserve date
```

```
- story: reservation 3
steps:
- intent: book_resource+book_date
- action: action_reserve_resource
- action: action_reserve_date
- checkpoint: confirm_reservation
```

```
- story: confirmation_cancelled
steps:
- checkpoint: confirm_reservation
- intent: cancel
- action: action_goodbye
```

```
- story: confirmation_confirmed
steps:
- checkpoint:
confirm_reservation
- intent: confirm
- action:
utter_final_reservation
- action: action_goodbye
```

Les stories principales utilisées dans le prototype

Selon les intents actuellement présents dans le prototype, le Chatbot est capable de :

- Reconnaître quelques types de salles/terrains de sports selon le sport
- Reconnaître une demande de réservation étant soit une demande de réservation + date ou une demande de réservation simple suivie d'une date
- Reconnaître des dates grâce à l'utilisation de Duckling

Grâce ensuite aux stories décrites après, le Chatbot est capable de :

- Recevoir une demande de réservation et si aucune date n'est spécifiée, le bot demandera à quelle date
- Recevoir une demande de réservation et si une date est spécifiée, le bot va procéder à la suite de la conversation directement
- Une fois qu'une des deux manières de réserver est effectuée, le bot demandera une confirmation de la réservation et le client pourra accepter ou l'annuler.

b) EXEMPLE DE CONVERSATION

```
Votre message ->Je souhaite réserver un terrain de badminton
Rasa's response -> [{'recipient_id': 'user', 'text': 'Et à quelle
date ?'}]
Votre message ->Le 12 janvier cette année
Rasa's response -> [{'recipient_id': 'user', 'text': 'Souhaitez-vous
donc réserver pour le 2024-01-12T00:00:00.000-08:00 un terrain de
badminton ?'}]
Votre message ->oui
Rasa's response -> [{'recipient_id': 'user', 'text': 'Vous avez bien
réservé pour le 2024-01-12T00:00:00.000-08:00 un terrain de
badminton'}, {'recipient_id': 'user', 'text': 'Au revoir'}]
```

```
Votre message ->Je souhaite réserver un terrain de pétanque le
29/05/2024
Rasa's response -> [{'recipient_id': 'user', 'text': 'Souhaitez-vous
donc réserver pour le 2024-05-29T00:00:00.000-07:00 un terrain de
pétanque ?'}]
Votre message ->Oui
Rasa's response -> [{'recipient_id': 'user', 'text': 'Vous avez bien
réservé pour le 2024-05-29T00:00:00.000-07:00 un terrain de pétanque'},
{'recipient_id': 'user', 'text': 'Au revoir'}]
```

Deux exemples de conversation avec le Chatbot du prototype

De par ces deux exemples de conversation avec le bot, on peut remarquer dans un premier temps le fait que le parsing des dates est géré correctement par Duckling, même quand une date comme “Le 12 janvier cette année” est fournie et que Duckling arrive à parser cette dernière correctement en “12 janvier 2024”.

De plus, on remarque bien avec les divers messages de rappels et de confirmation que le Chatbot enregistre bien les choix de l'utilisateur et les utilise pour déterminer les prochains messages comme le fait d'inclure ou non la date de la réservation directement dans le message initial.

c) CONCLUSION

Grâce à ce prototype simpliste de Chatbot conversationnel avec Rasa Open Source, on remarque à quel point il est aisé, même avec un petit jeu de données, de mettre en place un Chatbot ainsi que tout un dialogue de réservation de terrain de sport avec des branchements selon comment l'utilisateur répond.

En plus de son fonctionnement actuel, il y a la possibilité de lui ajouter diverses interactions avec une API, base de données, système externe, etc. L'implémentation d'une ou plusieurs de ces actions agrandirait largement la quantité de branchements et interactions complexes pouvant être effectuées avec le bot. Avec l'implémentation d'un scripts de parsing de ressources fournies en CSV il serait très facile de remplir les jeux de données du Chatbot afin d'ajouter davantage de ressources et de rendre ce dernier facilement modulable dans cet aspect.

5.4. RÉFLEXION SUR L'ENSEMBLE DES PROTOTYPES

La réalisation de ces divers prototypes m'a permis de venir à quelques conclusions concernant la future implémentation de ce projet.

En premier lieu, concernant les prototypes NLP, j'étais de l'avis que si j'allais réaliser un programme faisant usage de Machine Learning que je le coderais en Python. Mais, dans l'optique de proposer une solution dans un langage pas d'habitude utilisé pour ce type de prototypes, j'ai tout de même décidé d'essayer d'en faire un en java. Bien que le prototype soit tout à fait performant et capable d'atteindre des résultats cohérents, il ne reste néanmoins pas aussi facile à utiliser et mettre en place que son compère en Python pour les raisons qui suivent :

- Davantage de solutions et bibliothèques sont uniquement créées en Python dû à sa popularité dans ce domaine.
- Pour la même raison qu'au-dessus, il y a bien moins de modèles pré-entraînés de disponibles et compatibles avec les bibliothèques Java de NLP. Nécessitant ainsi de réaliser une plus grande portion des jeux de données utilisés soi-même.
- Une certaine quantité de techniques NLP ou autres fonctionnalités sont manquantes en Java alors que très présentes en python.

Le prototype Python reste lui pertinent car comme précisé lors de la définition de ce qu'est un Chatbot, il y a toujours un traitement NLP qui se fait avant de procéder vers le moteur de réponses et pouvoir en appliquer soi-même à l'aide d'une librairie est toujours un plus.

Deuxièmement, concernant le prototype TAPAS, je pense que mes attentes de ce modèle étaient bien au-delà de mes espérances initiales et que j'ai donc bien fait de réaliser un prototype initial afin de mettre à l'épreuve ce modèle pour ne pas me retrouver bloqué par la suite. TAPAS est un très bon modèle dans ce qu'il fait étant la recherche de données dans un tableau à l'aide de langage naturel. Hélas, les obstacles tels que :

- La langue, par le fait que beaucoup plus de requêtes reçoivent des résultats erronés après mise en place du modèle Français qu'en anglais, ce qui nécessiterait davantage de Fine tuning et ajout de jeux de données pour compléter le modèle avec ses données de langue Française.
- L'impossibilité de faire usage de multiples tables à la fois sans avoir à faire recourt à une combinaison de tables et quand bien même j'ai essayé d'en combiner, les résultats étaient rarement fructueux.
- La nécessité d'être souvent réentraîné si l'on ajoute continuellement des nouvelles données et types de tableaux. Parcourir ces tableaux est fastidieux et plus coûteux que de faire des requêtes à une base de données car nécessitant des lectures de fichiers ou récupération de tableaux d'autres sources avant de même pouvoir les faire passer dans TAPAS.

Dernièrement, le prototype de Chatbot fait avec Rasa fut un des plus intéressants car permettant de simuler un véritable cas d'usage étant ici une réservation et de manière claire et efficace grâce à la documentation abondante disponible pour Rasa.

Chapitre 6 : CONCEPTION DU PROJET

Cette section et celles qui vont suivre vont à présent porter sur le travail réalisé lors de ce projet de Bachelor. Dans un premier temps, les divers technologies employées seront détaillées. Dans un second temps, nous allons voir comment est structuré l'application dans son ensemble. Ensuite, la méthode de travail employée pour réaliser le projet. Puis, nous allons nous intéresser aux composants principaux dans leur fonctionnement.

6.1. TECHNOLOGIES PRÉSENTES

a) DOCKER

Docker est un logiciel sorti en 2013 et ayant pour but premier de virtualiser un système et créant ainsi des containers pour y exécuter divers programmes par exemple. Un container dans ce contexte est comme une machine virtuelle qui va imiter tout un environnement de travail d'un système d'exploitation.

La différence ici étant qu'au lieu de virtualiser l'intégralité du système et son système d'exploitation, seul l'explorateur de fichiers sera virtualisé grâce à un moteur de conteneurs Docker. De ce fait, il n'est pas nécessaire de simuler tout un système complet et le conteneur peut librement (selon ce qui a été spécifié au préalable) faire usage des divers ressources système. Une différence à noter aussi est que contrairement à une machine virtuelle l'environnement en question n'est pas entièrement isolé de la machine hôte.

Cependant dans un cas comme le nôtre où nous souhaitons simplement déployer de manière constante avec les mêmes packages et paramètres toute l'infrastructure nécessaire

au fonctionnement du bot, ce n'est pas un problème.

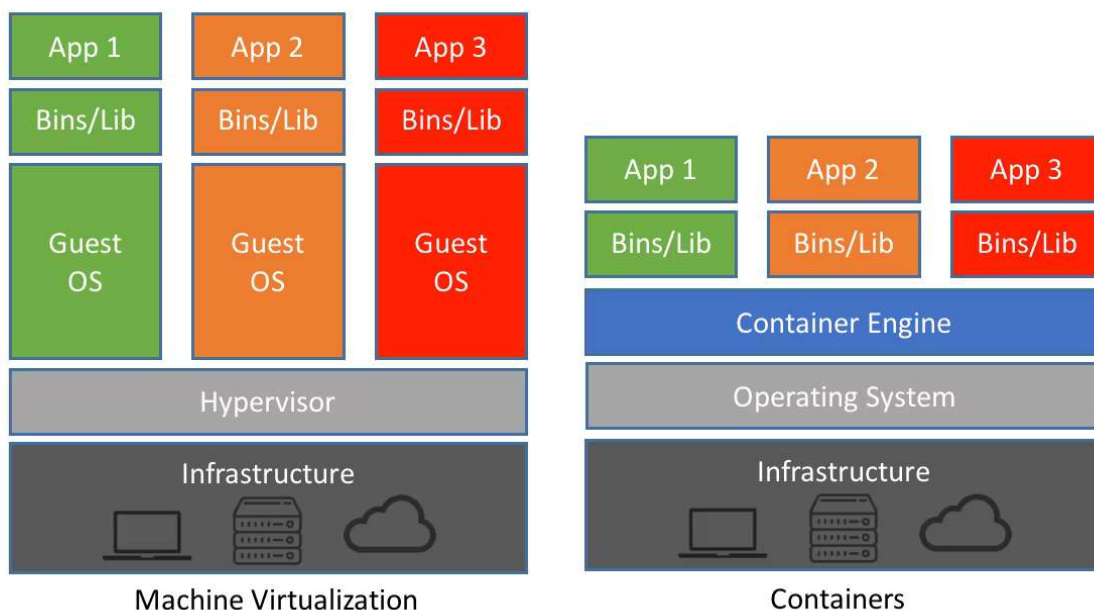


Illustration 18 : Illustration de comparaison entre une Machine Virtuelle et des conteneurs Docker

Source : Netapp.com Réf. : URL11

b) **FASTAPI**

Une API (Application Programming Interface) a pour but de mettre à dispositions divers fonctionnalités et services sans pour autant qu'il soit nécessaire de les télécharger comme dans le cas de packages. Dans ce projet, nous avons donc mis en place une API réalisée en Python avec FastAPI et fonctionnant grâce à Uvicorn.

Une API est généralement un URL auquel on envoie des requêtes et de laquelle on reçoit des informations en retour.

c) **EXPRESS JS**

Express JS remplit le même rôle que FastAPI à la différence du fait que le framework est disponible en Javascript/Typescript. Express a l'avantage de pouvoir aisément intégrer

une multitude de Middlewares et du fait qu’il est dans les langages mentionnés précédemment ce qui le rends intéressant lorsque l’on doit réaliser des projets web ayant des fonctionnalités autre que servir des routes. Et d’ailleurs, Express fait usage de NodeJS qui est un environnement serveur facilitant la mise en place et déploiement d’applications web en Frontend et Backend couramment utilisé et ce depuis bien longtemps.

d) **SOCKET.IO**

Afin de pouvoir mettre en place un système de Chat, il faut pouvoir recevoir et envoyer des messages en temps réel. C’est donc ce rôle que Socket.IO va remplir. Socket.IO est une librairie disponible en Javascript, Java, Python, Golang et Rust qui permet l’ouverture de canaux de communications entre un client et un serveur. Grâce à cela, le client peut communiquer instantanément et directement avec le serveur et recevoir une réponse uniquement adressé à ce-dernier. Très utile dans le cadre de la réalisation d’un Chat textuel entre deux partis.

e) **VITE + REACT**

Bien que Vite et React sont deux frameworks séparés, ils sont couramment utilisés pour la réalisation de Frontends dynamiques, sans latence et modulaires.

Vite a pour principal intérêt de fournir un serveur pour héberger un frontend très rapidement. Cependant ce n’est pas l’unique raison de ce choix de technologie car Vite possède une intégration native à deux autres framework de frontend single-page étant VueJS et React, ce qui dans notre cas est pertinent car nous utilisons React. Et finalement, l’avantage d’utiliser Vite avec React est que le serveur fonctionne en mode “Hot Reload” qui va rafraîchir la page dès qu’un changement survient dans les fichiers du serveur Frontend et facilitant ainsi le développement.

React est un framework Frontend en Javascript et compatible avec Typescript qui permet d’aisément mettre en place des sites web “Single Page”. Un site Single Page a

comme principal intérêt de ne pas nécessiter le rafraîchissement complet de la page car cette dernière comporte une multitude de composants qui sont mis à jour ou remplacés dynamiquement. La nature modale de ce type d'applications web le rends très attractif car le code sera naturellement modulaire et donc plus facile à maintenir et améliorer.

f) RASA

Comme observé grâce au prototype de chatbot produit lors du travail de semestre, le framework Rasa est optimal en ce qui concerne les besoins du projet et a par conséquent été sélectionné comme le pilier central.

g) DUCKLING

Explicité précédemment, Duckling a pour intérêt primaire de traiter divers données comme des dates, heures, numéros de téléphones, etc. aisément et de multiples manières. Ce logiciel est inclus dans la logique du Chatbot et du traitement des entrées utilisateur.

h) POSTGRESQL

Postgresql est un système de base de données OpenSource. Le choix d'utiliser ce dernier est car il possède autant ou plus de fonctionnalités qu'un serveur SQL, est open source et par conséquent non sujet à des licences commerciales et car il fut utilisé lors d'un projet précédent et afin d'apprendre à mieux l'employer il est devenu le système principal pour la base de données.

6.2. ARCHITECTURE GLOBALE

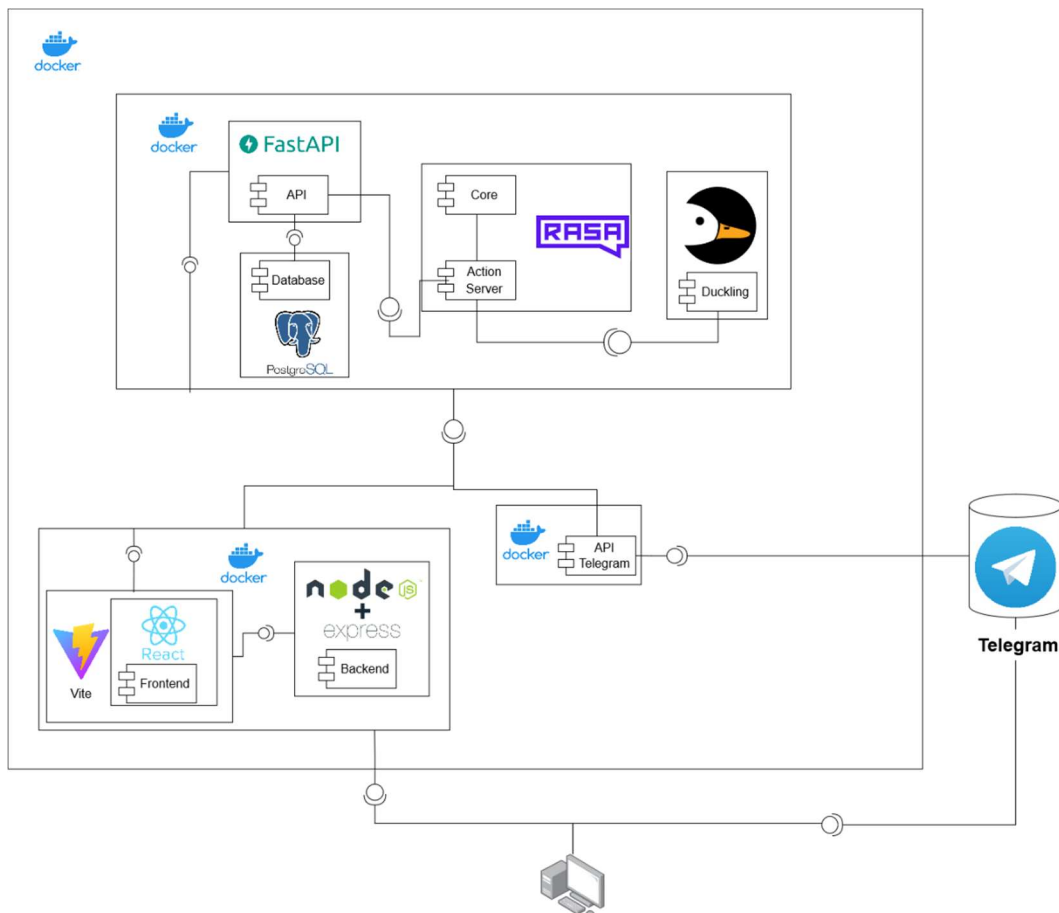


Illustration 19: Diagramme d'architecture général du projet

Source : Rodrigues dos Santos Fabio

Dans l'état, le projet peut être divisé en trois parties : Le Chatbot, l'interface Web et l'interface Telegram. Le chatbot concerne l'ensemble de la logique qui permet d'enregistrer des réservations et discuter avec l'utilisateur. Les deux interfaces proposées ont pour utilité de diversifier les points d'interaction avec le chatbot mis à disposition pour l'utilisateur.

Afin de faciliter l'usage du projet, il a été entièrement Dockerisé. Cela permet de réduire les soucis de répliquabilité de l'environnement requis pour utiliser les composants présents ainsi que rendre le démarrage de l'ensemble de ces derniers très aisé et rapide.

Les divers fichiers nécessaires à la dockerisation du projet se trouvent dans les dossiers spécifiés ci-dessous et tous appelés dans un fichier docker-compose commun permettant de démarrer tous les conteneurs dans un seul groupe de conteneurs liés au projet.

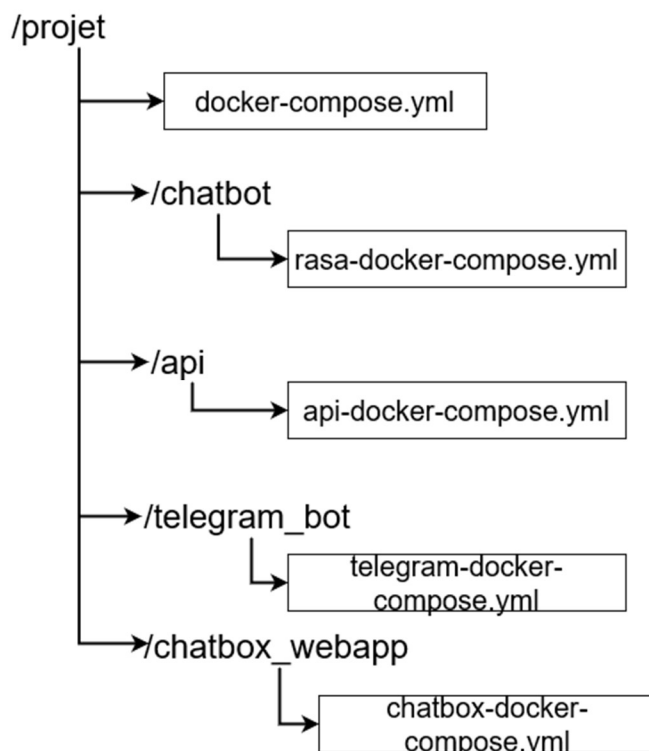


Illustration 20 : Diagramme en arbre de la structure Docker Compose du projet

Source : Rodrigues dos Santos Fabio

6.3. DÉROULEMENT DE RÉALISATION DU PROJET

Afin de s'assurer du bon déroulement du projet, la plateforme d'hébergement de GITs « Gitlab » possédée par la HES de Genève a été l'élément central concernant la gestion du projet de bachelor.

Un outil mis à disposition par Gitlab est l'issue board :

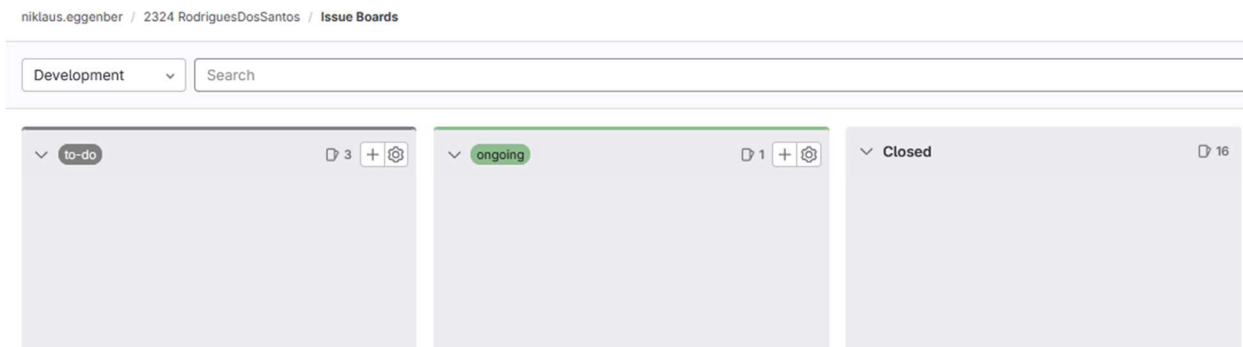


Illustration 22 : Capture d'écran d'un issue board sur Gitlab

Source: Rodrigues dos Santos Fabio

Sur GIT, un moyen de définir des tâches à réaliser est au travers d'**issues**. Une issue comporte un titre, une description, possiblement des sous-tâches et peut être labellisée.



Illustration 21 : Liste de labels sur le GIT du projet

Source : Rodrigues dos Santos Fabio

Afin de mieux définir mes tâches, j'ai attribué 2 labels principalement par tâches :

- Un label de difficulté

- Un label de priorité

Le label “ To-Do “ et “ Ongoing “ sont attribués automatiquement lorsque les tâches sont déplacées d’une colonne à l’autre de l’issue board. De ce fait, l’utilisation de l’issue board permet de donner un aspect similaire à celui de l’agile bien qu’aucun sprint ne fût défini chaque semaines pour la sélection des tâches à réaliser.

De sorte qu’un compte rendu du travail effectué chaque jour soit gardé, un journal de bord a été rédigé tout au long de la réalisation du projet.

Tout comme lors du projet de semestre, des rendez-vous hebdomadaires ont eu lieu avec le professeur attitré au projet avec pour but de confirmer l’avancée et orienter dans la bonne direction au besoin.

6.4. CHATBOT RASA

À présent, nous entamons la section dans laquelle nous déterminerons comment le bot doit se comporter de manière générale et quelles sont les diverses interactions qu’il va avoir entre chaque composants du Bot et ce qui va être finalement retourné à l’utilisateur.

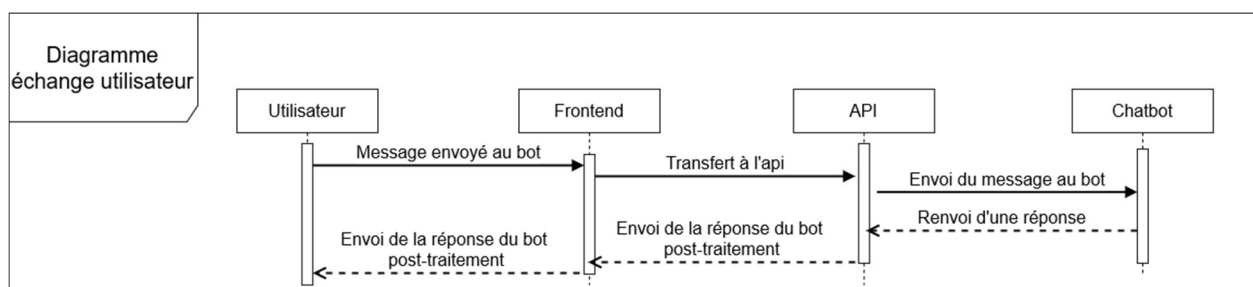


Illustration 23 : Diagramme représentant une interaction quelconque entre l'utilisateur et le bot

Source : Rodrigues dos Santos Fabio

De manière simpliste, Trois composants sont en jeu ici et permettant à l'utilisateur d'avoir des échanges. Le frontend ici pouvant être dans le cadre de ce projet à la fois celui fourni par Telegram ou le projet Web; la partie backend de chacun étant compris dans l'élément

Frontend dans le diagramme ci-dessus. L'API ici fait référence à la seule et unique interface communiquant avec le Chatbot directement. De plus, l'API communique aussi lorsque nécessaire avec la base de données pour récupérer les informations nécessaires au traitement du message. Le message reçu par le frontend de l'API sera le même qu'importe le frontend en question mais ce qui sera retourné à l'utilisateur différent car chaque frontend possède des implémentations différentes.

Intéressons-nous de plus près à un échange basique qui sera présent dans presque tous les échanges avec le bot.

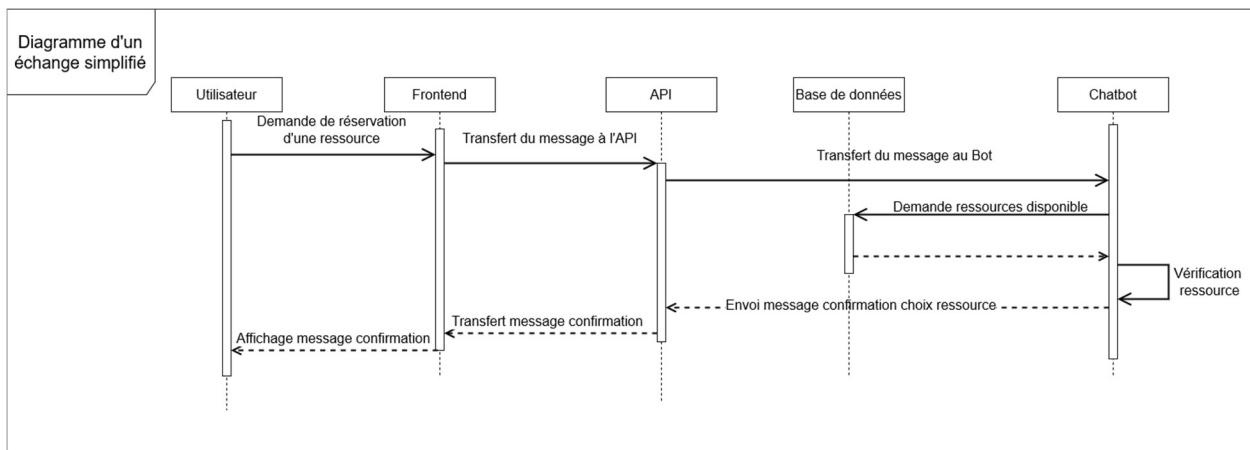


Illustration 24 : Diagramme d'un échange portant sur la demande de ressource

Source : Rodrigues dos Santos

Ci-dessus est représenté un cas où un utilisateur demande à réserver une ressource. Afin de clarifier dans un premier temps les acteurs en jeu lors d'un échange Utilisateur-Chatbot, il est assumé qu'aucune erreur ne survient pendant ce dernier. Dans un premier temps, l'utilisateur va poser la question depuis une interface. Cette dernière transmet immédiatement le message à l'API et cette dernière retransmet directement le message au Chatbot. Après avoir reçu la requête, le Chatbot effectue une multitude de traitements et peut faire appel à la base de données pour récupérer des informations telles qu'ici la liste des ressources. Une fois les informations réceptionnées, il va confirmer que la ressource

existe et est disponible puis envoyer la confirmation ainsi que d'autres informations à afficher à l'utilisateur. Le/les message(s) de confirmation seront retransmis à la chaîne jusqu'à parvenir à l'utilisateur au bout.

Ce cycle se répète presque à l'identique lors de l'intégralité de l'échange entre l'utilisateur et le Bot, à quelques différences près selon le cas et la requête.

Afin de davantage comprendre le fonctionnement interne du Chatbot, nous nous penchons

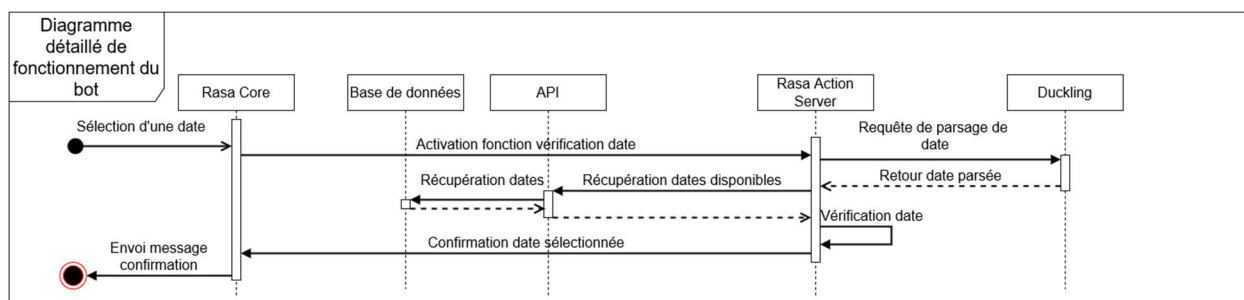


Illustration 25 : Diagramme du fonctionnement du Chatbot Rasa

Source : Rodrigues dos Santos Fabio

à présent sur la logique liée à Rasa ci-dessus. Dans ce scénario, l'utilisateur a choisi une date et cette date a été transférée jusqu'au bot qui doit à présent valider et enregistrer le choix.

Pour se faire, après avoir analysé le message de l'utilisateur et récupéré son intent et ses entités par Rasa Core, l'action de validation de date est appelée dans le Rasa Action Server. La date est envoyée au composant Duckling afin de parser la date en un format compréhensible par le code python. Une fois cela fait, l'action server va récupérer les dates disponibles en demandant à l'API qui est chargée de communiquer avec la base de données. Lorsque l'action server a reçu les dates disponibles et validé la date de l'utilisateur, elle est déclarée comme confirmée auprès de Rasa Core qui va finalement envoyer un message de confirmation à l'utilisateur et passer à l'étape suivante de la discussion.

6.5. API

L'API au centre du fonctionnement du Chatbot possède une multitude de routes permettant de récupérer diverses informations. Ci-dessous se trouve la liste des routes mises à disposition :

Route	Méthode HTML	Description	Retour de route
/add-reservation/	POST	Ajout d'une réservation à la base de données. Les données à fournir à la route sont : nom , prenom , numéro de téléphone , date , heure et liste d'options .	{ 'message' : ..., 'data' : {'lien_google' : ..., 'reservation' : ..., 'reservation_ressource' : ..., 'reservation_choix' : ..., 'id_fichier_ical' : ...} }
/communicate-rasa	POST	Route permettant de communiquer directement avec le bot Rasa. Les données à fournir sont : un id client et un message .	[{'sender' : <id_client>, 'message' : <message>}, ...]
/add-temp-reservation	POST	Ajout d'une réservation temporaire lorsqu'un client est au milieu d'une réservation afin	{ 'message' : <message>, 'data' : {'reservation' : <réservation ajoutée>}

		de ne pas perdre son créneau horaire. Les données à fournir sont : ressource , heure et date réservation	}
/get-options-choix/{ressource}	GET	Récupération des choix disponibles pour une ressource. Les données à fournir sont : Nom ressource	{ 'options' : <options> }
/get-reservations-ressources	GET	Récupération de toutes les réservations	[[id_reservation ,ressource ,date_reservation, heure], ...]
/get-reservations-ressources/{ressource}	GET	Récupération de toutes les réservations pour une ressource. Les données à fournir sont : Nom de ressource	<Même retour que la route précédente>
/get-reservations-ressources-from-	GET	Récupération de toutes les réservations	<Même retour que la route précédente>

date/{ressource}/ {date}		pour une ressource et date. Les données à fournir sont : Nom de ressource et date	
/get-jours- semaine/{ressource}/ {num_jours}	GET	Récupération des dates disponibles pour une ressource et allant d’aujourd’hui à x jours plus tard. Les données à fournir sont : Nom de ressource et nombres de jours	{ ‘dates’ : <suite de dates>, ‘horaires’ : <objets horaires> }
/get-jours- semaine/{ressource}/ {num_jours}/ {heure}	GET	Récupération des dates disponibles pour une ressource et heure et allant d’aujourd’hui à x jours plus tard. Les données à fournir sont : Nom de ressource, nombres de jours et heure	<Même retour que la route précédente>

/get-horaires/{ressource}	GET	Récupération des horaires et heures disponibles pour une ressource. Les données à fournir sont : Nom de ressource	{ ‘heures_dispo’ : <liste d’heures disponibles>, ‘horaires’ : <liste d’horaires> }
/get-horaires/{jour}/{ressource}	GET	Récupération des horaires et heures disponibles pour une ressource et date. Les données à fournir sont : Nom de ressource et date	<Même retour que la route précédente>
/get-ressources	GET	Récupération de toutes les ressources disponibles	[<liste de ressources>]
/get-ics-file/{ics_name}	GET	Récupération du fichier ICS en format base 64. Les données à fournir sont : Nom du fichier ics (ID)	{ ‘file_base64’ : <fichier base64> }

Tableau 5 : Tableau des routes disponibles sur l’API

Source : Rodrigues dos Santos Fabio

6.6. BASE DE DONNÉES

Le composant restant à expliciter concernant sa conception est la base de données. Bien qu'un calendrier Google soit mis en place afin de stocker les réservations des clients, il est bien plus sûr et facile d'avoir une base de données répertoriant tout cela. D'autant plus que même si les divers ressources sont spécifiées dans des fichiers fournis au Chatbot, il est davantage plus rapide d'appeler une base de données que de les ouvrir à chaque fois.

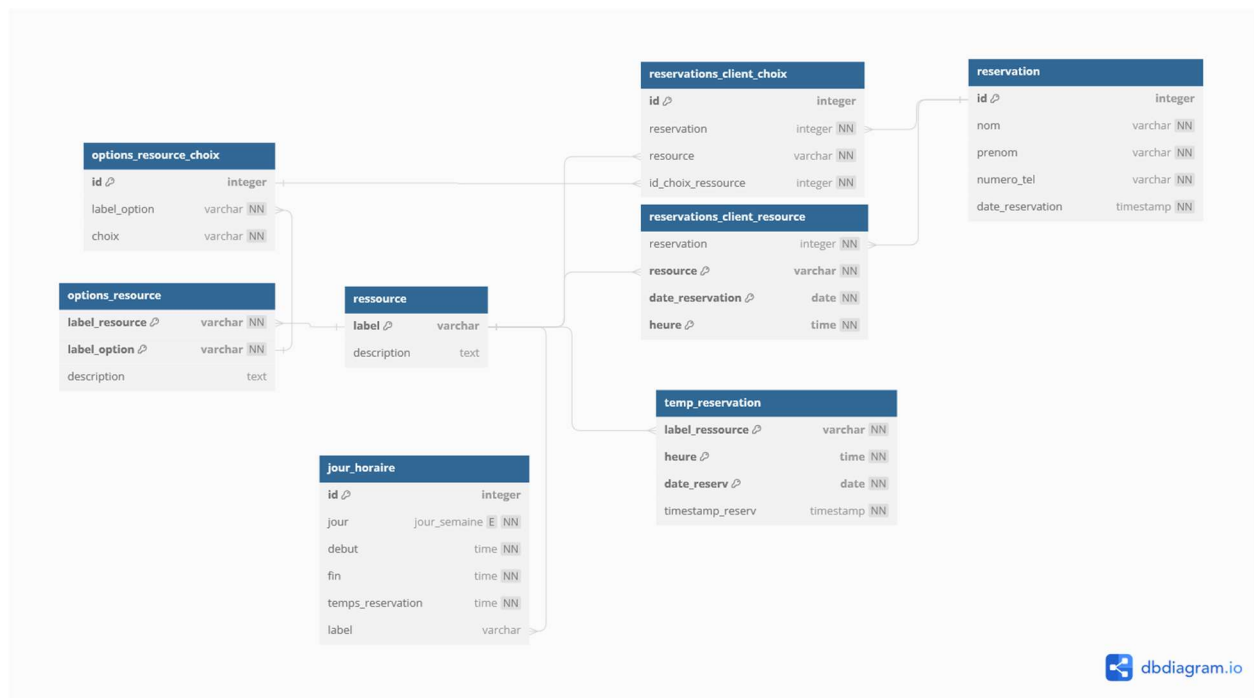


Illustration 26 : Diagramme Entité-Association de la Base de données

Source : Rodrigues dos Santos Fabio

Ci-dessus est une représentation en diagramme d'Entité-Association de la base de données utilisée dans le projet. Dans un premier lieu, la table qui sera régulièrement sélectionnée est celle des ressources. Celle-ci est liée à toutes les tables concernant les informations nécessaires à une réservation telles que :

- Jour_horaire : Table contenant les horaires par jour de semaine pour une ressource donnée.
- Options_ressource : Table contenant chaque option lié à une ressource.

- Options_ressources_choix : Table contenant cette fois-ci les choix liés à une option.
- Temp_reservation : Table contenant les réservations enregistrées temporairement afin d'éviter de multiples réservations sur un même créneau pour une ressource.

Ensuite, lorsque les informations nécessaires ont été récupérées, une réservation est en premier lieu créée dans la base de données au nom du client ainsi que d'autres informations supplémentaires. Après la création de la réservation, les tables suivantes sont remplies :

- Reservations_client_ressource : Table enregistrant l'heure, la date et la ressource réservée par une réservation.
- Reservations_client_choix : Table enregistrant les choix effectués par le client lorsqu'il y en a pour une ressource.

6.7. INTERFACES À DISPOSITION

Pour pouvoir plus aisément communiquer avec le Chatbot, deux interfaces ont été implémentées. Une application Web et un Chatbot Telegram qui est géré grâce à un package Telegram en Python.

a) TELEGRAM

La première interface, le Bot Telegram, est disponible pour tout utilisateur de Telegram à l'adresse de « https://t.me/WiREV_bot ».

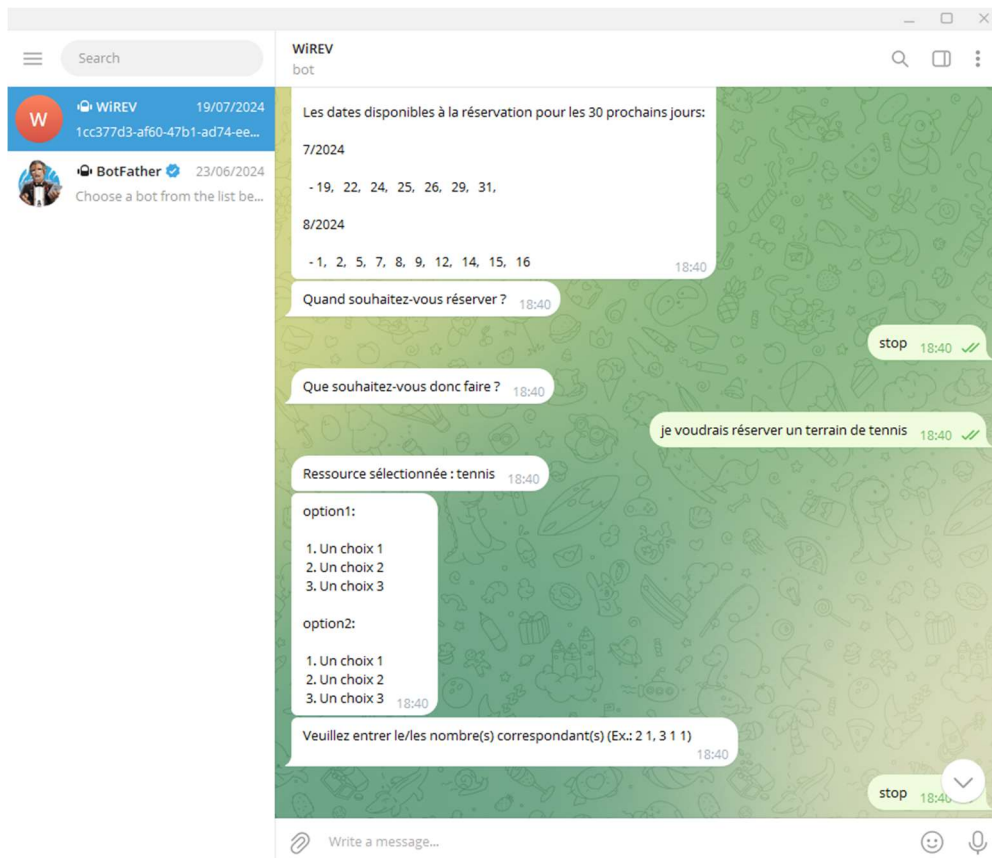


Illustration 27 : Capture d'écran du bot Telegram

Source : Rodrigues dos Santos Fabio

Lorsque le bot est ajouté comme contact Telegram, il ne suffit que d'interagir avec sans autres actions à effectuer au préalable.

b) APPLICATION WEB

La seconde interface, une application web, hébergé localement peut être aussi accédé de toute part si le réseau local est correctement configuré pour accepter des connexion externes.

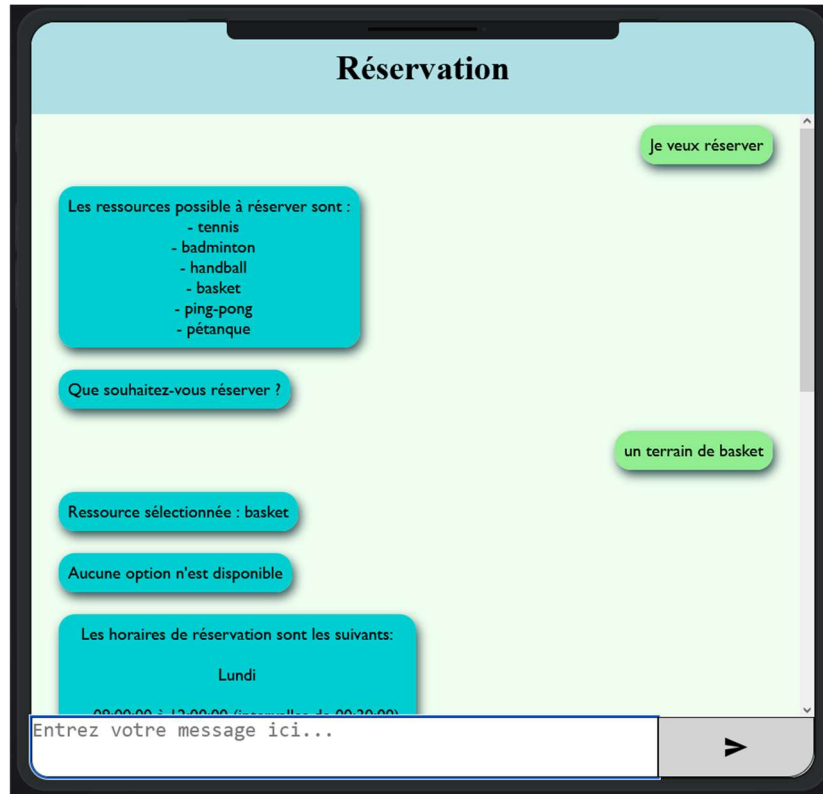


Illustration 28 : Capture d'écran d'une conversation avec le bot web.

Source : Rodrigues dos Santos Fabio

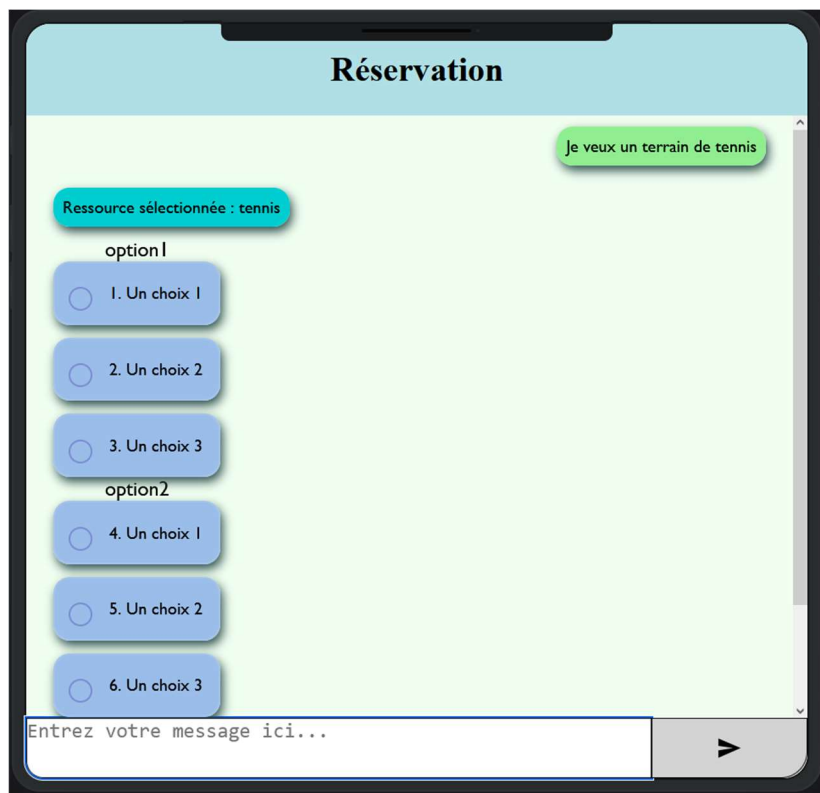


Illustration 29 : Capture d'écran de sélection de choix lors d'une réservation

Source : Rodrigues dos Santos Fabio

Chapitre 7 : RÉALISATION DU PROJET

Dans cette dernière section, nous abordons enfin comment ont été réalisé les divers composants formant le Chatbot.

7.1. IMPLÉMENTATION CHATBOT

Comme décrit précédemment, le bot au centre de ce projet est réalisé à l'aide du Framework de Bot Rasa. Framework qui on le rappelle sert à mettre en place des assistants conversationnels grâce à du NLU.

De la même manière que pour le prototype réalisé, le bot présent dans le projet actuel contient un composant Rasa Core et Action server et aussi un composant crucial étant un conteneur Duckling.

a) STORIES

Le cœur de la logique conversationnelle est réparti en quelques stories dont une principale.

- salutation_bot : Le bot salue l'utilisateur et propose à l'utilisateur quelques options sur ce qu'il souhaite faire.
- ask_horaire : Permet de directement demander l'horaire d'une ressource sans avoir besoin d'initier tout le système de réservation.
- stop_convo : Permet de mettre fin à une conversation subitement.
- goodbye : Le bot répond au revoir et mets fin à la conversation courante.
- selection_ressource : La story principale décrivant les étapes et interaction nécessaires pour faire une réservation.

Regardons de plus près un extrait de **selection_ressource** afin de comprendre comment une réservation peut être effectuée par un utilisateur.

```
- story: selection_ressource
  steps:
    - intent: ressource_gather
    - action: predefine_ressource
    - action: get_ressource_form
    - active_loop: get_ressource_form
    - slot_was_set:
      - ressource: Basket
    - slot_was_set:
      - requested_slot: null
    - active_loop: null
    - action: utter_ressource
    - slot_was_set:
      - accept_deny: None
    - action: get_options_reserv_form
    . . .
```

Ci-dessus, nous avons un extrait de la story précédemment mentionnée. Observons les premières lignes de plus près :

```
. . .
- intent: ressource_gather
- action: predefine_ressource
. . .
```

Dans la première ligne est précisé un des intents qui, s'il est détecté dans l'entrée

```
nlu:
- intent: ressource_gather
  examples: |
    - Je veux réserver un terrain de [tennis](ressource)
    - J'aimerais réserver un terrain de [badminton](ressource)
    - Je souhaite réserver un terrain de [handball](ressource)
    - Je veux réserver une table de [basket](ressource)
  ...
```

utilisateur, indiquera qu'une réservation veut être effectuée.

Lorsqu'une phrase similaire à ce qui est spécifié dans l'intent **ressource_gather** (dont un extrait est donné ci-dessus) est détecté par Rasa Core, l'action **predefine_ressource** est appelée dans le serveur d'action Rasa.

```

Action predefine_ressource :
    pre_ressource = tracker.get_slot("ressource_prereserv")
    . . .
    ressource = None
    final_return = []
    if pre_ressource is not None:
        pre_ressource_processed = str(pre_ressource).lower()
        if pre_ressource_processed not in get_ressource_list():
            dispatcher.utter_message(f"La ressource
{pre_ressource_processed} n'est pas disponible")
            final_return.append(SlotSet("ressource_prereserv",None))
        else:
            ressource = pre_ressource_processed
            final_return.append(SlotSet("ressource",pre_ressource_pr
ocessed))
    . . .

```

L'action **predefine_ressource** permet de valider une multitude d'options en une fois pour une réservation, cependant dans l'extrait présenté nous prêterons attention qu'à la réservation de ressources. Tout d'abord, nous récupérerons le contenu du slot **ressource_prereserv**. Après s'être assuré que le slot n'est pas vide, on confirme que la ressource est bien présente dans la liste des ressources avec l'appel de la fonction **get_ressource_list()**. Cette dernière fait un appel à l'API pour récupérer toutes les ressources mises à disposition dans la base de données. Si la ressource est bien présente, elle est directement affectée au slot **ressource** ce qui permet de directement présélectionner une ressource ; dans le cas contraire aucune ressource ne sera présélectionnée.

```

. . .
- action: get_ressource_form
- active_loop: get_ressource_form
- slot_was_set:
- ressource: Basket
- slot_was_set:
  - requested_slot: null
- active_loop: null
. . .

```

Ce qui suit est l'initialisation d'un **Form** qui on le rappelle permet de créer une boucle demandant une question à l'utilisateur et ce jusqu'à ce que les informations souhaitées soient fournies. Dans le cas ci présent, l'utilisateur va se voir demander quelle ressource il

souhaite réserver. Si comme expliqué précédemment l'utilisateur a déjà préréservé une ressource, alors le form prendra fin et continuera l'exécution de la conversation.

```
. . .
- action: utter_ressource
- slot_was_set:
  - accept_deny: None
- action: get_options_reserv_form
. . .
```

Ce qui suit est l'énonciation de la ressource choisie et l'initialisation du prochain form.

Une multitude d'actions vont suivre jusqu'à arriver à une réservation complète.

b) FORMS

La réalisation du bot a nécessité l'ajout de Forms afin d'assurer que l'utilisateur fournisse les informations nécessaires à la réservation. Ci-dessous se trouve quelques-uns des forms.

```
forms:
  get_ressource_form:
    required_slots:
      - ressource
    ignored_intents:
      - inform_date
      - inform_heure
      - ressource_gather

  get_date_heure_form:
    required_slots:
      - date
      - heure
    ignored_intents:
      - ressource_gather

  . . .
```

c) RULES

Les règles possèdent un rôle similaire aux stories à la différence qu'elles peuvent être appelées n'importe quand même pendant des stories. Elles servent ici principalement à assurer le début et la fin de la boucle de chaque form et permettre l'exécution de certaines actions selon des **intents** ou actions précises.

```

- rule: Activer messages de debug
  steps:
    - intent: debug
    - action: action_activate_debug_mode

- rule: recommencer
  steps:
    - intent: stop
    - action: restart_convo

- rule: activate Form ressource
  steps:
    - intent: ressource_gather
    - action: predefine_ressource
    - action: get_ressource_form
    - active_loop: get_ressource_form

```

Les deux premières règles sont ces cas où l'on veut pouvoir activer une suite d'action peu importe où l'on se trouve telle que l'activation de messages de débogage ou de pouvoir recommencer la conversation. La règle qui suit est une des nombreuses règles qui sont presque requises pour faire fonctionner les forms, sans quoi il se pourrait qu'ils ne s'activent pas ou que des comportements inattendus se produisent.

d) ACTION SERVER

L'action server est composé d'une série de classes qui possèdent un nom d'action et des fonctions associées. L'action serveur du projet possède principalement deux types d'actions :

- FormValidationAction : Une classe appelée lors de la validation de forms
- Action : Une simple fonction

Ces derniers se présentent sous cette forme :

```

class ValidateHeuresForm(FormValidationAction):
    def name(self)->Text:
        return "validate_get_date_heure_form"
    def validate_date(
        self,
        slot_value: Any,
        dispatcher: CollectingDispatcher,
        tracker: Tracker,
        domain: DomainDict,
    ) -> Dict[Text, Any]:
        . . .

```

```
class AskForResourceAction(Action):
    def name(self) -> Text:
        return "action_ask_ressource"

    def run(
        self, dispatcher: CollectingDispatcher, tracker: Tracker,
        domain: Dict
    ) -> List[EventType]:
        . . .
```

Dans le cas d'actions de type **FormValidationAction**, le nom de la classe est important car il doit toujours être de la forme « validate_<nom du form> » car sinon les actions liées à ce dernier ne seront pas correctement appelées lors de la validation du form en question. Ensuite, chaque slots qui doit être validé doit avoir une fonction de la forme « validate_<slot> ». Comme pour la classe, si le nom de fonction ne respecte pas ce format, elle ne sera pas appelée.

L'exemple d'action choisi ici n'est pas aléatoire car c'est pour démontrer un autre cas particulier toujours lié aux forms. Si l'on définit une classe avec le nom « action_ask_<nom slot> », lorsque l'utilisateur entre dans un form possédant ce slot à remplir, le code présent dans l'action sera exécuté et permettant ainsi de poser une question.

Toutes les actions possèdent trois paramètres donnant accès à certaines données importantes :

- Dispatcher : l'interface permettant de diffuser des messages à l'utilisateur avec la fonction **dispatcher.utter_message()** par exemple.
- Tracker : Objet contenant une multitude d'informations liés au message et état de divers données stockées lors de la conversation courante tels que :
 - o L'état courant du bot (Dans un form, dans un événement, affectation de slot, etc.)
 - o Le dernier message reçu

- Les intents parsés lors du dernier message de l'utilisateur et plus tôt encore
 - Les slots de la conversation courante
- Domain : Un dictionnaire contenant les diverses information spécifiées lors de la configuration du bot comme les slots, intents, réponses, forms, règles et actions présents.

7.2. IMPLÉMENTATION API

7.3. IMPLÉMENTATION INTERFACES

CONCLUSION

En ce qui concerne le travail effectué pour ce travail de semestre, les principaux intérêts furent la compréhension de ce qu'était un Chatbot concrètement parlant et comment on pouvait le décomposer en plusieurs parties distinctes qui une fois combinée permettent d'avoir un bot fonctionnel. Le focus initial a été principalement mis sur la compréhension du moteur de traitement de données étant un moteur NLP, les diverses recherches approfondies réalisées et la programmation de quelques prototypes qui ont permis de donner un bon moyen de valider ma compréhension théorique du NLP.

Ensuite, grâce à la réalisation de davantage de prototypes, j'ai pu me rendre compte que la technologie que je pensais initialement pouvoir m'être bien utile n'était finalement pas ce qui correspondait le mieux à la réalisation d'un Chatbot de réservations. Cependant, cela fut un moyen d'en apprendre davantage sur les mécanismes inhérents de divers autres applications faisant usage de NLP dans des modèles de Machine Learning comme le moteur de recherche Google avec BERT et que certains aspects étudiés lors de mes recherches sur TAPAS sont revenus comme les transformers dans ma pipeline de traitement NLU avec Rasa. Et bien que cela n'ait pas porté ses fruits en fin de compte, j'ai pu effectuer un travail d'exploration des possibilités pour la réalisation de mon futur Chatbot et déterminé plus concrètement ce qui allait contribuer à sa réalisation ou non.

Pour ma part, la rédaction de ce rapport fut très instructive car bien qu'ayant déjà fait usage de réseaux convolutifs, perceptrons ou de Réseaux neuronaux récurrents ou RNN dans d'autres cours qui m'ont permis d'analyser des images ou prédire des valeurs numériques, je n'avais pas encore fait usage de modèles de machine learning dans le cas de traitement de texte comme certaines des techniques en NLP comme du POS Tagging, NER, analyses de sentiments, etc. D'ailleurs, grâce aux rendez-vous hebdomadaires avec M. Niklaus Eggenberg, j'ai pu à la fois préciser davantage la direction dans laquelle j'allais mener ce projet et dans les moments plus troubles recevoir de la guidance afin de ne

jamais rester complètement bloqué, ce qui a un moment donné arriva presque lors de mes premières tentatives d'implémentation d'un prototype utilisant TAPAS. Additionnellement, ces rendez-vous ont eu pour effet de m'encourager à garder un rythme régulier de travail tout au long de la rédaction et de mes recherches.

Finalement, l'ensemble du travail associé à la rédaction de ce rapport m'a permis d'avoir une idée plus claire sur la façon dont je vais m'y prendre pour réaliser le Chatbot de réservations lors de mon travail de Bachelor et possiblement de scripts supplémentaires pouvant servir à rendre le bot plus modulaire, nécessitant ainsi que peu de modifications du code. En effet, en décrivant les divers ressources dans des fichiers séparés, il ne suffira que de faire usage de ces scripts pour les récupérer et faciliter la mise en place du bot. Cela permettrait donc d'aisément changer de type de réservation comme passer d'un Chatbot de réservation de rendez-vous pour dentiste à un bot de réservation de tables dans un restaurant. Une intégration intéressante au projet serait la possibilité d'exporter les réservations existantes dans le format Caldav précédemment mentionné ou sur un des services les plus couramment utilisés comme Google Calendar.

RÉFÉRENCES DOCUMENTAIRES

1. A Beginner's Introduction to NER (Named Entity Recognition), [en ligne]. Disponible à l'adresse : <https://www.analyticsvidhya.com/blog/2021/11/a-beginners-introduction-to-ner-named-entity-recognition/> [consulté le 4 janvier 2024].
2. A complete Guide to Named Entity Recognition (NER) in 2024, 2022 *Nanonets Intelligent Automation, and Business Process AI Blog* [en ligne]. Disponible à l'adresse : <https://nanonets.com/blog/named-entity-recognition-with-nltk-and-spacy/> [consulté le 12 février 2024].
3. ALBURGER, Jenna. Rule-Based Chatbots vs. AI Chatbots: Key Differences. *Hubtype* [en ligne]. Disponible à l'adresse : <https://www.hubtype.com/blog/rule-based-chatbots-vs-ai-chatbots> [consulté le 4 janvier 2024].
4. Apache OpenNLP, [en ligne]. Disponible à l'adresse : <https://opennlp.apache.org/> [consulté le 8 février 2024].
5. AYER, Benjamin, 2020. What Makes a Chatbot Conversational? *Inbenta* [en ligne]. 20 mars 2020. Disponible à l'adresse : <https://www.inbenta.com/articles/what-makes-a-chatbot-conversational/> [consulté le 4 janvier 2024].
6. BOCKLISCH, Tom et al., 2024. *Task-Oriented Dialogue with In-Context Learning* [en ligne]. arXiv:2402.12234. arXiv. arXiv:2402.12234. Disponible à l'adresse : <http://arxiv.org/abs/2402.12234> [consulté le 29 février 2024]. arXiv:2402.12234 [cs]
7. CAPACITY, Team. How does an AI chatbot work, and what does it mean for the future? *Capacity* [en ligne]. Disponible à l'adresse :

- <https://capacity.com/learn/ai-chatbots/how-does-an-ai-chatbot-work/> [consulté le 4 janvier 2024].
8. Chatbot Evolution: ChatGPT Vs. Rule-based - Analytics Vidhya, [en ligne]. Disponible à l'adresse : <https://www.analyticsvidhya.com/blog/2023/05/chatbot-evolution-chatgpt-vs-rule-based/> [consulté le 4 janvier 2024].
9. CHOUDHRY, Arjun et al., 2023. *Adversarial Adaptation for French Named Entity Recognition* [en ligne]. arXiv:2301.05220. arXiv. arXiv:2301.05220. Disponible à l'adresse : <http://arxiv.org/abs/2301.05220> [consulté le 11 février 2024]. arXiv:2301.05220 [cs]
10. CHURCH, Bella, 2023. 5 types of chatbot and how to choose the right one. *IBM Blog* [en ligne]. 5 septembre 2023. Disponible à l'adresse : <https://www.ibm.com/blog/chatbot-types/www.ibm.com/blog/chatbot-types> [consulté le 4 janvier 2024].
11. Command Line Interface, 2024 [en ligne]. Disponible à l'adresse : <https://rasa.com/docs/rasa/command-line-interface/> [consulté le 21 février 2024].
12. CONTRIBUTORS ([HTTPS://GITHUB.COM/HUGGINGFACE/TRANSFORMERS/GRAPHS/CONTRIBUTORS](https://github.com/huggingface/transformers/graphs/contributors)), The Hugging Face team (past and future) with the help of all our. *transformers: State-of-the-art Machine Learning for JAX, PyTorch and TensorFlow* [logiciel]. Version 4.36.2. [consulté le 4 janvier 2024]. Disponible à l'adresse : <https://github.com/huggingface/transformers> [consulté le 4 janvier 2024].
13. Conversational AI with Language Models, 2024 [en ligne]. Disponible à l'adresse : <https://rasa.com/docs/rasa-pro/calm/> [consulté le 28 février 2024].

14. Difference between a bot, a chatbot, a NLP chatbot and all the rest? | Botpress Blog, [en ligne]. Disponible à l'adresse : <https://botpress.com/blog/nlp-chatbot> [consulté le 4 janvier 2024].
15. DUTTA, Bhumika. What is Topic Modelling in NLP? | Analytics Steps. [en ligne]. Disponible à l'adresse : <https://www.analyticssteps.com/blogs/what-topic-modelling-nlp> [consulté le 4 janvier 2024].
16. *facebook/duckling* [logiciel] [en ligne]. 21 février 2024. Meta. [consulté le 21 février 2024]. Disponible à l'adresse : <https://github.com/facebook/duckling> [consulté le 21 février 2024].
17. Getting Started with Rasa X as an Existing Rasa User | Rasa Blog, *Rasa* [en ligne]. Disponible à l'adresse : <https://rasa.com/blog/rasa-x-getting-started-as-a-current-rasa-user/> [consulté le 3 mars 2024].
18. *google-research/tapas* [logiciel] [en ligne]. 29 février 2024. Google Research. [consulté le 3 mars 2024]. Disponible à l'adresse : <https://github.com/google-research/tapas> [consulté le 3 mars 2024].
19. HERZIG, Jonathan et al., 2020. TAPAS: Weakly Supervised Table Parsing via Pre-training. In : *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4320-4333. 2020. DOI 10.18653/v1/2020.acl-main.398. arXiv:2004.02349 [cs]
20. How AI chatbots like ChatGPT or Bard work – visual explainer | Technology | The Guardian, [en ligne]. Disponible à l'adresse : <https://www.theguardian.com/technology/ng-interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer> [consulté le 4 janvier 2024].
21. How do Chatbots work? A Guide to the Chatbot Architecture, *Maruti Techlabs* [en ligne]. Disponible à l'adresse : <https://marutitech.com/chatbots-work-guide-chatbot-architecture/> [consulté le 27 février 2024].

22. How Menu-Based Chatbots Can Streamline Your Business Processes, 2023

[en ligne]. Disponible à l'adresse :

<https://www.docomatic.ai/blog/chatbots/menu-based-chatbot/> [consulté le 4 janvier 2024].

23. IBM TECHNOLOGY, 2022. *What are Transformers (Machine Learning Model)?*

[en ligne]. 11 mars 2022. Disponible à l'adresse :

<https://www.youtube.com/watch?v=ZXiruGOCn9s> [consulté le 2 mars 2024].

24. IBM TECHNOLOGY, 2023. *Generative vs Rules-Based Chatbots* [en ligne].

5 juin 2023. Disponible à l'adresse :

https://www.youtube.com/watch?v=lZjUS_8btEo [consulté le 4 janvier 2024].

25. Introduction to Rasa Open Source & Rasa Pro, 2024 [en ligne]. Disponible à

l'adresse : <https://rasa.com/docs/rasa/> [consulté le 21 février 2024].

26. JAISWAL, Abhishek, 2022. NLP Tutorials Part -I from Basics to Advance.

Analytics Vidhya [en ligne]. 13 janvier 2022. Disponible à l'adresse :

<https://www.analyticsvidhya.com/blog/2022/01/nlp-tutorials-part-i-from-basics-to-advance/> [consulté le 4 janvier 2024].

27. JAY ALAMMAR, 2021. *Language Processing with BERT: The 3 Minute Intro*

(Deep learning for NLP) [en ligne]. 15 juin 2021. Disponible à l'adresse :

<https://www.youtube.com/watch?v=ioGry-89gqE> [consulté le 2 mars 2024].

28. KAKODKAR, Saish, 2022. Different Types of Chatbots. *ClaySys Technologies*

[en ligne]. 31 octobre 2022. Disponible à l'adresse :

<https://www.claysys.com/blog/types-of-chatbots/> [consulté le 4 janvier 2024].

29. KAVLAKOGLU, Eda, 2020. NLP vs. NLU vs. NLG: the differences between three natural language processing concepts. *IBM Blog* [en ligne].

12 novembre 2020. Disponible à l'adresse : <https://www.ibm.com/blog/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing->

- concepts/www.ibm.com/blog/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts [consulté le 28 février 2024].
30. KHANNA, Chetna, 2021. Text pre-processing: Stop words removal using different libraries. *Medium* [en ligne]. 10 février 2021. Disponible à l'adresse : <https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a> [consulté le 4 janvier 2024].
31. KLARA | Logiciels de gestion pour les PME, [en ligne]. Disponible à l'adresse : <https://www.klara.ch/fr-ch/> [consulté le 28 février 2024].
32. KUMARI, Kajal, 2023. Building Language Models: A Step-by-Step BERT Implementation Guide. *Analytics Vidhya* [en ligne]. 29 juin 2023. Disponible à l'adresse : <https://www.analyticsvidhya.com/blog/2023/06/step-by-step-bert-implementation-guide/> [consulté le 4 janvier 2024].
33. Lemmatization, *Engati* [en ligne]. Disponible à l'adresse : <https://www.engati.com/glossary/lemmatization> [consulté le 4 janvier 2024].
34. Meetme - Simplify Your Appointment Scheduling, 2023 [en ligne]. Disponible à l'adresse : <https://www.meetme.io/en> [consulté le 28 février 2024].
35. MEHTA, Ansh et al., 2021. SPELL CORRECTION AND SUGGESTION USING LEVENSHTAIN DISTANCE. . Vol. 08, no 09.
36. Models & Languages · spaCy Usage Documentation, *Models & Languages* [en ligne]. Disponible à l'adresse : <https://spacy.io/usage/models> [consulté le 21 février 2024].
37. MUNNANGI, Manikanta, 2019. A Comprehensive Guide To NLP. *Medium* [en ligne]. 22 novembre 2019. Disponible à l'adresse : <https://towardsdatascience.com/nlp-with-spacy-part-1-beginner-guide-to-nlp-4b9460652994> [consulté le 1 mars 2024].

38. Named Entity Recognition, 2021 *GeeksforGeeks* [en ligne]. Disponible à l'adresse : <https://www.geeksforgeeks.org/named-entity-recognition/> [consulté le 4 janvier 2024].
39. Natural Language Processing (NLP) Tutorial, 2023 *GeeksforGeeks* [en ligne]. Disponible à l'adresse : <https://www.geeksforgeeks.org/natural-language-processing-nlp-tutorial/> [consulté le 4 janvier 2024].
40. Natural Language Processing (NLP): What Is It & How Does it Work?, *MonkeyLearn* [en ligne]. Disponible à l'adresse : <https://monkeylearn.com/natural-language-processing/> [consulté le 4 janvier 2024].
41. NLTK: A Beginners Hands-on Guide to Natural Language Processing, [en ligne]. Disponible à l'adresse : <https://www.analyticsvidhya.com/blog/2021/07/nltk-a-beginners-hands-on-guide-to-natural-language-processing/> [consulté le 4 janvier 2024].
42. NLTK :: Natural Language Toolkit, [en ligne]. Disponible à l'adresse : <https://www.nltk.org/> [consulté le 11 février 2024].
43. PIRGE, Gursev, 2023a. TAPAS — Question Answering from Tables in Spark NLP. *John Snow Labs* [en ligne]. 26 avril 2023. Disponible à l'adresse : <https://www.johnsnowlabs.com/tapas-question-answering-from-tables-in-spark-nlp/> [consulté le 4 janvier 2024].
44. PIRGE, Gursev, 2023b. TAPAS — Question Answering from Tables. *John Snow Labs* [en ligne]. 2 mai 2023. Disponible à l'adresse : <https://medium.com/john-snow-labs/tapas-question-answering-from-tables-1796a7b90f49> [consulté le 4 janvier 2024].
45. Prise de rendez-vous en ligne avec une application professionnel, [en ligne]. Disponible à l'adresse : <https://agenda.ch> [consulté le 28 février 2024].

46. JABLONSKI, Joanna. Natural Language Processing With Python's NLTK Package – Real Python. [en ligne]. Disponible à l'adresse :
<https://realpython.com/nltk-nlp-python/> [consulté le 4 janvier 2024].
47. RAJ, Nikhil, 2021. Starters Guide to Sentiment Analysis using Natural Language Processing. *Analytics Vidhya* [en ligne]. 15 juin 2021. Disponible à l'adresse :
<https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/>
[consulté le 4 janvier 2024].
48. RASA, 2021. *Conversational AI with Rasa: Custom Actions* [en ligne].
7 décembre 2021. Disponible à l'adresse :
<https://www.youtube.com/watch?v=VcbfcsjBBIg> [consulté le 21 février 2024].
49. Rasa Documentation, [en ligne]. Disponible à l'adresse :
<https://rasa.com/docs/rasa/pages/http-api/> [consulté le 21 février 2024].
50. Rasa Platform: Rasa Pro, 2022 *Rasa* [en ligne]. Disponible à l'adresse :
<https://rasa.com/product/rasa-pro/> [consulté le 4 mars 2024].
51. RasaRestaurantBot/README.md at
master · kothiyayogesh/RasaRestaurantBot, *GitHub* [en ligne]. Disponible à
l'adresse :
[https://github.com/kothiyayogesh/RasaRestaurantBot/blob/master/README.](https://github.com/kothiyayogesh/RasaRestaurantBot/blob/master/README.md)
md [consulté le 4 janvier 2024].
52. REDDY, Ajay Vardhan, 2021. Getting started with NLP using NLTK Library. *Analytics Vidhya* [en ligne]. 30 juillet 2021. Disponible à l'adresse :
[https://www.analyticsvidhya.com/blog/2021/07/getting-started-with-nlp-using-](https://www.analyticsvidhya.com/blog/2021/07/getting-started-with-nlp-using-nltk-library/)
nltk-library/ [consulté le 11 février 2024].
53. REITER, Ehud et DALE, Robert, 1997. Building applied natural language
generation systems. *Natural Language Engineering*. Vol. 3, no 1, pp. 57-87.
DOI 10.1017/S1351324997001502.

54. RODRIGUEZ, Jesus, 2020. Google TAPAS is a BERT-Based Model to Query Tabular Data Using Natural Language. *DataSeries* [en ligne]. 7 septembre 2020. Disponible à l'adresse : <https://medium.com/dataseries/google-tapas-is-a-bert-based-model-to-query-tabular-data-using-natural-language-2435a386b43f> [consulté le 4 janvier 2024].
55. SAEED, Mehreen, 2022. A Gentle Introduction to Positional Encoding in Transformer Models, Part 1. *MachineLearningMastery.com* [en ligne]. 19 septembre 2022. Disponible à l'adresse : <https://machinelearningmastery.com/a-gentle-introduction-to-positional-encoding-in-transformer-models-part-1/> [consulté le 6 février 2024].
56. SANAD, Mohdsanadzakirizvi@gmail com, 2019. Demystifying BERT: A Comprehensive Guide to the Groundbreaking NLP Framework. *Analytics Vidhya* [en ligne]. 25 septembre 2019. Disponible à l'adresse : <https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/> [consulté le 4 janvier 2024].
57. SHILEDARBAXI, Nikita, 2021. Guide To TAPAS (Table PArSing) - A Technique To Retrieve Information From Tabular Data Using NLP. *Analytics India Magazine* [en ligne]. 13 février 2021. Disponible à l'adresse : <https://analyticsindiamag.com/guide-to-tapas-table-parsing-a-technique-to-retrieve-information-from-tabular-data-using-nlp/> [consulté le 4 janvier 2024].
58. SHRIDHAR, Kumar, 2020. Natural Language Understanding for Chatbots. *NeuralSpace* [en ligne]. 26 décembre 2020. Disponible à l'adresse : <https://medium.com/neuralspace/natural-language-understanding-for-chatbots-2eb7a81b9390> [consulté le 3 mars 2024].

59. Software - The Stanford Natural Language Processing Group, [en ligne].
Disponible à l'adresse : <https://nlp.stanford.edu/software/> [consulté le 8 février 2024].
60. Stages of Natural Language Processing (NLP) | byteiota, 2020 [en ligne].
Disponible à l'adresse : <https://byteiota.com/stages-of-nlp/> [consulté le 1 mars 2024].
61. Stemming, *Engati* [en ligne]. Disponible à l'adresse :
<https://www.engati.com/glossary/stemming> [consulté le 4 janvier 2024].
62. STOY, Lazarina, 2023. Five phases of NLP and how to incorporate them into your SEO journey. *Oncrawl - Technical SEO Data* [en ligne]. 28 février 2023.
Disponible à l'adresse : <https://www.oncrawl.com/technical-seo/five-phases-nlp-how-incorporate-them-into-seo-journey/> [consulté le 1 mars 2024].
63. TAPAS, [en ligne]. Disponible à l'adresse :
https://huggingface.co/docs/transformers/model_doc/tapas [consulté le 4 janvier 2024].
64. The 3 Kinds of Chatbots You'll Meet | OvationCXM, [en ligne]. Disponible à l'adresse : <https://www.ovationcxm.com/blog/3-kinds-chatbots-youll-meet> [consulté le 4 janvier 2024].
65. The Stanford Natural Language Processing Group, [en ligne]. Disponible à l'adresse : <https://nlp.stanford.edu/software/tagger.shtml> [consulté le 5 mars 2024 a].
66. The Stanford Natural Language Processing Group, [en ligne]. Disponible à l'adresse : <https://nlp.stanford.edu/software/tagger.shtml> [consulté le 11 février 2024 b].

67. The Ultimate Guide to Transformer Deep Learning, [en ligne]. Disponible à l'adresse : <https://www.turing.com/kb/brief-introduction-to-transformers-and-their-power> [consulté le 2 mars 2024].
68. Understanding Part-of-Speech Tagging in NLP: Techniques and Applications - Shiksha Online, [en ligne]. Disponible à l'adresse : <https://www.shiksha.com/online-courses/articles/pos-tagging-in-nlp/> [consulté le 4 janvier 2024].
69. Using Neural Networks to Find Answers in Tables, 2020 [en ligne]. Disponible à l'adresse : <https://blog.research.google/2020/04/using-neural-networks-to-find-answers.html> [consulté le 4 janvier 2024].
70. VERSLOOT, Christian. Machine Learning for Table Parsing: TAPAS. *GitHub* [en ligne]. Disponible à l'adresse : <https://github.com/christianversloot/machine-learning-articles/blob/main/easy-table-parsing-with-tapas-machine-learning-and-huggingface-transformers.md> [consulté le 5 mars 2024].
71. What Is a Chatbot? | IBM, [en ligne]. Disponible à l'adresse : <https://www.ibm.com/topics/chatbots> [consulté le 27 février 2024].
72. What is BERT | BERT For Text Classification, [en ligne]. Disponible à l'adresse : <https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/> [consulté le 4 janvier 2024].
73. What is BERT (Language Model) and How Does It Work?, *Enterprise AI* [en ligne]. Disponible à l'adresse : <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model> [consulté le 4 janvier 2024].
74. What is Named Entity Recognition (NER)? Methods, Use Cases, and Challenges, [en ligne]. Disponible à l'adresse : <https://www.datacamp.com/blog/what-is-named-entity-recognition-ner> [consulté le 4 janvier 2024].

75. What Is Natural Language Processing, 2020 *MonkeyLearn Blog* [en ligne].
Disponible à l'adresse : <https://monkeylearn.com/blog/what-is-natural-language-processing/> [consulté le 4 janvier 2024].
76. What is Natural Language Understanding (NLU)? | Definition from TechTarget, *Enterprise AI* [en ligne]. Disponible à l'adresse :
<https://www.techtarget.com/searchenterpriseai/definition/natural-language-understanding-NLU> [consulté le 1 mars 2024].
77. What Is Natural Language Understanding (NLU)?, 2022 [en ligne]. Disponible à l'adresse : <https://thelevel.ai/blog/natural-language-understanding/> [consulté le 26 février 2024].
78. WUTTKE, Laurenz, 2023. NLP vs. NLU vs. NLG: Unterschiede, Funktionen und Beispiele. *datasolut GmbH* [en ligne]. 24 mai 2023. Disponible à l'adresse :
<https://datasolut.com/natural-language-processing-vs-nlu-vs-nlg-unterschiede-funktionen-und-beispiele/> [consulté le 1 mars 2024].
79. YANNIC KILCHER, 2020. *TAPAS: Weakly Supervised Table Parsing via Pre-training (Paper Explained)* [en ligne]. 5 mai 2020. Disponible à l'adresse :
<https://www.youtube.com/watch?v=cIUtRNhY6Rw> [consulté le 6 janvier 2024].
80. YOG, 2021. *kothiyayogesh/RasaRestaurantBot* [logiciel] [en ligne].
14 septembre 2021. [consulté le 4 janvier 2024]. Disponible à l'adresse :
<https://github.com/kothiyayogesh/RasaRestaurantBot> [consulté le 4 janvier 2024].