MINISTERE DE L'ENSEIGNEMENT SUPERIEURET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE SAAD DAHLEB – BLIDA 01

FACULTE DES SCIENCES

DEPARTEMENT D'INFORMATIQUE

MEMOIRE DE MASTER II

Spécialité : Ingénierie des logiciels

THEME

Génération automatique de questions en arabe à l'aide de systèmes de réponses aux questions

Présenté par :

- ✓ **Rahma Amira**
- ✓ **Bahri Lylia**

Proposé et Encadré par :

- ✓ **Mme. Ouahrani Leila**

Soutenu : 24/09/2022

Devant le jury Composé de

Mme. MESKALDJI Présidente
Mr. BENAISSI Examinateur

Année Universitaire : 2021/2022

Remerciements

Nous remercions tout d'abord dieu, pour la force et la santé qu'il nous a donné pour pouvoir terminer ce mémoire.

Nous remercions notre promotrice **Mme OUAHRANI Leila**, pour son encadrement professionnel, sa présence, ses conseils et ses remarques en or, sans les quels ce travail n'aurait pas pu avoir le jour et ne serait pas aussi riche.

Nous tenons à remercier les membres du jury pour nous avoir consacré de leur temps pour la correction de notre mémoire.

Nous remercions ainsi, nos chers parents pour leur amour et leur soutien qui a toujours été présents depuis notre premier jour de nos études. Que dieu vous garde pour nous.

Nous remercions l'ensemble des enseignants qui nous ont permis d'acquérir autant d'informations durant toutes ces années pour pouvoir arriver à ce point.

Nous adressons à la fin nos remerciements à tous les membres des familles **RAHMA** et **TAHRAOUI**, **BAHRI** et **SAADAOUI.** À nos chères sœurs et frères. Ainsi que nos chers amis. Et à toute personne qui a contribué de prés et de loin à la réussite de ce travail. Un merci spécial à **BENHAMIDA Abdennour** pour ses supports tout au long de l'année.

ABSTRACT

Automatic question generation is a task of Natural language processing on which researchers aim to achieve better results. We are interested in our work on realizing an automatic question generation system using question answering systems for the Arabic language in order to enrich the research of existing works that deal with this language. We aim to improve the user's query or the quality of a given question by suggesting better ones. The realization of our work is based on a data-driven approach using deep learning techniques, more precisely the Encoder/Decoder architecture with attention mechanism.

Our model was also trained on an English dataset in order to compare the obtained results with those of the existing models. The two chosen datasets are of a community type which allows us to satisfy the objective of our work.

Keywords: Automatic question generation, Question answering systems, Natural language processing, Attention mechanism, Deep learning.

Résumé

La génération automatique des questions est une tâche de traitement automatique du langage naturel sur lequel les chercheurs visent toujours à atteindre de meilleurs résultats.

Nous nous intéressons dans notre travail à réaliser un système de génération automatique des questions pour la langue arabe afin d'enrichir les travaux qui traitent cette dernière. Notre objectif est d'améliorer la requête de l'utilisateur ou la qualité de la question en suggérant des questions meilleures.

La réalisation de notre système se base sur une approche axée sur les données en utilisant les techniques de l'apprentissage profond, bien précisément l'architecture encodeur/décodeur avec mécanisme d'attention.

Notre modèle à été également entraîné sur un ensemble de données en langue anglaise pour pouvoir comparer les résultats obtenus avec les autres systèmes existants. Les deux ensembles de données que nous avons choisi sont de type communautaire, ce qui nous permet de satisfaire le besoin de notre travail.

Mots clés : Génération automatique des questions, Systèmes de réponses aux questions, traitement automatique du langage naturel, mécanisme d'attention, apprentissage profond.

الملخص

يعد التوليد التلقائي للأسئلة أحد مهام معالجة اللغة الطبيعية، والتي يهدف الباحثون فيها إلى تحقيق نتائج أفضل.

نهتم في عملنا على تحقيق نظام آلي لتوليد الأسئلة باستخدام أنظمة الإجابة على الأسئلة باللغة العربية، وذلك من أجل إثراء البحث والأعمال التي تتعامل مع هذه اللغة. يهدف هذا النظام إلى تحسين استعلام المستخدم أو تحسين جودة سؤال معين من خلال اقتراح أسئلة أفضل.

يعتمد تحقيق عملنا على نهج يعتمد على البيانات باستخدام تقنيات التعلم العميق وبصورة أدق يعتمد على تقنية التشفير وفك التشفير مع آلية الانتباه.

تم تدريب نموذجنا أيضا على مجموعة بيانات باللغة الإنجليزية وذلك لمقارنة النتائج التي تم الحصول عليها مع تلك الخاصة بالنماذج الحالية.

كلى المجموعتان المختارتان هما من نوع "المجتمع" وهذا يسمح لنا بتحقيق هدف عملنا.

الكلمات المفتاحية: التوليد التلقائي للأسئلة، معالجة اللغة الطبيعية، آلية الانتباه، التعلم العميق، أنظمة الإجابة على الأسئلة

Table des figures

Liste des tableaux

Glossaire de termes

Abréviation	Signification
AQG	Génération Automatique des Questions (Automatic Question Generation)
GRU	Unité récurrente fermée (Gated Recurrent Unit)
MSA	Arabe Moderne Standard (Modern Standard Arabic)
BOW	Sac de mots (Bag Of Words)
AI	Intelligence Artificielle (Artificial intelligence)
IR	Récupération d'information (Information Retrieval)
LSTM	Longue mémoire à court terme (Long Short-TermMemory)
NER	Reconnaissance des entités nommées (Named EntityRecognition)
POS	Partie de discours (Part Of Speech)
QA	Réponse aux questions (Question answering)
QAS	Système de réponse aux question (Question AnsweringSystem)
RNN	Les réseaux de neurones récurrents (Recurrent NeuralNetworks)
SRL	Étiquetage sémantique de rôle (Semantic Role Labeling)
NLP	Traitement Automatique du Langage naturel (Natural Language Processing)
WE	Plongement lexical (Word Embedding)
WER	Taux d'erreur de mots (Word Rate error)

Tables des matières

Introduction Générale

De nos jours, lorsqu'on a besoin d'acquérir plus d'informations sur un sujet, l'un des moyens les plus pratiques qui vient à l'esprit est l'utilisation d'Internet. La quantité des données est si importante qu'elle devient une source d'information très populaire. Le processus de recherche d'information est simple et direct avec l'utilisation des moteurs de recherche tels que 'Google', 'Bing'. Après la saisie d'une requête dans un moteur de recherche, on se trouve face à des milliers de liens fournis qui correspondent à la requête. Cependant, certains de ces liens peuvent être proches de la requête mais ne fournissent pas les informations exactes requises pour fournir une réponse.

La deuxième étape pour trouver la réponse à une requête consiste à trouver la réponse souhaitée, pour cela, l'utilisateur doit parcourir un certain nombre de ces liens. Le processus peut rapidement devenir frustrant si l'utilisateur n'a pas saisi une requête correcte et obtient de nombreux résultats indésirables. De plus, parcourir chacun de ces liens peut prendre beaucoup de temps.

Parfois, les utilisateurs ne peuvent pas dire précisément ce qu'ils pensent donc les résultats qu'ils obtiennent peuvent différer de leur intention. La saisie d'un requête correcte et informative est indispensable et assez importante pour obtenir les résultats souhaités.

Une solution à ce problème consiste à fournir des exemples de questions associées à la requête de l'utilisateur. L'utilisateur peut soit en sélectionner une parmi ces suggestions, soit, à l'aide des suggestions, proposer une requête plus précise, et donc plus proche de ses intentions. Et c'est la motivation principale de notre travail.

Un système de génération de question automatique doit être capable de produire des questions naturelles à partir d'une phrase ou un paragraphe. Et donc suggérer des exemples de questions connexes aux utilisateurs lorsqu'ils saisissent une requête. Ce genre de système peut être considéré comme indispensable dans le domaine d'éducation, car les questions sont le moyen dont un professeur teste les connaissances des étudiants, donc avec un tel système en peut diminuer le temps et l'effort fourni pour produire des questions manuellement car la construction des questions est difficile et complexe et demande de l'expérience. Mais avec ces questions générées automatiquement nous aurons avec une grande variété des questions qui

elles même peuvent être utilisé comme ressources pour d'autres systèmes d'apprentissage automatique.

Dans ce travail, nous nous intéressons à la génération des questions en langue arabe, la langue propose pleins de difficultés avec sa complexité grammaticale, le manque des ressources et outils par rapport à d'autres langues telle que l'anglais.

Généralement, ce que les gens saisissent comme requête est leur discours informel qui peut ne pas suivre la structure grammaticale correcte. D'où la motivation derrière le développement d'un système de génération de questions pour de telles données. Nous avons décidé d'utiliser des systèmes communautaires de questions-réponses (CQA). À notre avis, ces systèmes sont les plus proches en ce qui concerne la structure de ce que les gens saisissent comme requête.

Objectifs

Notre objectif principal est de construire un système de générations automatique des questions dont il fournit des exemples de la requête (question) initiale de l'utilisateur dans un moteur de recherche. Donc le but est d'aider les utilisateurs en fournissant des suggestions de questions liées à leur requête.

Notre approche est complétement data-driven (axée sur les données). Elle prend en entrée une réponse d'un système de réponses aux questions. Vu la nature de problème, nous explorons les réseaux neuronaux séquence-à-séquence pour construire notre système.

Les objectifs qui doivent être atteint durant ce travail sont les suivants :

- Explorer les approches de génération de questions et celles des systèmes de réponses aux questions.
- Explorer les modèles d'apprentissage profond « séquence à séquence » pour en adapter à la génération de questions (Transformer, Encodeur-décodeur, Lstm, Bi-Lstm, …)
- Acquisition & collecte du dataset d'entrainement (qui doit être issu d'un système de réponses aux questions communautaire).
- Concevoir et implémenter le modèle de génération de questions
- Evaluer la qualité du modèle de génération des questions (Evaluation quantitative par des métriques automatiques & une évaluation humaine par une expertise humaine).

Structure de mémoire

Afin d'atteindre nos objectifs fixés, nous structurons notre mémoire en 5 chapitres distincts :

- Chapitre 1 : nous présentons les concepts fondamentaux du Deep Learning liés à notre problématique.
- Chapitre 2 : nous présentons un état de l'art sur les travaux et approches existants traitant la génération automatique des questions et ainsi des systèmes de réponses aux questions.
- Chapitre 3 : vise à présenter les modèles basés sur le deep learning réalisés pour résoudre la problématique de génération des questions en langue arabe.
- Chapitre 4 : porte sur les performances de notre outil ainsi que les résultats de l'évaluation automatique et humaine
- Chapitre 5 : nous terminons par une conclusion mettant le point sur le travail réalisé ainsi que des perspectives futures.

Chapitre I : Concepts fondamentaux de Deep Learning

1. Introduction

De nos jours, l'Intelligence Artificielle (IA) est devenue de plus en plus importante et présente dans notre vie quotidienne et dans plusieurs domaines tels que (traitement du langage naturel, traitement d'image, détection de fraudes, systèmes décisionnels, services client virtuel, etc.). Et avec l'apparition du deep learning les avantages de son utilisation sont de plus en plus nombreux.

L'apprentissage profond (deep learning) est un sous-domaine de l'intelligence artificielle, son fonctionnement se base sur l'utilisation des réseaux de neurones. Dans ce chapitre nous allonsaborder les différents concepts de l'apprentissage profond

2. Apprentissage profond

L'apprentissage profond, aussi appelé "apprentissage hiérarchique" est une sous-catégorie de l'IA, qui est un ensemble de techniques permettant à une machine de s'inspirer du comportement humain [1].

L'apprentissage profond est aussi une approche de l'apprentissage automatique, qui est un ensemble d'algorithmes qui apprennent et améliorent leurs performances par les données qu'ils traitent [2]. Il se base sur l'utilisation des algorithmes d'apprentissage en plusieurs niveaux de représentations qui permet de modéliser des relations complexes entre les données [3].

Ce concept est géré par les réseaux de neurones et permet à une machine d'apprendre de manière autonome.

Il s'agit de 4 différents modes d'apprentissage qui sont :

- **Apprentissage supervisé** : qui s'appuie sur les données étiquetées. Ou les modèles de séquences sont utilisés tels que (encodeur/décodeur, Transformer) qui vont être détaillés dans une prochaine étape.
- **Apprentissage non supervisé** : contrairement à l'apprentissage supervisé les données dans ce cas ne sont pas étiquetées.
- **Apprentissage par transfert :** c'est l'utilisation des outils et des techniques qui permettent le transfert de connaissances des tâches sources vers des tâches cibles.

- Apprentissage par renforcement : qui est inspiré par l'apprentissage humain et qui permet à un agent d'apprendre à choisir quelle action et ceci à partir des expériences.

3. Les réseaux de neurones

Un réseau de neurones est une succession de couches composées de neurones, où le neurone est une fonction mathématique et les sorties d'une couche sont les entrées de la couche qui suit.

Les réseaux de neurones ont aujourd'hui un impact remarquable dans toutes sortes d'applications et dans différents domaines (robotique, télécommunication, systèmes pour la synthèse de la parole, traitement de signal, etc.).

Les réseaux de neurones existent en différents types :

- Réseaux de neurones de base (feed-forward en anglais**) :** dans ce type de réseaux de neurones l'information passe directement de l'entrée aux nœuds de traitement. De là, l'information est acheminée à la sortie sans retour en arrière [4]. Disposant une ou plusieurs couches cachées. La figure 1 suivante représente l'architecture d'un réseau de neurones.

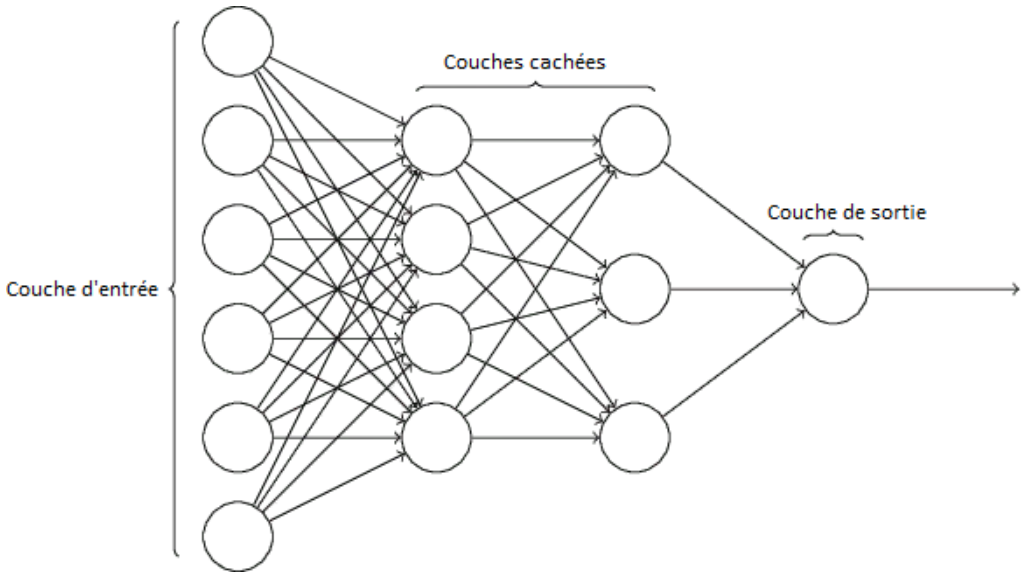


Figure 1: Architecture d'un réseau de neurones [5]

Le deuxième type de réseaux de neurones c'est les réseaux de neurones récurrents.

-Réseaux de neurones récurrents (Recurrent Neural Networks)

Contrairement aux réseaux de base, l'information dans ce type de réseaux est traitée en cycle, c'est-à-dire l'information peut toujours revenir en arrière avant d'atteindre la couche "sortie". Ce type de réseau est utilisé pour le traitement des données séquentielles, lors de la génération des sorties. Ils prennent en considération la relation entre les séquences et la notion de temps, autrement dit, il prend non seulement l'entré à l'état "T" mais aussi la valeur de l'état "T-1". La figure suivante montre l'architecture d'un réseau de neurones.

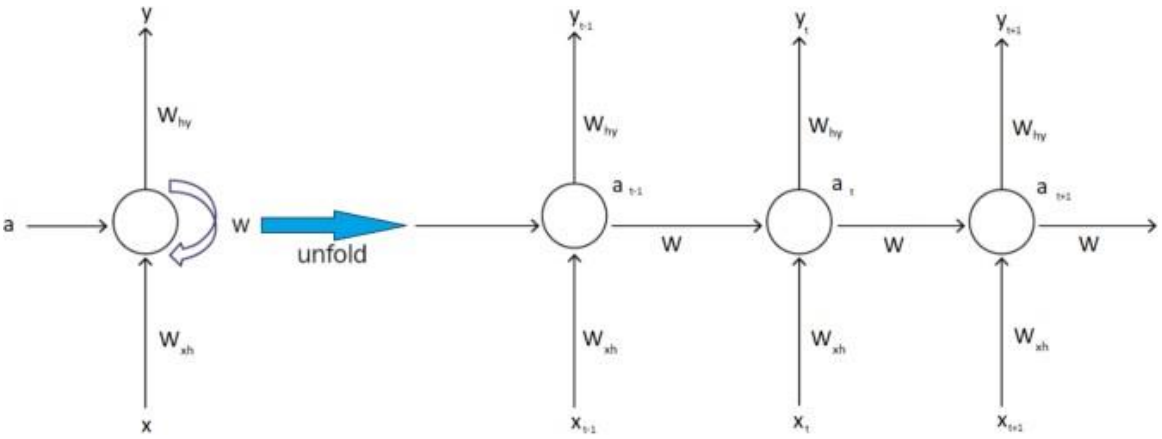


Figure 2: Représentation d'un RNN [6]

Un réseau de neurones récurrent peut prendre plusieurs architectures selon son application. La figure -3- montre les différentes architectures d'un RNN :

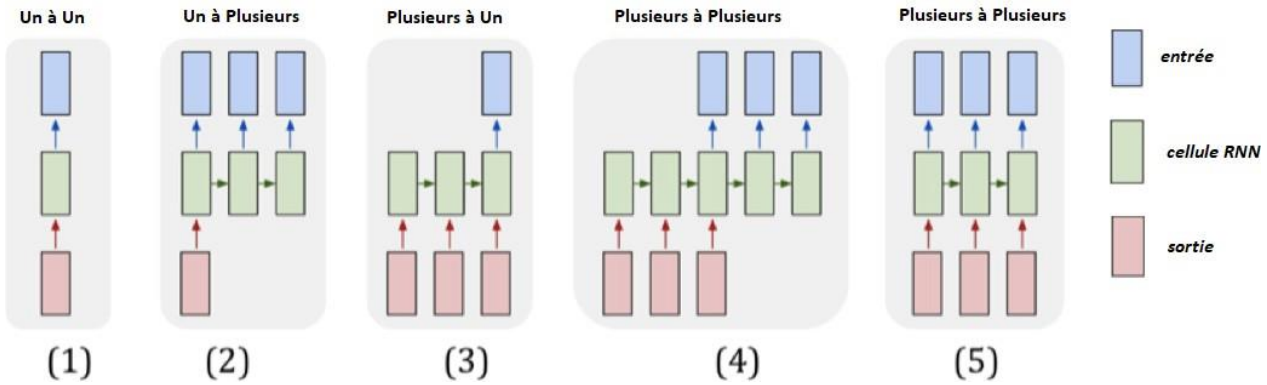


Figure 3: Les différentes architectures d'un RNN [7]

Exemples d'utilisations des différentes architectures d'un RNN :

- Un à un (Classification d'une image)

- Un à plusieurs (Générer une phrase à partir d'un mot en entrée)
- Plusieurs à un (Analyse des sentiments d'après des messages)
- Plusieurs à plusieurs (Traduction d'un texte)

3.1 Long Short-Term Memory (LSTM)

Ces réseaux ont été inventés pour régler la limitation des RNNs qui est la mémoire courte et pour résoudre le problème de disparition de gradient ceci grâce au mécanisme de portes. Ce genre de réseaux a été initialement introduit en 1997 par Hochreiter & Schmidhuber [8]. Ils ont été développés pour maintenir une ancienne information pendant une longue durée surtout quand la séquence est très longue est ceci grâce à la cellule mémoire. Une représentation détaillée de la cellule LSTM est illustrée dans la figure 4 :

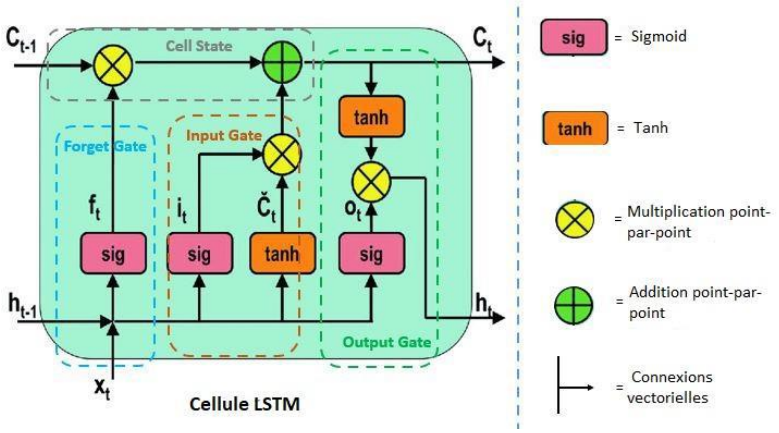


Figure 4: Architecture d'une cellule LSTM [9]

Chaque cellule LSTM contient les portes suivantes :

- **Porte d'oubli (Forget Gate) :**

Cette porte sert à décider quelle information doit avoir plus d'attention et laquelle peut être ignorée. Une information de l'état caché précédent $h_{(t-1)}$ et une information de l'état actuelle x_t passent par la fonction $sigmoid$, les valeurs près de 0 sont à oubliés et les valeurs près de 1 sont à retenir. La figure 5 montre l'opération de la porte d'oubli.

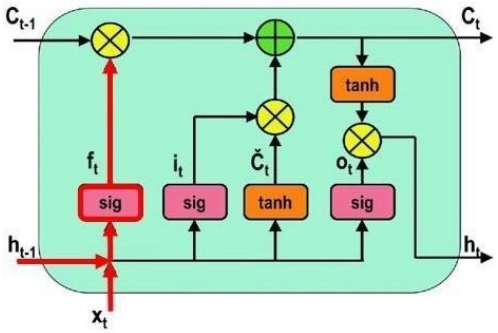


Figure 5: Opération de la porte d'oubli [9]

Avec :

$f_t = \sigma\,(W_f \cdot [ht_{-1}\,;\, x_t] + b_f$

t : le pas de temps.

f_t: la porte d'oubli à l'instant t.

x_t : l'entrée

h_{t-1} : l'état caché précédent.

W_f : la matrice de poids entre la porte d'oubli et la porte d'entrée.

b_t: biais de connexions à l'instant t.

- **Porte d'entrée (Input Gate) :**

Cette porte sert à mettre à jour la cellule d'état (Cell State) en faisant passer en premier l'état caché précédent h_{t-1} et l'entré en état actuelle x_t par la fonction Sigmoïd qui décide quelles valeurs seront mises à jour en les transformant entre 0 (pas important) et 1 (important).

Par la suite les mêmes informations $h_{(t-1)}$ et x_t passent par la fonction $tanh$ où un vecteur sera créé contenant toutes les valeurs possibles entre -1 et 1. Les sorties fournis par la fonction Sigmoïd et $tanh$ passent par la multiplication point-par-point.

$i_t = \sigma\,(W_i \cdot [h_{t-1}\,;\, x_t] + b_i$

$\check{C}t = tanh\,(Wc \cdot [h_{t-1}\,;\, x_t] + b_c$

W_i = matrice de poids de l'opérateur sigmoid entre la porte d'entrée et la porte de sortie.

Wc = matrice de poids de l'opérateur tanh entre l'information de la cellule d'état et la sortie du réseau. La figure 6 montre l'opération de la porte d'entrée.

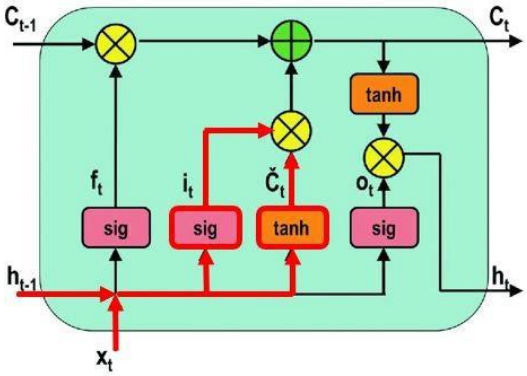


Figure 6: Opération de la porte d'entrée [9]

- **Porte de sortie (Output Gate) :**

Cette porte a pour but de déterminer la valeur du prochain état caché h_t qui est la sortie de la cellule LSTM par la fonction suivante : $h_t = O_t * \tanh(C_t)$

Avec : O_t : la sortie de la fonction sigmoid qui est égal : $Ot = \sigma(W_O [h_{t-1}; x_t] + b_O)$. La figure 7 montre l'opération de la porte de sortie.

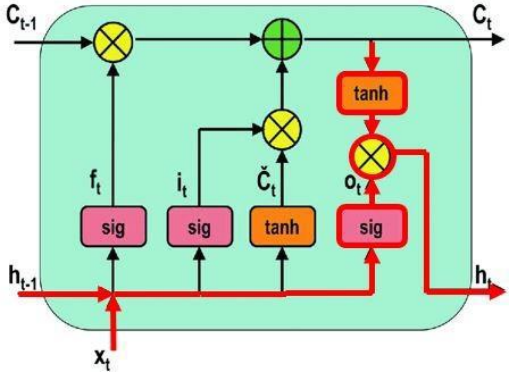


Figure 7: Opération de la porte de sortie [9]

3.2 Gated Reccurent Unit (GRU)

C'est la nouvelle génération des RNN, ils ont été proposés pour régler le problème des RNN qui souffre de la disparition ou l'explosion de gradient et ne se souvient pas des états précédents une longue durée. Ces cellules disposent de 2 portes : une porte de réinitialisation "reset-gate" et une porte de mise à jour "update-gate".

La figure 8 représente la structure d'une unité GRU :

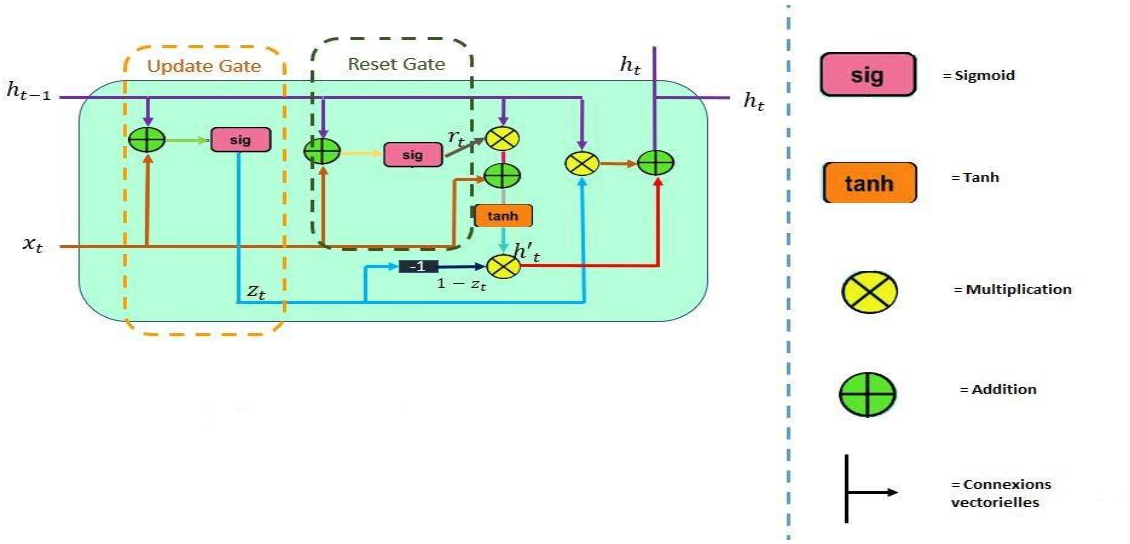


Figure 8: Structure d'une cellule GRU [9]

- **Porte de réinitialisation (Reset Gate) :**

Cette porte sert à décider combien d'informations sont passées, le réseau doit oublier. L'état caché précédent, concaténé avec les données d'entrée, passe par une fonction sigmoïde (pour ne conserver que les coordonnées pertinentes) puis est multiplié par l'ancien état caché : on n'en conserve donc que les coordonnées importantes (telles qu'elles) de l'état précédent (on a donc perdu une partie de l'état précédent dans cette porte). La figure 9 montre l'opération de la porte de réinitialisation.

Cette opération se calcule par la formule : $r_t = \sigma\,(W_r \cdot [h_{t-1}, x_t] + b_r$, d'où :

t = le pas de temps.

r_t = la porte de réinitialisation à l'instant t.

x_t =le vecteur d'entrée.

h_{t-1} = l'état caché précédent.

Wr = la matrice de poids.

b_t = le biais de connexion à l'instant t.

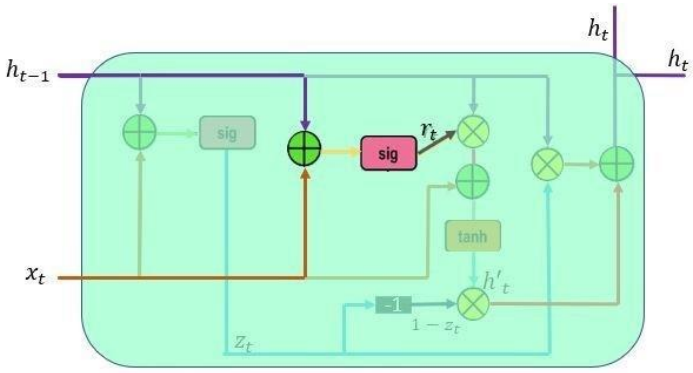


Figure 9: Opération de la porte de réinitialisation [9]

➢ **Porte de mise à jour (Update gate) :**

Elle sert à décider quelles informations sont à conserver et quelles informations à oublier. Les données d'entrées et l'ancien état caché sont concaténés et passent par une fonction sigmoïde dont le rôle est de déterminer quelles sont les composantes importantes.

$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z$

Avec : $z_t =$ la porte de mise à jour à l'instant t

La figure 10 montre l'opération de la porte de mise à jour.

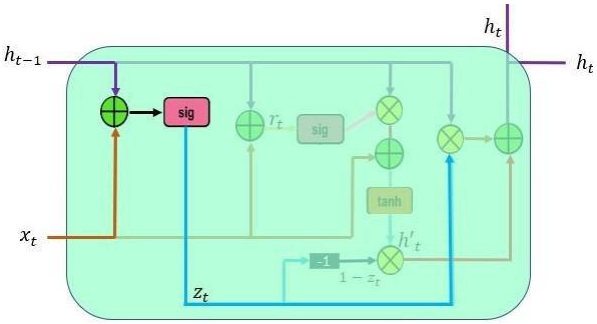


Figure 10: Opération de la porte mise à jour [9]

4. Les fonctions d'activation

Une fonction d'activation est une fonction mathématique qui décide l'état d'un neurone (s'il doit être activé ou non) par le calcul de la somme pondérée. Elle sert à modifier la représentation d'une donnée grâce à sa particularité non linéaire. Autrement dit, elle nous permet de changer notre manière de voir une donnée [10].

4.1 Les fonctions d'activations linéaires

C'est tout simplement une ligne droite, elle fournit en sortie un résultat proportionnel à l'entrée (somme pondérée de neurones). Elle est de la forme : $f(x) = x$

Elle a 2 majeurs problèmes :

- Il n'est pas possible d'utiliser la rétropropagation car la dérivée est une constante.

- Elle transforme tout le réseau en une seule couche car toutes les couches s'affrontent en une seule (la dernière couche sera toujours en fonction de la première). Le graphe d'une fonction linéaire est représenté dans la figure 11.

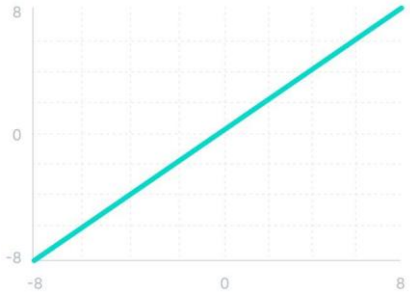


Figure 11: Graphe d'une fonction linéaire [11]

4.2 Les fonctions d'activations non-linéaire

Nous nous intéressons à ce type de fonctions car elles permettent au modèle de s'adapter facilement à une variété de données avec différents résultats. Parmi ces fonctions nous citons :

- **Sigmoid :**

La fonction la plus utilsée, surtout avec les modéles dont le résultat doit être une probabilité prédite. La valeur de sortie est comprise entre [0, 1]. Elle de la forme de l'équation 1:

$$F(x) = \frac{1}{1 + e^{-x}} \qquad (1)$$

Le graphe de la fonction Sigmoid est représenté dans la figure 12.

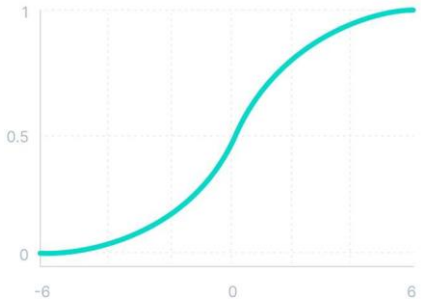


Figure 12: Graphe de la fonction Sigmoïd [11]

- **Tanh (Tangente Hyperbolique) :**

Les résultats de Tanh sont entre -1 et 1. Avec Tanh plus l'entrée est grande (positive), la valuer de sortie sera proche à 1 et vice versa. Il faut noter cici que Tanh souffrent de probléme de disparition de gradient (vanishing gradient) car dans l'étape de rétropropagtion les gradients devient de plus en plus petits ce qui résulte un apprentissage lent . Elle est de la forme de l'équation 2 :

$$F(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} = \frac{2}{1 + e^{-2x}} - 1 \qquad (2)$$

Le graphe de la fonction Tanh est représenté dans la figure 13.

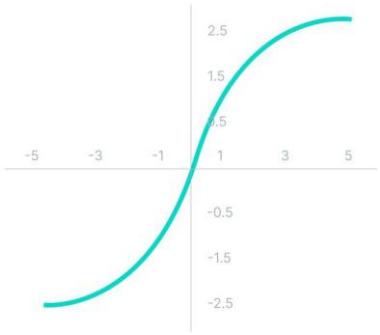
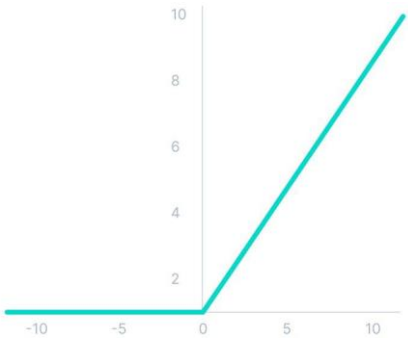


Figure 13: Graphe de la fonction Tanh [11]

- **RELU (Rectified Linear Unit) :** Son avantage c'est quelle n'active pas tous les neurons au même temps, cela veut dire que les neurons seront désactivés si le résultat de sortie est moins de 0. Cette fonction popose une solution au dispertion du gradient, mais elle conduit à l'explosion du gradient car elle peut prendre une infinité de valeurs [0 , ∞[.

  f(x) = 0 si x < 0

  f(x) = x si x >= 0



Le graphe de la fonction RELU est représenté dans la figure 14.

Figure 14: Graphe de la fonction RELU [11]

- **Softmax**

Certains ont décrit Softmax comme la combinaison de plusieurs focntions Segmoid, Elle permet de normaliser les valeurs rééls d'un vecteur à des valeurs entre [0,1]. Dans ce cas le total de la distribution de probabilité égal à 1. Softmax est de la forme de l'équation 3 :

$$S(z)j = \frac{e^{z\,j}}{\sum_{k=1}^{k} e^{z\,k}} \qquad (3)$$

Avec : j ∈ {1 ….. k}

k : nombre de classe dans un classifieur

z_j : le vecteur d'entrée avec des valeurs réelles.

5. Modèle de séquences

Les modèles de séquences servent à mapper une séquence (une suite de valeurs) d'entrée vers une séquence de sortie, les deux séquences peuvent être de même taille ou de taille

différentes. Le traitement dans un modèle de séquence diffère selon la taille des séquences, nous distinguons plusieurs solutions les RNN, GRU, LSTM, Encodeur/Décodeur, transformer, etc.

5.1 Encodeur-Décodeur

Cette architecture est composée de deux réseaux RNN dans deux parties :

- Encodeur : Une ou plusieurs couches RNN pour traiter la séquence d'entrée uniquement. Puis renvoie son état interne qui va représenter le contexte (le vecteur de contexte avec une taille fixe) qui sera transmis vers le décodeur.
- Décodeur : Une ou plusieurs couches RNN qui prennent en entrée le vecteur de contexte généré par l'encodeur.

Le rôle du décodeur est de prédire les prochains caractères de la séquence cible en lui donnant les caractères précédents.

La figure -15- représente l'architecture générale d'un encodeur décodeur basique.

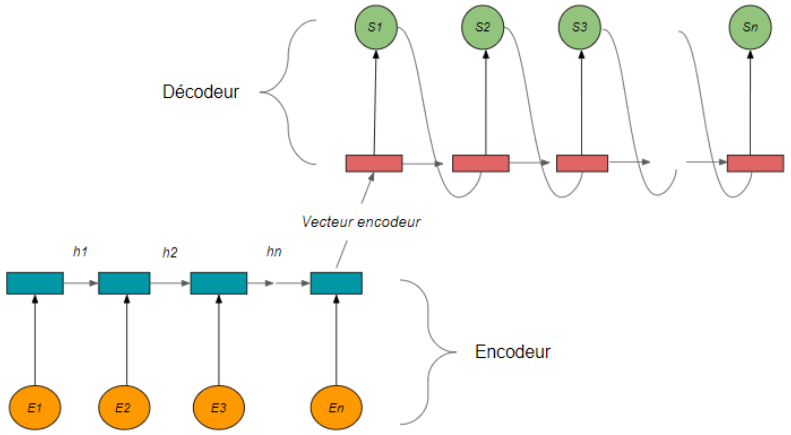


Figure 15: Architecture encodeur/décodeur [12]

Chaque cellule de l'encodeur prend un élément de la séquence passée en entrée.

Exemple : E1 → Premier mot de la séquence d'entrée.

- hi : les états cachés.
- Vecteur encodeur : le dernier état caché de l'encodeur et le premier état caché du décodeur. Il aide le décodeur à faire les prédictions et ceci en résumant toutes les informations pour tous les éléments de la séquence d'entrée.

Pour le décodeur : chaque cellule prend l'état caché précédent et produit une sortie Si et un état caché.

5.2 Transformer

Cette architecture a été introduite en 2017 dans ‘Attention is all You Need’ développé par des chercheurs de Google. C’est un modèle encodeur-décodeur qui utilise un type d’attention qui s’appelle -multi têtes. Ce type fait entrer les séquences en parallèle au lieu de les faire entrer une par une. Ensuite, les séquences passent par l'espace d’embedding où le mot se transforme en vecteur. Le vecteur d’attention (z) est calculé par la suite avec la fonction Softmax et les valeurs (Valeur, Clé, Requête), à la fin on se trouve avec plusieurs vecteurs d’attention (z) pour chaque mot. Mais pour le réseau de neurone (Feed-Forward) pour l’étape suivante, il faut un seul vecteur, donc une matrice Wz est utilisée pour que la sortie soit un vecteur d’attention pour chaque mot. Puis, la sortie finale est obtenue [13].

Pour la partie décodeur, c’est pareil, sauf qu’il existe une partie de plus c’est « Masked head-attention". Elle représente le premier bloc du décodeur et fait passer dans ce bloc les résultats du décodeur « le résultat précédent », le résultat est une Requête qui va être concaténer avec (Valeur et clé) de la sortie de l’encodeur pour passer aux étapes suivantes qui sont les mêmes que celles de l’encodeur sauf la dernière étape, c’est de passer par une fonction Softmax pour obtenir le résultat avec la plus grande probabilité. L’architecture d’un transformer est présentée dans la figure suivante [14].

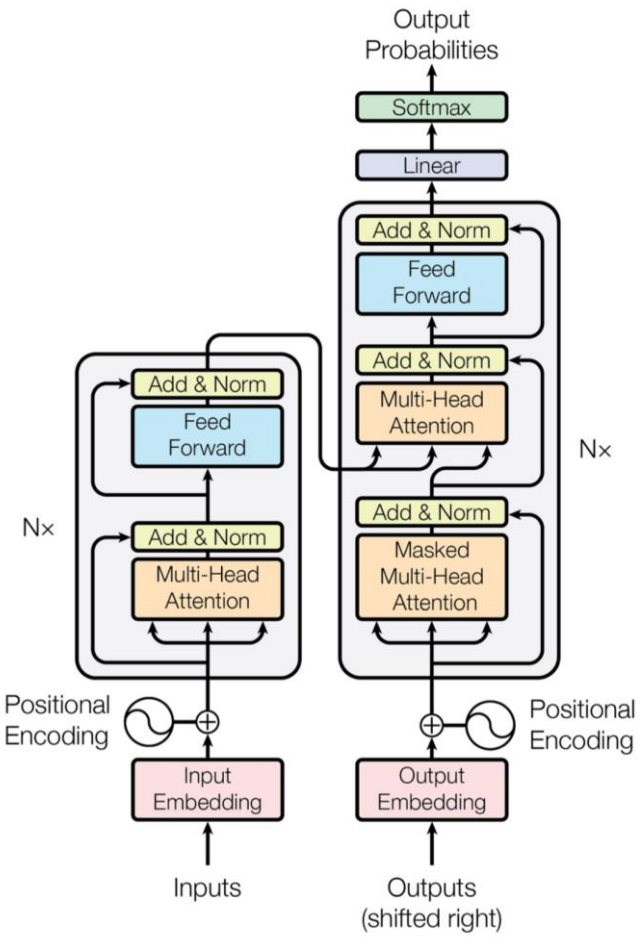


Figure 16: L'architecture d'un transformer [14]

6. Mécanisme d'attention

Le mécanisme d'attention est une technique qui vise à résoudre la limitation de l'architecture Encodeur/Décodeur qui est la difficulté de traiter une longue séquence avec un vecteur de contexte de taille petite et fixe. La solution est d'utiliser un vecteur de contexte de taille variable qui permet d'accéder à toute la séquence d'entrée au lieu du dernier état caché de l'encodeur seulement [15].

Le principe consiste à faire passer tous les états cachés de l'encodeur et l'état caché du décodeur dans une fonction pour calculer des scores qui servent à mesurer la similarité entre deux vecteurs, cela permet au modèle de se concentrer sur les parties pertinentes en calculant l'alignement entre eux. Par la suite, la fonction Softmax est utilisée pour convertir les scores en

probabilités ce qui va permettre au décodeur de savoir sur quelles parties il doit se concentrer [16].

7. Conclusion

Nous avons abordé dans ce chapitre les différents concepts de l'apprentissage profond. Nous nous intéressons dans ce travail à une tâche de traitement automatique du langage naturel, bien particulièrement à la génération automatique des questions. Où nous allons utiliser des techniques de l'apprentissage profond afin de proposer une solution. Le prochain chapitre est consacré à la génération automatique de questions, ses différentes approches, les travaux connexes les plus récents, etc.

Chapitre II : Etat de l'art

1. Introduction

L'être humain est de nature curieux et pose pleins de questions, un étudiant pose des questions pour accueillir de nouvelles connaissances. Le professeur pose des questions pour tester la compréhension de ses étudiants sur un sujet donné. Le processus de rédaction des questions de bonne qualité peut prendre beaucoup de temps et d'effort, ainsi que la partie d'évaluation de ces réponses. Pendant la pandémie en 2020 les moyens de e-learning ont été beaucoup plus utilisés, des systèmes d'évaluation automatique ont été développés ainsi que des systèmes pour la génération automatique des questions. Nous abordons dans ce chapitre le concept de génération automatique des questions, ses différentes approches, ainsi que les revues de la littérature et les travaux connexes les plus récents.

2. Traitement automatique du langage naturel

Le traitement automatique du langage (TAL), c'est un domaine de recherche de l'intelligence artificielle. Il se base sur des techniques de calculs pour développer des logiciels qui sont automatiquement capables de traiter des données qui représentent le langage humain naturel.

Il se trouve à l'intersection de l'informatique et linguistique, ce domaine est apparu en 1950 avec Alan Turing quand il a publié l'article « Computing Machinery and Intelligence ». Le TAL a plusieurs applications, par exemple : la traduction, l'indexation, l'extraction de l'information, la reconnaissance vocale, le paraphrasing, etc.

3. Défis de la langue arabe

L'Arabe représente la langue natale de plus de 450 millions de personnes, elle peut être classée en : Arabe classique (elle est complexe et avec laquelle le livre sacré du coran est écrit). Arabe moderne standard (elle contient des signes diacritiques pour simplifier la lecture pour ne pas perdre le sens du mot, AMS représente la langue officielle des pays qui utilisent l'arabe actuellement) [17].

Malgré que l'arabe est une langue très utilisée, ces ressources linguistiques ne sont pas très exploitées à cause de :

● Le manque de ressources pour la langue arabe. Très peu de datasets pour la tâche de génération automatique de questions et avec des tailles insuffisantes. Pour cette raison,

plusieurs chercheurs s'orientent vers la création de leurs propres bases de données, ce qui demandent une annotation manuelle et vérification avant son utilisation [18].

● L'extraction des noms propres d'un texte arabe est une tâche difficile et ambiguë, le système aura besoin de plus d'outils pour reconnaître les noms. Contrairement aux autres langues ou les noms commencent toujours par une majuscule et qui nécessite uniquement de la méthode NER (Named Entity Recognition) pour les extraire à partir d'un texte.

● Les symboles de dialecte et les voyelles sont considérés comme des défis en langue arabe écrite. Les signes diacritiques sont un ensemble des symboles d'orthographe qui nous aide à mieux prononcer le mot, ils sont placés au-dessus ou au-dessous de la lettre. Par contre dans les textes arabes modernes, ils sont rarement utilisés ce qui pose un problème d'ambiguïté, car un seul mot peut avoir plusieurs sens. Ça été un des défis du TAL pour avancer les recherches en langue arabe. Plusieurs développeurs ont appliqué la diactirisation automatique dans leurs systèmes comme solution à ce problème.

● Sa morphologie : La langue arabe contient 28 lettres, chacune peut prendre jusqu'à 4 formes selon sa position dans le mot. Par exemple, la lettre jim : ج se prononce [j] et peut s'écrire de trois manières :

- Au début : جـ
- Au milieu : ـجـ
- À la fin : ـج

4. Génération automatique des questions

La génération automatique des questions est une tâche de traitement automatique du langage naturel. Ce processus est apparu en 1966, il sert à générer des questions de n'importe quel type à partir des phrases ou des paragraphes pour différents domaines (éducatif ou autres), suivant plusieurs approches et techniques [19].

5. Les différents types de questions

Nous distinguons deux types de questions :

5.1 Subjective

Ce type de questions sert à examiner les connaissances approfondies du système éducatif traditionnel. Ça demande la rédaction des réponses où chacun peut répondre à sa manière (questions ouvertes). Il se divise en 4 sous-types :

● **Questions de rédaction** (la réponse est longue).

● **Questions à réponse courte** (la réponse est une phrase ou un petit paragraphe).

● **Questions de définition** (la réponse est une définition).

● **Questions d'opinion** (la réponse à ce type de question est l'opinion à propos d'un sujet donnée)

5.2 Objectives

Ce type de questions est fameux pour l'évaluation automatique pour un examen car c'est rapide et facile. Il inclut :

● **Questions à choix multiple** où la question possède plusieurs réponses dont une ou plusieurs sont correctes.

● **Questions vraies/fausses** où la réponse peut prendre une valeur soit vraie soit fausse.

● **Questions remplir-les-lacunes** où un ou plusieurs mots sont retirés de la phrase

6. L'évaluation des systèmes de génération de question

Pour dire qu'un système de génération de questions est bon ou pas, les questions générées par ce système doivent être évaluées. Et pour ce faire, deux types d'évaluations sont utilisés :

6.1 Evaluation automatique

Où un ensemble de calculs appelé métriques est appliqué sur cet ensemble de questions générées. Généralement les 3 métriques (BLEU, ROUGE et METEOR) sont utilisées.

- **BLEU [20]**

C'est une métrique rapide et simple qui sert à évaluer une phrase générée (candidate) par rapport à la phrase référence en calculant les mots (n-grammes qui peuvent être des uni-grammes, bi-grammes, trigrammes et quadrigrammes) avec la précision P.

La précision : P = n/c. Avec :

n : nombre de n-grammes de la phrase candidate présents dans la référence.

c : représente le nombre de n-grammes de la phrase candidate.

Cette formule pose un problème dans un cas où le mot se répète. Exemple :

Phrase candidate: the the the the

Phrase référence : the cat is on the mat

Si nous utilisons la formule précédente de précision, P = 4 / 4 et c'est faux.

Cette limitation de calcul est résolue, en utilisant la précision modifié f où $f = n/c$.

Avec :

n : - Si le nombre d'occurrence du n-gramme dans la phrase candidate ≤ le nombre d'occurrence du n-gramme dans la phrase référence, n= nombre d'occurrence du n-grammes dans la phrase candidate

- Sinon, n = nombre d'occurrence de n-gramme dans la phrase référence.

c : représente le nombre de n-grammes de la phrase candidate.

Donc avec l'exemple précédent : f = 1 /4.

- **ROUGE [21]**

ROUGE est un ensemble de métriques incluant : ROUGE-N, ROUGE-L et ROUGE-S.

ROUGE-L est la plus utilisée dans l'évaluation des systèmes de génération de questions. Elle consiste à mesurer le $F1 - Score$ en se basant sur la plus longue sous-séquence commune entre la phrase générée par le modèle et la phrase référence.

$$F1 - Score = 2 * \frac{Précision * Rappel}{Précision + Rappel} \qquad (4)$$

Avec :

$$Rappel = \frac{Nombre\ de\ n - grammes\ dans\ la\ plus\ longue\ sous - séquence}{Nombre\ total\ de\ n - grammes\ dans\ la\ référence} \qquad (5)$$

$$\text{Précision} = \frac{\text{Nombre de n−grammes dans la plus longue sous−séquence}}{\text{Nombre total des n−grammes dans la phrase candidate}} \qquad (6)$$

Où le n-grammes est une sous-séquence de n éléments à partir d'une séquence donnée d'où l'ordre des mots n'est pas important.

Exemple :

-La phrase candidate: The cat and the dog

-La phrase référence : The cat is on the mat

-La plus longue sous-séquence : 3 'the cat the'

-Nombre de n-grammes dans la plus longue sous-séquence : 3

Donc : la précision : 3/5 = 0.6

Le rappel : 3/6 = 0.5

Et F1-Score = $2 * \frac{0.6 * 0.5}{0.6 + 0.5}$ = 0.55

- **METEOR [22]**

Cette métrique calcule le niveau de similarité entre la phrase référence et la phrase hypothèse en utilisant le rappel et la précision et en créant un alignement entre les deux chaînes de caractères, METEOR attribue un score entre 0 et 1 ce score est calculé comme suit :

$$Score = Fmean.(1 - Pen) \qquad (7)$$

Avec :

$$\text{Fmean} = \frac{10\,PR}{R + 9P} \qquad (8)$$

$$\text{R} = \frac{\text{Nombre des unigrammes trouvés dans la phrase candidate et référence}}{\text{Nombre total des unigrammes dans la phrase référence}} \qquad (9)$$

$$\text{P} = \frac{\text{Nombre des unigrammes trouvés dans la phrase candidate et référence}}{\text{Nombre total des unigrammes dans la phrase candidate}} \qquad (10)$$

Et : Pen c'est Pénalité de bloc, le bloc est une suite de n-grammes adjacents.

$$Pen = 0.5\left(\frac{C}{m}\right) \qquad (11)$$

C : Nombre de n-grammes.

m : Nombre totale des mots dans la phrase candidate

6.2 Evaluation manuelle

Plusieurs critères sont pris en considération lors de l'évaluation manuelle (humaine) des systèmes de génération de questions, parmi ces critères nous citons ceux les plus utilisés :

- **La correction syntaxique** : indique la grammaticalité et la fluidité.

- **La grammaticalité** : sert à déterminer si la question respecte les règles grammaticales d'une langue.

- **La fluidité** : consiste à évaluer la simplicité et le naturel des termes dans les questions générées.

- **La pertinence** : sert à déterminer si la question générée est significative et est reliée au passage en entrée ou non

- **La réponse** : sert à déterminer si la question générée possède une réponse claire ou non.

- **La difficulté** : sert à déterminer si la question est facile à comprendre ou non, ou à déterminer son niveau de difficulté.

7. Les revues de la littérature

L'importance du e-learning a augmenté ces dernières années, pour automatiser le système d'évaluation, il préférable d'abord de penser à automatiser la génération des questions et aussi le processus d'évaluation automatique des réponses. Pleins de travaux ont été accomplis pour la deuxième partie, cependant la Génération Automatique des Questions en langue arabe, peu de travaux ont été réalisés c'est dû au manque de ressources, les outils du TAL pour la manipulation et la complexité de la langue. La GAQ a un rôle très important dans l'éducation, car les questions écrites par la main demandent beaucoup de travail et de temps.

Des analyses ont été faites sur plusieurs questions montrant la mauvaise qualité de ces dernières, donc le challenge a été de créer des questions de bonne qualité qui ont un sens et avec le plus grand nombre de questions possibles. Alors la solution est la génération automatique des questions. D'après [23] « La génération automatique des questions est une technique qui construit des algorithmes responsables de la production des questions depuis des ressources textuelles ou visuelles. »

1- A Systematic Review of Automatic Question Generation for Educational Purposes [23]

Les auteurs dans [23], ont eu pour but de proposer des améliorations dans le domaine de la GAQ ainsi que des nouvelles méthodes pour l'évaluation. Ils ont cité toutes les revues depuis 2014 jusqu'à début 2019, ils ont mis en évidence les mesures de difficultés des questions, automatisation des templates de construction pour la communauté qui s'intéressent à la GAQ.

2- Automatic Multiple Choice Question Generation From Text : A Survey [24] :

Les auteurs dans [24], s'intéressent aux questions à choix multiples. Ils ont cité les techniques de chaque phase de génération de questions à choix multiples (QCM) adoptées dans la littérature.

3- Evaluation methodologies in Automatic Question Generation 2013-2018 [25] :

Un Framework standard a été proposé pour les méthodes d'évaluations dans les systèmes de GAQ. Les auteurs dans [25], ont fait leurs études sur 37 autres études. Et ils ont concentré sur deux dimensions qui sont :

- Les méthodes d'évaluation intrinsèque qui servent à mesurer la performance du système par l'évaluation humaine et automatique de la sortie.
- Les méthodes d'évaluation extrinsèques qui servent à mesurer la performance du système en évaluant la sortie par rapport à la tâche qui doit être accomplie par ce système.

4- Automatic question generation approaches and evaluation techniques [26] :

Les auteurs dans [26], ont mis l'accent sur les méthodes des générations de questions existantes qui sont :

- Méthodes basées sur la syntaxe.
- Méthodes basées sur la sémantique.
- Méthodes basées sur un modèle.
- Méthodes basées sur des règles.
- Méthodes basées sur un schéma.

Et pour savoir également pourquoi le domaine de GAQ est très populaire ces dernières années.

8. Les Approches de génération automatique de questions

Les approches de génération automatiques des questions se divisent en deux catégories :

8.1 Les approches traditionnelles

- **Basées sur la syntaxe** : Cette approche se focalise et s'opère sur la grammaire de l'entrée, prenant en considération les relations entre les unités de l'entrée. La méthode POS est généralement utilisée dans cette approche pour guider la génération des questions. Cette approche à été utilisée par **(Dhole et Manning)** en 2020 dans leur framework **Syn-QG** où un ensemble de règles syntaxiques est utilisé pour améliorer la syntaxe des questions générées et éliminer les questions avec une fausse grammaire.
- **Basées sur la sémantique** : Cette approche concerne le sens et la signification de l'entrée. Elle demande le raisonnement pour extraire l'information car elle opère dans

un niveau profond. Pour cela des ressources de connaissance tel que les ontologies et taxonomies sont utilisées. Cette approche à été utilisée par **(Alberti et al.2019)**.

- **Mixtes** : Dans cette approche les deux niveaux syntaxique et sémantique sont pris en considération lors de la génération des questions. Cette approche à été utilisée par **(Pan et al)** en 2020, dans leur système **MQA-QG**.
- **Basées sur des règles :** Cette approche est basée sur la construction et extension des règles existantes, le système ne nécessite donc pas un grand corpus. Elle utilise le texte comme entrée, et la notation pour faire correspondre l'entrée avec une règle spécifique (manuelle).
- **Basées sur le modèle :** Cette approche utilise des modèles composés de "texte fixe" et des "espaces réservés" qui seront remplis ou renseignés à partir de l'entrée. Elle a été utilisée en 2020 par (**Fabbri** et **al**).

8.2 Les approches basées sur les données

Cette approche nécessite comme entrée des données sous format des paires question-réponse. Avec l'utilisation des approches de l'apprentissage automatique tel que les modèles séquence-à-séquence. Cette approche est plus moderne et a été utilisée dans les travaux connexes mentionnés dans la section 10.

9. Datasets

Plusieurs datasets existent pour la réalisation des systèmes de (génération automatique de questions) et de (réponses automatiques aux questions). Parmi ces datasets, nous avons recensé des datasets en anglais et en arabe présentés dans les tableaux 1 & 2 respectivement. Nous présentons pour chaque dataset les propriétés suivantes :

Le nom, le type, la taille, la source, le domaine, la description, et le lien par lequel le dataset est accessible.

Nom	Type	Taille	Source	Domaine	Description	Lien
SQuAD (2016)	N.C	150K	Wikipédia	Éducatif	**SQuAD 1.1 [27] :** contient 100K paires questions-réponses. **SQuAD 2.0 [28] :** contient les 100.000 paires de SQuAD 1 + 50.000 questions sans réponse.	https://www.kaggle.com/datasets/ananthu017/squad-csv-format
TriviaQA (2017) [29]	N.C	950K	Wikipédia	Éducatif	Contient 95k paires questions-réponses de. Se compose de sous ensemble de questions-réponses générées par lamachine et vérifiées par l'homme.Les réponses peuvent ne pas être directement obtenues par prédiction et le contexte	http://nlp.cs.washington.edu/triviaqa/
WikiQA (2015) [30]	N.C	3047	Wikipédia	Domaine-ouvert	Contient : 3047 questions, 29258 phrases où 1473 phrases ont été étiquetées comme phrases-réponses à leur questions correspondantes.	https://huggingface.co/datasets/wi
MS Macro (2018) [31]	C	1M	Bing	Compréhension écrite automatique	Une collection de datasets contenant 1.010.916 questionsréelles de Bing Avec chaque question avec une réponse générée par l'homme	https://huggingface.co/datasets/ms_marco

Amazon (2019) [32]	C	1.4M	Amazon	Commercial	Contient 1.4 million paires questions- réponses concernant des produits sur le site Amazon.	http://jmcauley.ucsd.edu/data/amazon/
Race (2017) [33]	N.C	100K	Examens d'anglais à la Chine	Éducatif	Contient 28K passages et 100K questions.	https://huggingface.co/datasets/race
LearningQ (2018) [34]	N.C	230K	Plateformes Apprentissage enligne	Éducatif	Contient 7K questions générées par des instructeurs et 223K questions par des étudiants	https://drive.google.com/file/d/1D9Qp6BB3prKYSKSM4B Eo04cw-71JFt/view
COM-QA (2019) [35]	C	11214	WikiAnswers	Modélisation de language	Contient 11214 questions groupées en 4834 paraphrase-clusters. -les clusters sont jumelés avec leurs réponses.	http://qa.mpi-inf.mpg.de/comqa/

Tableau 1: Datasets anglais avec N.C = Non Communautaire, C = Communautaire

Nom	Type	Taille	Source	Domaine	Description	Lien
XQuAD (2020) [36]	N.C	1190	SQuAD v1.1	Éducatif	Contient un sous- ensemble de 240 paragraphes + 1190 paires questions- réponses traduites de SQuAD 1.1 en 11 langues y compris l'arabe.	https://github.com/deepmind/xquad
MLQA (2019) [37]	N.C	5742	Wikipédia	Éducatif	Constitué de 12k instances en anglais et 5k dans 6 langues. - sa forme est de : contexte-question-réponse.	https://github.com/facebookresearch/MLQA
CQA-MD (SemEval) (2016) [38]	N.C	45164	Sites web (WebTeb, Al-Tibbi, IslamWeb)	Médical	Contient 1.531 contextes avec 30 paires de questions réponses pour chaque contexte dans le domaine médical.	https://alt.qcri.org/semeval2016/task3/index.php?id=data-and-tools
AskFM (2017) [39]	C	98K	AskFM	Recherche d'information	Contient 98K questions islamiques avec leurs réponses.	https://github.com/Omarito2412/ASKFM/blob/master/full_dataset.csv
AQAD (2020) [40]	N.C	+17K	Wikipédia	N/A	+17000 question- réponse. -nombre de questions : 17.911 -nombre d'articles : 299 -nombre de paragraphes : 3.381	https://github.com/adelmeleka/AQAD/blob/master/AQQAD%201.0/FINAL_AAQAD-v1.0.json

DAWQAS (2018) [41]	C	3205	Sites arabes publiques	Éducatif	Contient 3205 paires questions-réponses	https://github.com/masun/DAWQAS/blob/master/DAWQAS_Masun_Nabhan_Homsi.xlsx
AR-ASAG (2020) [42]	N.C	2133	Examens de cours cybercriminalité de 3 classes	Éducatif	Contient 48 questions & 48 réponsesmodèles -2133 réponses d'étudiants -5 types de questions (1- Définir. 2-Expliquer. 3- Quelles sont les Conséquences. 4- Justifier. 5- Quelleest la différence).	https://data.mendeley.com/datasets/dj95jh332j/1
ARCD (2019) [43]	N.C	1.4K	Wikipédia + Arabic-SQuAD	Éducatif	Contient 1395 questions et segments de texte) – sa forme : passage - questions - réponses.	https://github.com/husseinmozannar/SOQAL/blob/master/data/arcd.json
Arabic-SQuAD (2019) [43]	N.C	48344	Traduction de SQuAD	Éducatif	Contient 48.344 passage-question-réponse. -nombre de paragraphe : 10.364 -nombre d'articles : 231	https://github.com/husseinmozannar/SOQAL/bl ob/master/data/Arabic-SQuAD.json

Tableau 2: Datasets arabe avec N.C = Non Communautaire, C = Communautaire

10. Travaux connexes

Dans cette partie nous nous intéressons aux travaux ayant utilisé l'approche basée sur les données et les modèles séquences à séquences. Nous citons :

- **Learning to Ask: Neural Question Generation for Reading Comprehension [44]:**

Les auteurs dans [45], ont utilisé l'architecture (encodeur/décodeur) avec un mécanisme d'attention. Leur système a 2 variantes. Un qui prend en entrée une phrase et l'autre prend en entrée un paragraphe.

Le système est inspiré de la façon dont l'homme pense. La solution était de créer un modèle RNN avec un mécanisme d'attention pour focaliser sur les parties pertinentes de l'entrée. La partie du décodeur utilise LSTM. La partie de L'encodeur est basée sur mécanisme d'attention avec bi-LSTM.

Expérimentation :

Datasets : l'ensemble de données utilisé est SQuAD.

Évaluation : pour l'évaluation de ce système les deux types d'évaluationont été utilisées (automatique et manuelle).

Pour l'évaluation automatique, ils ont utilisé les métriques BLEU (1,2,3,4),ROUGE-L et METEOR.

Pour l'évaluation manuelle, ils ont pris en considération les critères suivants(difficulté, simplicité). Notés de 1 à 5 (5 pour le meilleur résultat).

Exemples de questions générées :

Entrée 1: the largest of these is the eldon square shopping centre, one of the largestcity centre shopping complexes in the uk.

Question générée 1: what is one of the largest city centers in the uk ?

Entrée 2: free oxygen first appeared in significant quantities during the paleoproterozoic -lrb- between 3.0 and 2.3 billions years ago -rrb-.

Question générée 2: how long ago did the paleoproterozoic exhibit?

- **Machine Comprehension by Text-to-Text Neural Question Generation [45]**:

Les auteurs dans [45], ont combiné l'apprentissage supervisé avec l'apprentissage par renforcement. Ils ont utilisé une architecture encodeur-décodeur avec un mécanisme d'attention et un mécanisme de copie. L'encodeur utilise le Bi-LSTM et prend 2 entrées : document D et

réponse A. Pour le décodeur, ils ont utilisé LSTM avec un mécanisme de copie. Pour la partie d'apprentissage par renforcement, ils ont décrit 2 méthodes : l'utilisation du REINFORCE qui est un ancien algorithme depuis 1992 qui sert à maximiser la récompense de chaque question générée, ou bien l'utilisation d'un système de réponse aux questions.

- Expérimentation :

- Datasets : pour l'ensemble de données utilisé dans ce modèle c'est SQuAD.

- Évaluation : Pour l'évaluation de leur système, ils ont appliqué l'évaluation automatique et l'évaluation manuelle.

Pour l'évaluation automatique les métriques suivantes ont été utilisées : BLEU, F1-score.

Pour l'évaluation manuelle aucun critère n'a été mentionné.

- Exemples de questions générées:

- **Entrée:** the court of justice accepted that a requirement to speak gaelic to teach in a Dublin design college could be justified as part of the public policy of promoting the Irish language.
- **Question générée:** what language did the court of justice say should be justified as part of the public language?

A la comparaison de leur système avec le système Séquence-à-séquence de base, les résultats ont été meilleurs par rapport aux métriques utilisées BLEU et F1-score.

- **A Joint Model for Question Answering and Question Generation [46]**

Les auteurs dans [46], ont réalisé ce système en 2017. Ils ont utilisé un mécanisme d'attention et un décodeur avec un pointeur Softmax. Ce système prend en entrée un document et une condition qui correspond à une séquence de mots de la question dans le mode de génération de réponse. Ou à une séquence de mots de la réponse dans le cas de la génération des questions. L'encodeur utilise une couche de bi-LSTM. Le décodeur est basé sur RNN avec un mécanisme de pointeur Softmax implémenté avec 2 cellules LSTM.

Expérimentation :

- **Datasets :** pour l'ensemble de données utilisé dans ce modèle c'est SQuAD. En prenant 5158 paires (questions-réponses) de l'ensemble d'apprentissage de SQuAD pour faire la validation.

Et l'ensemble de données de développement dans SQuAD pour rapporter les résultats de leur test.

- **Évaluation :** Pour l'évaluation de leur système de génération de questions, ils ont utilisé l'évaluation automatique par les métriques (BLEU-4 et F1-score).

- Exemples de questions générées:

- **Entrée 1:** in the 1960 election to choose his successor, eisenhower endorsed his own vice president, republican richard nixon against democrat john f. kennedy .

- **Question générée:** what was the name of eisenhower 's own vice president?

- **Entrée 2:** in 1870, tesla moved to karlovac, to attend school at the higher real gymnasium, where he was profoundly influenced by a math teacher martin sekulic´.

- Question générée 2: what did tesla do at the higher real gymnasium?

Les résultats du modèle JointQA ont été bons pour la génération de réponses par rapport au modèle A-Gen qui génère uniquement des réponses.

Mais ils ont été mauvais pour la génération de questions par rapport au modèle Q-Gen qui sert à générer des questions seulement.

- **Neural Generation of Diverse Questions using Answer Focus, Contextual and Linguistic Features [47]:**

Les auteurs dans [47], ont utilisé une architecture encodeur/décodeur avec un mécanisme d'attention et mécanisme de copie, pour copier les mots directement depuis la phrase vers la question. Et les propriétés :

- Signal de la réponse (Ans) : sert à guider le modèle sur quelle information dans la phrase il doit se concentrer.

- Case : une valeur binaire qui sert à déterminer si l'entrée contient des caractères majuscules ou non.

- NER (Named Entity Recognition) qui veut dire la reconnaissance des entités nommées, cette propriété sert à identifier les entités importantes dans une phrase.

- CoRef : la coréférence calculée lors de la phase du prétraitement sert à identifier quelle mention dans un texte renvois à une entité.

L'encodeur utilise plusieurs couches de bi-LSTM. Le décodeur utilise des LSTM.

- Expérimentation :

- Datasets : l'ensemble de données utilisé dans ce modèle est SQuAD.

- **Evaluation :** Pour l'évaluation de leur système, ils ont utilisé les métriques (BLEU-4, METEOR et ROUGE) dans l'évaluation automatique. Et les critères (Grammaticalité, Information correcte, Simplicité, Réponse) pour l'évaluation humaine.

- **Exemples de questions générées :**

- **Entrée:** The character of midna has the most voice acting – her on-screen dialog is often accompanied by a babble of pseudo-speech , which was produced by scrambling the phonemes of english phrases [better source needed] sampled by Japanese voice actress akiko komoto.

- **Questions générées :**

  - 1 / what character has the most voice acting in english?
  - 2 / what is the name of the japanese voice actress?
  - 3 / what is the nationality of akiko komoto?

En comparaison avec le système de (Du et al.2017), FOCUS a obtenu de meilleurs résultats par rapport à toutes les métriques d'évaluation automatiques utilisées.

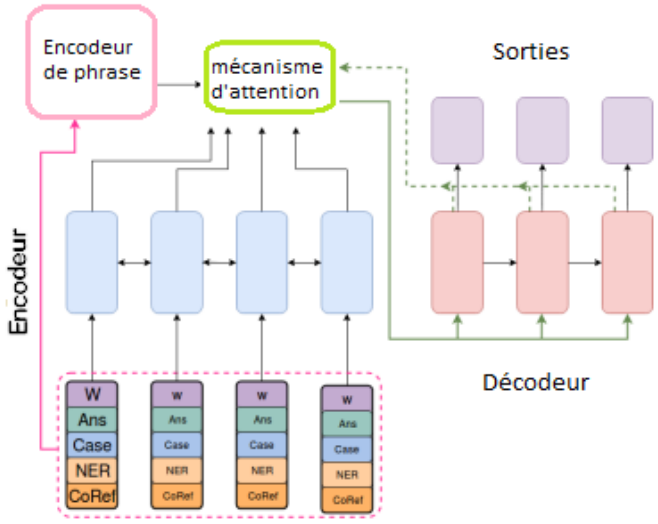


Figure 17: Diagramme du modèle FOCUS [47]

- **Improving Neural Question Generation Using Answer Separation (ASs2s) [48]**:

Les auteurs dans [48], Ont réalisé un système en 2019 qui consiste à séparer la réponse du passage initial, cette architecture est composée d'un encodeur-décodeur avec un mécanisme d'attention. L'encodeur de son tour est composé de 2 parties : un pour encoder le passage, et l'autre pour la réponse. Les deux utilisent une couche de bi-LSTM. Pour le décodeur, il utilise

les caractéristiques contextuelles de passage avec le mécanisme d'attention et les mots-clés sur la réponse cible, pour générer une question liée à la réponse depuis le passage.

- Expérimentation :

- Datasets : l'ensemble de données utilisé est SQuAD.

- Evaluation : Pour l'évaluation de leur système, ils ont utilisé les métriques d'évaluation automatique (BLUE-4, METEOR et ROUGE-L).

En comparaison avec le modèle de (Du et al.2017), ASs2s propose de meilleurs résultats par rapport à toutes les métriques.

- **Question-type Driven Question Generation NQG++ [49] :**

Les auteurs dans [49], proposent un modèle basé sur les types de questions, qui prédit automatiquement le type de la question, selon l'entrée qui est constituée de la réponse et le passage. Pour qu'à la fin, générer des questions. Ils adoptent une architecture encodeur-décodeur avec un mécanisme d'attention.

L'encodeur est utilisé pour encoder la phrase et la réponse correspondante. Les propriétés utilisées : POS, NER, ensuite les WE et la position de la réponse. Les propriétés lexicales sont concaténées pour former l'entrée.

- Expérimentation :

- Datasets : Les données utilisées sont des triplets de phrase-réponse-question depuis les datasets SQuAD et MARCO. Pour le dataset SQuAD (entraînement : 86,635), (développement : 8,965), (test : 8,964). Pour le dataset MARCO (entraînement : 74,097),(développement : 4,539),(test : 4,539).

- Evaluation : pour l'évaluation de leur système, ils ont utilisé la métrique BLEU pour l'évaluation automatique (BLEU 1-2-3-4) uniquement.

- Exemples de questions générées:

- Entrée 1: In land plants, chloroplasts are generally lens-shaped, 5–8 m in diameter and 1–3 m thick.

- Question générée 1: How are chloroplasts in land plants?

- Entrée 2: Article 65 of the agreement banned cartels and article 66 made provisions for concentrations, or mergers, and the abuse of a dominant position by companies.

- Question générée 2: Which article made provisions for concentrations, or mergers?

- **Learning to Generate Questions by Learning What not to Generate (CQC-QG) [50]:**

Les auteurs dans [50], ont réalisé un système en 2019, dont le but est de générer des questions automatiquement. C'est une architecture décodeur-encodeur avec les mécanismes d'attention et de copie.

Il est constitué de 3 composants : prédicteur de mot indice, qui sert à prédire des indices qui sont pertinentes pour générer des questions depuis le passage. L'encodeur du passage qui utilise le mot indice prédit avec des propriétés sur l'entrée tel que, la position de la réponse ou les caractéristiques lexicales. Le dernier composant c'est le décodeur qui apprend la probabilité de générer un mot et copier le mot depuis le passage initial.

- Expérimentation :

- Datasets : deux ensembles de données ont été utilisés (SQuAD & NewsQA).

- Evaluation : pour l'évaluation de leur système, ils ont utilisé les métriques d'évaluation automatiques (BLEU (1-2-3-4), ROUGE-L et METEOR) sur les deux datasets SQuAD et NewsQA.

Pour le but d'évaluer la performance de leur système, ils ont comparé les résultats d'évaluation automatique avec plusieurs d'autres systèmes y compris NQG++. Les résultats de CGC-QG ont été meilleurs par rapport à toutes les métriques.

- **Paragraph-level Neural Question Generation with Maxout Pointer and Gated Self-Attention Networks [51]:**

En 2018, les auteurs dans [51], ont réalisé un système de génération de questions qui se focalise sur le niveau 'paragraphe' et non pas au niveau 'phrase'. Ils utilisent un modèle séquence à séquence avec un mécanisme d'auto attention fermé (gated self attention) qui est conçu pour agréger les informations de tout le passage et encastrer les dépendances dans ce dernier pour affiner la représentation codée (passage-réponse) à chaque pas de temps. Un mécanisme pointeur-maxout qui sert à éliminer la répétition. Et un marquage des réponses.
C'est le premier modèle qui montre une amélioration avec un paragraphe en entrée par rapport à une phrase en entrée.

-Expérimentation :

- Datasets : les deux ensembles de données SQuAD et MS MARCO ont été utilisés pour l'entraînement et le teste de ce système.

- Evaluation : Pour l'évaluation automatique, ils ont utilisé les trois métriques (BLEU 1-2-3-4), ROUGE-L et METEOR.

- Exemples de questions générées :

-Entrée: a problem is regarded as inherently difficult if its solution requires significant resources, whatever the algorithm used. The theory formalizes this intuition, by introducing mathematical models of computation to study these problems and quantifying the amount of resources needed to solve them, such as time and storage. Other complexity measures are also used, such as the amount of communication (used in communication complexity), the number of gates in a circuit (used in circuit complexity) and the number of processors (used in parallel computing). One of the roles of computational complexity theory is to determine the practical limits on what computers can and can not do.

Question générée: what is another name for circuit complexity?

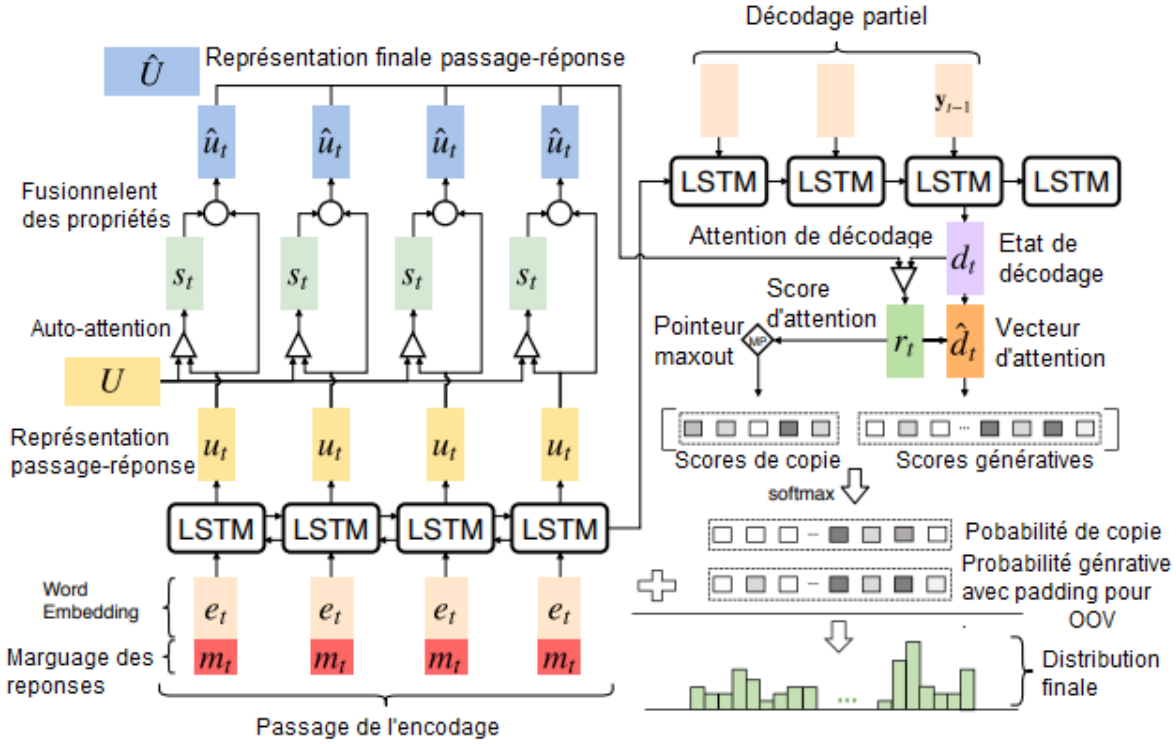


Figure 18 : Diagramme du modèle [51]

- **Self-Attention Architectures for Answer-Agnostic Neural Question Generation [52]:**

Dans ce système, ils ont utilisé les transformers pour générer des questions sauf que dans ce cas ils n'ont pas pris la réponse comme entrée et leur justification était que le système ne sera pas libre car la réponse impose la génération des questions pertinentes. L'apport de ce travail est qu'ils ont pris l'architecture de base d'un transformer et ils ont également ajouté le mécanisme de copie, tenue de place, l'incorporation des mots contextuels.

- Expérimentation :

- Datasets : l'ensemble de données utilisé est : SQuAD.

- Évaluation : évaluation automatique et évaluation humaine.

Pour l'évaluation automatique, ils ont calculé les métriques (BLEU (1-2-3-4), ROUGE-L).

Pour l'évaluation humaine, les critères pris en considération sont (grammaticalité, fluidité, réponse, pertinence).

- Exemples de questions générées:

- Entrée 1: Chopin was of slight build, and even in early childhood was prone to illnesses.

- Question genre 1: what type of disease did chopin have?

- Entrée 2: Montana contains thousands of named rivers and creeks, 450 miles (720 km) of which are known for" blue-ribbon" trout fishing.

- Question générée 2: how many miles of rivers does montana contain?

- **Arabic Question Generator (AQG) [53] :**

Le premier générateur de questions pour la langue arabe, ce système n'utilise ni l'architecture Encodeur/Décodeur, ni transformer, ni finetuning. Il a été initialement créé pour la plateforme QUIZZITO dédiée à l'éducation. Ce système passe par plusieurs étapes avant d'arriver à la génération de questions. Dans un premier temps, le système combine le SRL pour la génération de rôles sémantiques, qui sert à décrire comment le mot est utilisé et quel est son but dans une phrase. Et ceci à été réalisé par l'utilisation de $PropBank$[1] 1, après l'étape de prétraitement ou les deux outils du TAL (MADAMIRA2[2] et STAR3[3]) ont été utilisés pour la tokenization et la

[1] PropBank : un corpus de texte annoté avec des informations sur les propositions verbales sémantiques
[2] MADAMIRA : un système d'analyse morphologique et de désambiguïsation de la langue arabe.
[3] STAR : un outil de TAL, qui sert à segmenter des textes arabes.

segmentation. Un ensemble d'arguments à été assigné aux différentes catégories grammaticales. Cet ensemble est :

- Arg 0 (A0) → le sujet
- Arg 1 (A1) → l'objet
- AM_LOC → locatif
- AM_TMP → temporel
- AM_PRP → préposition
- Predicate → verbe

L'étape suivante consiste à faire la conception des motifs des phrases (phrases nominales ou phrases verbales).

- Exemple : une phrase nominale peut être écrite sous la forme A0+A1

[الله لطيف بعباده]

A1 + A0

Ces motifs ont été pris en compte dans l'étape suivante lors de la conception des modèles, avec l'utilisation des interrogatifs de la langue arabe, pour que à la fin arriver à la génération de questions.

- Expérimentation :

- Datasets : l'ensemble de données utilisé et (corpus-QUIZZITO) qui contient 40435 questions dérivées des livres pour enfants.

- Evaluation :

- Evaluation humaine : en considérant les critères (grammaticalité, réponse).

- Evaluation automatique : en calculant F-mesure. Les résultats étaient différents selon le texte sur lequel les questions ont été générées.

- Exemples de questions générée :

- ماذا سأل الفهد؟
- من سمحت بتمشيط شعرها؟

Il n'est pas possible de comparer ce modèle avec les autres, car c'est le seul qui concerne la langue arabe.

- **Simplifying Paragraph-level Question Generation via Transformer Language Modelsl [54] :**

Les auteurs dans [54], ont utilisé le modèle pré-entraîné GPT2[4]. Donc ils ont fait un fine-tuning avec changement de format de dataset choisie, SQuAD dataset version 1.1 été utilisé, ils ont gardé le passage mais les questions sont devenues enchaînées dans une seule phrase séparée par des espaces.

Pour l'évaluation de ce système, ils ont appliqué les métriques (BLEU (1-2-3-4), ROUGE-L et METEOR).

En comparant les résultats de l'évaluation automatique avec le système NGQ par (Du et al) et le système NQG++ par (Zhao et al), ce système a obtenu de meilleurs résultats par rapport aux métriques ROUGE-L et METEOR, mais le résultat de BLEU était le moins d'entre eux.

- **Question Generation by Transformers [55] :**

Ce système a été développé en 2019, il est basé sur le mécanisme d'attentions en utilisant des transformers. Pour l'évaluation de leur système Ils ont utilisé WER[5] comme métrique d'évaluation automatique sur l'ensemble de données SQuAD.

Comme c'est le seul système parmi ceux que nous avons abordé qui a utilisé WER comme métrique d'évaluation, il n'est pas possible de faire la comparaison.

- **Learning to Answer by Learning to Ask: Getting the Best of GPT-2 and BERT Worlds [56]:**

Dans ce système, les deux méthodes (génération de questions) et (réponses aux questions) ont été appliquées par l'utilisation des deux transformers GPT2 et BERT[6] afin de générer des questions valides et qui possèdent des réponses. Le transformer GPT2 pour la (génération de questions). Le transformer BERT pour la (réponses aux questions). Ce modèle à été entrainer et évaluer sur le dataset SQuAD 1.1.

Pour l'évaluation de ce système, les métriques d'évaluation BLEU (1.2.3.4), ROUGE-L et le score F1 ont été utilisées.

Par rapport aux métriques BLEU et ROUGE-L, les résultats de ce système ont été mauvais en comparaison avec les autres travaux connexes. Mais pour le score F1, en comparaison avec les

[4] GPT2 : est un transformer pré entraîne sur un large ensemble de données (texte brut non étiqueté par l'homme). Il est spécialement conçu pour la modélisation de langage.

[5] WER : une métrique d'évaluation automatique généralement utilisée dans les systèmes de reconnaissance de la voix.

[6] BERT : un modèle de représentation de langage, développé par google en 2018.

autres systèmes mentionnés qui ont utilisé cette mesure d'évaluation, ce système à obtenu le meilleur résultat.

11. Synthèse des travaux

Après avoir récupéré les résultats d'évaluations obtenus dans les travaux connexes mentionnés précédemment, nous représentons ces résultats sur le dataset les plus utilisé : SQuAD dans le tableau suivant :

Dataset	SQUAD					
	BLEU				Rouge-L	METEOR
Modèle	Bleu1	Bleu2	Bleu3	Bleu4		
NQG [44]	41	23.78	15.71	10.8	37.95	15.17
FOCUS [47]	/	/	/	19.98	48.23	22.26
ASs2s [48]	/	/	/	16.20	43.96	19.92
CGC-QG [50]	46.58	30.90	22.82	17.55	44.53	21.24
JointQA [46]	/	/	/	10.20	/	/
NQG++ [49]	43,11	29,13	21,39	16,31	/	/
Zhao et al [51]	43,47	28,23	20,4	15,32	43,91	19,29
Scialom et al. [52]	43.33	26.27	18.32	13.23	40.22	/
Lopez et al.[54]	/	/	/	8.26	44.38	21.2
Yuan et al.[45]	/	/	/	10.5	/	/
Klein et Nabi [56]	31.46	19.50	12.41	7.84	34.51	/

Tableau 3: Résultats d'évaluation automatique des différents systèmes sur le dataset SQuAD

Nous pouvons déduire à partir de ce tableau que le système FOCUS a obtenu les meilleurs résultats par rapport aux trois métriques utilisées.

Nous avons remarqué que la majorité des travaux ont utilisé un dataset formel car il est grammaticalement correct, et ne contient pas des fautes d'orthographes et c'est pour cette raison qu'ils obtiennent des meilleurs résultats par rapport aux datasets communautaires.

Dans la suite, nous allons synthétiser les travaux et systèmes mentionnées ci-dessus dans l'approche axée sur les données, nous allons faire une étude comparative où nous extrayons de chaque système l'ensemble des caractéristiques suivantes : propriété pertinente du système, l'année, l'utilisation du mécanisme d'attention, dataset utilisée, les métrique d'évaluation et l'évaluation humaine.

Le tableau -4- contient les détails de cette analyse comparative :

	Langue	Année	Dataset	Propriété	Mécanisme d'attention	Métrique d'évaluation	Evaluation humaine
NQG [44]	Anglais	2017	SQuAD	Encodeur/décodeurencodeur : bi-Lstm décodeur : Lstm	Oui	BLEU (1.2.3.4) ROUGE-L METEOR	Difficulté Fluidité Grammaticalité
FOCUS [47]	Anglais	2018	SQuAD	Encodeur avec bi-Lstm et décodeur avec Lstm et mécanisme de copie	Oui	BLEU (4) ROUGE-L METEOR	Grammaticalité Fluidité Information correcte Réponse
ASs2s [48]	Anglais	2019	SQuAD	2 encodeurs : bi-Lstm Décodeur avec mots clés	Oui	BLEU (4) ROUGE-L METEOR	/
NQG++[49]	Anglais	2019	SQuAD	Encodeur/décodeur	Oui	BLEU (1.2.3.4)	Sens
CGC-QG [50]	Anglais	2019	SQuAD	Encodeur/décodeur Avec mécanisme de copie	Oui	BLEU (1.2.3.4) ROUGE-L METEOR	/
JointQA [46]	Anglais	2017	SQuAD	Encodeur : bi-Lstm, décodeur : Lstm	Oui	BLEU (4) F1-score	/
Yuan et al [45]	Anglais	2017	SQuAD	Encodeur : bi-Lstm Décodeur : Lstm + mécanisme de copie	Oui	BLEU (4) F1-score	/

Scialom et al [52]	Anglais	2019	SQuAD	Transformer avec mécanisme de copie	Non	BLEU (1.2.3.4) ROUGE-L	Grammaticalité Fluidité Réponse Pertinence
Lopez et al [54]	Anglais	2020	SQuAD	Finetuning GPT2	Non	BLEU (1.2.3.4) ROUGE-L METEOR	/
Klein et Nabi [56]	Anglais	2019	SQuAD 1.1	Finetuning GPT2 et BERT	Oui	BLEU (1.2.3.4) ROUGE-L F1	/
Zhao et al. [51]	Anglais	2018	SQuAD MS Marco	Encodeur/décodeur avec mécanisme pointeur maxout et auto-attention	Oui	BLEU (1.2.3.4) METEOR ROUGE-L	/

Tableau 4: Résultats de comparaison des modèles

Comme nous nous somme intéresser uniquement aux travaux ayant utilisé l'approche basée sur les données, il n'est pas possible de comparer les résultats obtenus par ces travaux avec les travaux ayant utilisé d'autres approches car les mesures d'évaluation ne sont pas les mêmes.

Mais nous pouvons constater à partir de ces deux tableaux de synthèse que les modèles utilisant une architecture encodeur/décodeur + mécanisme d'attention ont obtenu de bons résultats pour la tâche de génération de questions pour la langue anglaise par rapport aux travaux ayant utilisé un finetuning ou des Transformers.

12. Les systèmes de réponse aux questions et les systèmes de réponse aux questions communautaires

Comme le but de notre système est de générer des questions de meilleure qualité, nous avons le besoin d'utiliser des données qui sont dédiés aux systèmes de réponses aux questions. Ces systèmes prennent en entrée non seulement des paragraphes contextuels mais aussi des réponses. En introduisant la réponse en entrée dans un système de génération de question, la question générée va dépendre de cette réponse et va en avoir une relation directe et donc la question est meilleure pour atteindre cette réponse. Autrement dit, le système de génération de question est influencé par le fonctionnement d'un système de réponses aux questions.

Mais cela sert beaucoup plus à atteindre la réponse qu'à générer une question meilleure à une qui existe déjà. Pour cela nous nous intéressons aux systèmes de réponses aux questions communautaire. Où les données sont des questions et des réponses générées par des utilisateurs. Qui dit utilisateurs dit phrases informelles, fautes grammaticales, fautes d'orthographes, etc. Ce qui va rendre visible la qualité des questions générées automatiquement par un système. Ainsi, la nature des données va affecter largement les résultats finals.

12.1 Système de réponse aux questions

Un système de réponse aux questions est une approche automatisée pour récupérer les réponses correctes aux questions posées en langage naturel [57].

Le terme Récupération d'informations RI signifie une tâche de récupération à partir d'une grande collection de données non structurées. Ce système est conçu pour mettre à la disposition d'un utilisateur une collection d'informations stockées.

L'un des sous-domaines de l'RI est la réponse aux questions. Une bonne alternative aux moteurs de recherche est le système de réponse aux questions qui est composé des phases suivantes : analyse des questions, récupération des passages et extraction des réponses.

12.2 Système de réponse aux questions communautaire

Ce sont des services web où les utilisateurs interagissent entre eux en posant et répondant aux questions.

L'objectif principal de ce genre de système de réponse aux questions est de fournir les réponses les plus appropriées dans les plus brefs délais. Ils existent en deux types qui sont [58]:

- Les systèmes dédiés à un domaine spécifique ou l'utilisateur peut poser des questions uniquement dans le domaine spécifié.
- Les systèmes sans restriction ou l'utilisateur peut poser des questions dans n'importe quel domaine.

Le tableau -5- montre la différence entre les systèmes de réponses aux questions et les systèmes de réponses aux questions communautaires par rapport aux types de la question, source de la réponse, la qualité de la réponse, la disponibilité des métadonnées et le temps d'attente [59].

	Système QA	Système QA communautaire
Type de question	Questions factoïdes à une seule phrase	Questions sur plusieurs phrases
Source de la réponse	Extrait d'un document d'un corpus	Apporté par les utilisateurs
Qualité de la réponse	Haute qualité, extraite de sources réputées	Varié car elle dépend de l'utilisateur
Disponibilité demétadonnée	Pas de métadonnée	Meilleure réponse sélectionnée par le demandeur, notes positives et négatives données par les sélecteurs.
Temps d'attente	Automatique et immédiate	Le demandeur doit attendre que d'autres utilisateurs postent des réponses.

Tableau 5: Comparaison en QA et CQA

13. Conclusion

Les différentes notions liées à la génération automatique de questions ont été présentées dans ce chapitre, tout en focalisant sur les approches existantes et les travaux connexes les plus récents. Le chapitre qui suit est consacré à notre modèle de génération de questions basé sur

l’architecture encodeur/décodeur avec plus de détails sur les datasets choisis pour les deux langues arabe et anglaise, l'entraînement du modèle ainsi que les différentes méthodes d’évaluation.

Chapitre III : Conception

1. Introduction :

Ce chapitre présente l'architecture générale de notre système de génération de questions. La recherche menée avant dans ce rapport indique l'existence de plusieurs travaux connexes dans ce cadre. Par contre nous avons remarqué qu'aucun travail qui intègre l'apprentissage profond n'est dédié à la langue arabe.

Nous détaillerons dans ce qui vient l'architecture du système, le choix des datasets avec les différents composants de système.

2. Contribution de notre travail

Le but de notre générateur de question est d'améliorer la qualité des résultats obtenus lorsqu'une personne saisit sa requête dans un moteur de recherche. Notre travail a pour objectif de produire des questions naturelles et correctes, et suggérer des exemples de questions associées à la requête de l'utilisateur.

Puisque notre travail est le premier dans la langue arabe, nous avons fait un modèle pour la langue anglaise aussi, pour pouvoir comparer les résultats obtenus avec les résultats existants, et montrer que le modèle peut être adaptée à une autre langue.

3. Définition de la tâche de génération de questions

Dans cette partie, nous définissons la tâche de génération de questions. C'est une des tâches les plus importantes dans le domaine de traitement automatique du langage naturel. La définition formelle est :

Etant donnée une réponse R qui contient N mots : R = (r_1, r_2, r_3, …, r_N), dont r_i est un token dans la réponse et N est la langueur de la séquence. Le but est de générer une question Q de M tokens avec Q = (q_1, q_2, q_3, …, q_M), et M la longueur maximale de la question. Afin de générer la question nous devrons maximiser la probabilité conditionnelle P(Q|R), dont :

$$P(Q|R) = \prod_{m=1}^{M} P(q_m | q_{1:m-1}, R)$$

Ceci indique que afin de trouver la probabilité de q_m, nous devrons prendre en considération toute la séquence générée auparavant et la réponse R.

4. Choix des Datasets

Chaque système intelligent a besoin d'acquérir des connaissances à partir d'un corpus de données, appelé aussi 'dataset'. Un dataset communautaire est issue des sites web et des forums ('StackOverflow', 'Yahoo !', 'Quora','Amazon'), où des utilisateurs posent leurs questions sur un sujet spécifique ou bien un produit et d'autres utilisateurs répondent. Pour réaliser notre travail, il nous faut un dataset communautaire pour l'arabe et un autre pour anglais.
Mais nous étions limités dans notre choix vu le manque des datasets en langue arabe, et surtout un dataset communautaire de grande taille qui répond aux exigences de l'apprentissage profond et peut être entrainé avec les moyens informatiques disponibles.

Nous avons choisi deux dataset informels pour l'anglais 'Amazon QA paires' disponible sur [60] et pour l'arabe 'SemEval 2016- task 3' disponible sur [61], pour que les paires des questions et réponses soient proches à la requête d'un utilisateur dans un moteur de recherche. Généralement ce que l'utilisateur introduit comme requête est un language informel qui n'est pas correct grammaticalement.
-Le dataset arabe est un dataset médical utilisé par SemEval 2016, il est issu de 3 sites web : WebTeb[7], Al-Tibbi[8], Islamweb[9]. La tâche principale de ce dataset est de reclasser les bonnes réponses pour une nouvelle question. Ce dataset est composé initialement de 1531 contextes et 45164 paires de questions-réponses [38]. Afin de générer le dataset de SemEval-2016, les 1531 contextes étaient pris depuis WebTeb, et les paires de questions-réponses associées depuis Al-Tibbi avec 69582 paires et Islamweb avec 31714 paires.
Le résultat final est plus de 100.000 paires, néanmoins ils ont seulement utilisé les 1531 contextes avec les 30 paires classées en premier [62].

Plusieurs défis ont été rencontré pour adapter le dataset pour la tâche de la génération automatique des questions tels que : l'apparition de la même paire de question-réponse avec plusieurs contextes, l'utilisation des différents dialectes, les fautes d'orthographe (suppression des lettres, fusionnement des mots), les paires très détaillées ou bien des questions avec des réponses très courte. Après avoir effectué les traitements -manuel et automatique- nécessaires (voir section 4.1), nous nous sommes retrouvés avec 20112 paires de questions-réponses

[7] https://www.webteb.com/
[8] https://altibbi.com/
[9] https://www.islamweb.net/ar/consult/

-Le dataset anglais est un dataset communautaire qui contient plus de 1.4 millions pairs de question-réponse où les clients posent leurs questions sur 191 milles produits regroupés dans 21 catégories disponible sur le site d'Amazon [63]. Nous étions aussi limités dans notre choix pour le dataset anglais car la majorité des datasets trouvés en anglais pour la génération des questions sont dans le domaine éducatif et donc non communautaire.

Nous avons choisi 6 catégories (produits bébé, produits de santé, accessoires téléphones, produit de bureau, produits de maison et cuisine et outils de maison et amélioration de l'habitat) afin de créer notre dataset qui contient 225000 paires de questions-réponses.

4.1 Adaptation et traitement du dataset arabe

Le dataset choisi est un dataset destiné à la tache de génération de réponse. Il a été proposé dans le cadre du concours SemEval (Semantic Evaluation) pour la tache de validation de réponses dans un système de réponse aux questions. Nous l'avons adapté à notre tache de génération de questions en passant par les prétraitement manuelle et automatique.

Le taille de dataset arabe après les prétraitements effectués a diminué par rapport à la taille initiale, Ainsi nous avons fixé une contrainte sur la longueur des entrées. Les contraintes sont :

Pour les questions : 4 tokens ≤ longueur de la question ≤ 35 tokens

Pour les réponses : 10 tokens ≤ longueur de la question ≤ 35 tokens

En appliquant ces conditions, la taille de dataset est passé de 45164 à 20112 paires. Par la suite, nous allons consacrer 85% des paires pour l'entrainement et 15% pour le test. Le tableau 6 contient plus de détails sur l'utilisation de dataset :

Dataset	Entrainement (85%)	Test (15%)	Total (100%)
SemEval 2016	17095	3017	20112

Tableau 6 : détails d'utilisation de dataset arabe

4.1.1 Prétraitement Manuel

Le dataset initial contient 30 paires de question-réponse pour chaque contexte, nous avons supprimé tous les contextes et gardé seulement les paires. Plusieurs défis ont été rencontrés pour que le dataset soit adapté et utilisé par le modèle, le dataset est de nature communautaire donc les paires de question – réponse ne sont pas grammaticalement correctes et ne sont pas contrôlées, Ceci va impacter largement les résultats obtenus. Voici les détails sur le traitement manuel effectué :

-Un des défis est les introductions et les salutations très longues, car c'est très commun en arabe de commencer et terminer par les salutations religieuses. D'après [62], 58% des questions commence par ce genre d'introductions et 50% des réponses se termine par des prières et des douaas. Pour ne pas perdre le sens de la question ou encore la réponse nous avons passé manuellement par le dataset pour supprimer ce genre d'introductions. Exemple : « السلام عليكم، الإجابة : تحية طيبة أم فاضل أما » ,«نشكركم لثقتكم و الله الموفق » ,«أما بعد نشكركم على هذا الفضاء لطرح انشغالاتنا «بعد نسأل الله السلامة و العافية. L'exemple suivant représente un exemple d'une paires QA depuis le dataset avec une introduction et salutation très longues :

<QApair QAID="300643" QArel="R">

<QAquestion>السؤال السلام عليكم ورحمة الله وبركاته حياكم الله إخواني الكرام القائمين على هذا الموقع الممتاز وجزاكم الله عنا خير الجزاء على ما تقدمون من جهد. تحياتي للدكتور محمد عبد العليم ودائما إلى الأمام -إن شاء الله-. استشارتي تتعلق بابن أختي وهو يبلغ من العمر اثنتي عشرة سنة ويشكو دائماً من ضعف التركيز وعدم القدرة على مراجعة الدروس بشكل جيد مع الشرود الذهني خاصة أثناء الدروس مع القيام بباقي الواجبات الحياتية بشكل طبيعي كأي طفل; وعدم ظهور أعراض مرضية أخرى قمنا بعرضه على متخصص وكان التشخيص أنه اكتئاب أطفال. علما بأن هناك أكثر من حالة اكتئاب في الأسرة ووصف له الطبيب بروزاك 20 مجم قرصا واحدا يوميا فهل هذا التشخيص صحيح وإن كان صحيحا فما العلاج المناسب لحالته وإن لم يكن التشخيص سليما فما تشخيصكم للحالة وجزاكم الله خيراً.

</QAquestion>

<QAanswer>الإجابــة بسم الله الرحمن الرحيم الأخ الفاضل محمد عبد الله حفظه الله. السلام عليكم ورحمة الله وبركاته وبعد فنحن سعداء بتواصلك معنا وبثقتك في هذا الموقع ونسأل الله تعالى أن ينفع بنا جميعًا. ابن أختك هذا نسأل الله تعالى له العافية والصحة وأن يكون من الصالحين الناجحين. ضعف التركيز لدى الأطفال في الطفولة المبكرة أو الطفولة المتأخرة له أسباب بالطبع من الأسباب الرئيسية هو أن تكون المقدرات المعرفية للطفل أي مستوى الذكاء أصلاً لديه منخفض لكن لا أعتقد أن هذا ينطبق على هذا الطفل. السبب الثاني هو أن يكون الطفل مرتبطًا بالألعاب والأشياء غير الملزمة له في الحياة بشكل مستمر ولذا حين نطلب منه القيام بأي أمور جادة كمراجعة الدروس مثلاً تجده ينفر وتركيزه يضعف. السبب الثالث: هو أن بعض الأطفال كما تعلم لديهم متلازمة مرضية تعرف بمتلازمة (فرط الحركة وضعف التركيز) وقد لا تكون الحركة زائدة قد يكون الأمر منحصرًا فقط في ضعف التركيز. أما السبب الرابع فهو: الاكتئاب النفسي لدى الأطفال وحقيقة أصبح الاكتئاب النفسي أمرًا مزعجًا لدى الأطفال كنا لا نشاهد هذه الحالات لكن خلال العشر سنوات الأخيرة اتضح تمامًا أن الاكتئاب لدى الأطفال هو تشخيص نفسي حقيقي وأصبح الآن هنالك مختصون في الطب النفسي للأطفال يركزون على علاج هذه الحالات بصورة ممتازة ومتأنية جدًّا. هذا الطفل حفظه الله لديه إشكالية حول الدروس ولكن كما ذكرت وتكرمت في رسالتك أنه يعيش حياته الأخرى في مناحيها المختلفة بشكل طبيعي. أعتقد أن هذا الطفل ربما يكون له أو لديه شيء من النفور حول الدروس والمذاكرة لذا لا بد أن نبحث في الطرق التي تُحببه نحو دراسته وأكثر شيء يُحبب الطفل في الدراسة هو أن تنظم له وقته وأن تجعله يستمتع باللعب وبالراحة ويجلس للكمبيوتر تخصص له أوقاتًا وبعد ذلك تخصص له وقتًا للدراسة أو تخصص له وقتًا للدراسة أولاً ثم بعد ذلك تطلب منه أن يذهب وأن يلعب وأن يقوم بكل الأشياء التي يستمتع بها هذا مدخل جيد نحن كثيرًا ما نُلح على أطفالنا في موضوع المذاكرة والدراسة وننسى حاجاتهم الأخرى لهم حاجات نفسية ووجدانية لا بد أن توفر للطفل فهذا أمر أعتقد أنه يجب أن يُعطى أهمية خاصة. هذا الطفل حتى وإن كان مصابًا باكتئاب نفسي حقيقي إلا أن هذا المنهج سوف يفيده كثيرًا. الأمر الآخر هو: أن نُشعر هذا الطفل بأنه محبوب بأنه مُقدّر داخل الأسرة وأن نعطيه فرصة أن يلعب مع أقرانه خاصة مع الأطفال المؤتمنين هذا يعطيه حيزًا كبيرًا جدًّا من المتعة وهذا قطعًا سوف يحسن تركيزه كثيرًا. بالنسبة لموضوع البروزاك: البروزاك يعتبر من الأدوية السليمة جدًّا لعلاج الاكتئاب النفسي لدى الأطفال والشيء المتفق عليه الآن أن الأطفال من عمر ست أو سبع سنوات يمكن أن يتناولون البروزاك لكن يفضل أن يكون ذلك تحت إشراف طبيب مختص وأنتم -الحمد لله تعالى- أقدمتم على هذه الخطوة. فأنا (حقيقة) أوافق تمامًا على منهج الأخ الطبيب من حيث التشخيص ومن حيث طريقة العلاج وأرجو فقط أن يتم التركيز على المناهج السلوكية التي تحدثنا عنها باختصار. بارك الله فيك وجزاك الله خيرًا ونسأل الله له العافية والتوفيق والسداد.

</QAanswer>

</QApair>

-Puisque les utilisateurs des sites web doivent s'identifier pour qu'ils puissent poser leurs questions, ils utilisent des pseudos. Ces pseudos peuvent être 'أم فاضل' ou bien 'user123', pour cette raison nous avons supprimé les pseudos aussi manuellement.

-Nous avons éliminé toutes les paires dont la question ou la réponse était en anglais ou bien en français. Dans l'exemple suivant, la question est écrite par l'utilisateur en anglais :

<QApair QAID="68195" QArel="?">

<QAquestion>السلام عليكم ارجوبيان خطورة هذا-Abnormal area of increased signal intensity seen at the medial condyle of the lower end of the femur due to bone contusion. -Normal joint alignment. & articular surface -Loss of the normal contour with intermediate hyperintense signal form the lateral compartment of the anterior cruciate ligament due to its subacute partial tear. -A localized subchondral hypointense(T1); hypointense (T2&STIR) area is seen in the upper end of the tibia; posteriorly with a bone defect ; subchondral fracture is not excluded. . -Abnormal linear increased signal intensity seen at anterior horn of lateral meniscus doesn t reach the inferior articular surface consisting with grade I meniscal tea

</QAquestion>

<QAanswer>من التقرير المرفق هناك ما يشير الى تمزق جزئي في الرباط الصليبي الامامي ; اثار وذمة او استسقاء للجزء الداخلي لعظم الفخذ نتيجة اصابة ; هناك خلل في الجزء العلوي من الساق قد يشير الى وجود كسر ; الغضروف الخارجي مزق جزئيا . بناء على المعطيات يجب الخضوع لعملية تنظير لاستكشاف وضع المفصل الداخلي حيث تكون الصورة اوضح وقد تكون بحاجة لترميم الوتر والغضروف وتثبيت كسر الساق وعدم البحث عن حلول مريحة لك الان ومتعبة مستقبلا</QAanswer>

</QApair>

-Le dataset initial contient la notion de relevance -QArel- pour chaque paire, pour indiquer est-ce-que la réponse est directe, pertinente ou non pertinente par rapport à la question. A cause de ça, plusieurs paires se répètent.

Exemple : Si le contexte parle des femmes enceintes, parmi les 30 paires des questions-réponses associées au contexte, 10 paires peuvent avoir une relevance 'non – pertinent' car elles répondent aux questions sur des femmes qui allaitent. Les mêmes 10 paires réapparaissent avec relevance 'directe' avec un contexte qui s'intéressent aux femmes allaitantes.

-Plusieurs réponses sont très courte de genre 'لقد تمت الإجابة', 'نعم يمكن', 'ارسل مزيد من التفاصيل', 'سؤال غير واضح', ceci ne contient aucune information utile donc toutes les paires avec des réponses pareils sont supprimées. Voici un exemple avec une réponse très courte :

<QApair QAID="59602" QArel="I">

<QAquestion>السلام عليكم ورحمة الله وبركاته الضغط عندى اغلب الاوقات 130 90 واوقات 140 90 واعيش بكلية واحدة بعد استئصال الكلية اليسرى جراحيا فهل احتاج الى علاج ضغط مع العلم ان وزنى حوالى 102 كيلو والطول 170سم والسن 34

</QAquestion>

<QAanswer>سبق أن تمت الاجابة</QAanswer>

</QApair>

Avec tous ces changements et modifications, nous nous retrouverons avec seulement 12k de paires QA. Pour cette raison un deuxième passage manuel sur le dataset était nécessaire afin d'ajouter des nouvelles paires depuis des paires existantes, et agrandir la taille de dataset.

-Dans la majorité des questions, le patient décrit son cas médical avec trop de détails et explique son histoire médicale avec les traitements qu'il a pris et donc la question devient très longue. D'autres part le médecin ne répond pas directement à la question mais il donne une définition sur le cas médical avant de répondre à la question, parfois le patient demande l'avis de plusieurs médecins donc la réponse sera très longue.

Toute modification dans ce genre de paires nous a causé la perte de sens et/ou la relation entre la question et sa réponse. La paire suivante contient une question et une réponse très longues :

<QApair QAID="103580" QArel="?">

<QAquestion> عضو : السلام عليكم..انا شاب ابلغ من العمر 24 عاما... بداية توهاني في عمق الأمراض النفسية كان وعمري 16 سنة تقريباً في الصف الأول ثانوي حيث وانا جالس مع الأصدقاء وبدأ الحديث عن الكرة التي لا اعرف عنها شيئاً ولكني أفصح عن ميولي مجاراة وتقليداً للناس ولكن في ذلك التجمع حصل أمراً لم أحس به من قبل البتة وذلك حينما ألقى أحد الأصدقاء علي سؤال عن الكرة أرتبكت حينها وبدأت ملامح وجهي تتغير واحس بذلك بنفسي وأنظر إليه في وجوه الآخرين حتى تدارك أحد الأصدقاء الموقف ليقول لهم أني حديث عهدٍ بالكرة...كان هذا الموقف بداية لكل ما أعاني منه الآن...أصبحت بعد هذا أرتجف وأرتبك عند مواقف لا تستحق ابداً الخوف...حتى أنني أصبحت قلقاً جداً لما يحدث لي...بدأت أتغير لا أكثر الكلام أصبحت هادئاً لا أمازح الناس كل ذلك من أجل أن اتجنب الآخرين فلا أقع بإحراجات...علمت بعد ذلك أن مابي هو رهاب أجتماعي...أنتقل الأمر الى أبعد من ذلك ودخلت في متاهات الوسواس القهري - في العقيدة - بدأت الأسئلة تنهال علي و تلح علي إلحاحاً لا أستطيع أن أنفك عنها أو أن أدفعها أو أن انفصل عنها...وإن أجبت عنها ظهر لي سؤال آخر...حتى بدأت أحس بغصات في حلقي...هنا بدأت أفقد مشاعري وأحاسيسي وعواطفي حتى أعتزلت الناس ولا أخلاطهم إلا للضرورة إما عمل أو مسجد أو بيت...أنفتح في ذهني باب لم أجربه من قبل ولم أعهده البتة وصرت أسرح به وأحقق ما أشتهي وتطلبه نفسي من خلاله...علمت بعد ذلك أن ما بي هو أحلام اليقضة...إذا أردت النوم ثم أغمضت عيني ودخلت فيه أقوم مفزوعاً ودقات قلبي تتسارع وكأنه الموت... بقيت على هذا الحال الى اليوم...هناك موقف حصل لي لن أنسان أبدا وهو الذي جعلني أضع اليد على الجرح بكل وثوق...وهو أنني كنت جالساً مع أحد الزملاء في العمل وفجأة وفي جزء من الثانية تغير بي شيئ حيث بدأت أحس بمن حولي وأنني طبيعي الآن...ولكن ذلك لم يلبث إلا قليلاً ثم ذهب...بعد ذلك قرأت عن الفصام البسيط و وجدت أن أعراضه هي أعراضي وأن المريض بهذا الداء فاقد البصيرة لا يعلم أنه مصاب به لإنفصاله عن الواقع...وأحسن تعبير للحالة التي فاجأتي وانا في العمل هو أنني أتصلت بالواقع ثم أنفصلت عنه مباشرة...ولو قرأت عن الفصام قبل ذلك الحدث لما آمنت بأني مصاب به...ولكن لطف الله بي دلني على ذلك والحمدلله...أنا أعاني بلادة في العواطف والمشاعر والوجدان فليس لي صديق علاقتي معه قوية ذلك أنني لا استطيع أن أندمج معه عاطفياً فعقلي هو الوحيد الذي يعمل وعامة الناس لا يستخدمون عقولهم...كما أن تعبيري ليس بالتلقائي وإنما هو عويص...وأتمنى أن لا يقاس تعبيري في هذه الرسالة على حالتي...عموماً أنا لا أحس بنفسي ولا بالآخرين...حتى أثر ذلك بعلاقتي مع أمي و والدي وأخواني وأخواتي...وذلك بإهمالي لهم...وأنا لم ازهد بهم في يوم من الأيام إلا أنني فاقد للحس وللمشاعر وللعواطف التي تدفعني نحو مساعدة من يريد المساعدة بلا طلب وذلك مستحيل على من هم مثل إذ أن ذلك يحتاج الى ذكاء عاطفي...وعواطفي منطفأة -لن أقول معدومة تفائلاً- ... مع العلم بأني كنت إنسان إجتماعي 100%...وذكي ذكاء عقلياً وعاطفياً ...ولكن بالتدريج بدأت أنزل من أعلى الى أسفل...أريد علاجاً أرجوكم...آسف على الإطالة والسلام عليكم ورحمة الله وبركاته</QAquestion>

<QAanswer>: السيلام عليكم اخي الكريم :) نود ان نعبر لك اخي الكريم عن مدي تفهمنا لالامك ومعاناتك النفسيه وندعو الله ان نستطيع مساعدتك ;;;; مما ترويه لما يظهر ان لديك مرض الرهاب الاجتماعي والوسواس القهري ولكننا نود ان نوضح لك انه لا يمكن الجزم بدون الذهاب للفحص العيادي لدى طبيب امراض نفسيه وعصبيه وذلك لاجراء فحص شامل لك لتاكيد التشخيص ومن ثم وضع خطه علاجيه مناسبه ;..... اما احتماليه اصابتك بالفصام فهي ضئيله نوعا ما مما ترويه لنا ولكن لا يمكن التاكد الا من خلال : يتم التشخيص من خلال : طرح الاسئلة من قبل الطبيب النفسى او العاملين فى مجال الرعاية النفسية : حول الهلوسة أو الأوهام الإكتئاب وأعراض القلق وتعاطي المخدرات ; كذلك بعض اضطرابات الشخصية (على سبيل المثال اضطراب الشخصية الفصامية) واضطرابات النمو(على سبيل المثال اضطرابات طيف التوحد). اجراء الفحص الطبى الكامل : للتأكد من سلامة العقل حيث أن بعض أعراض الفصام

يمكن أن تحدث أيضاً في الأمراض العقلية الأخرى. فحص الصحة العقلية : هو لتحديد ما إذا كان الفرد يعاني من إضطراب فصامي عاطفي أو اضطراب ذهاني آخر مثل : -الاكتئاب - الهوس الاكتئابي - اضطراب القلق - تعاطي المواد المخدرة - اضطراب في الشخصية أوأي اضطراب مقترن بالسلوك الغريب والمزاج أو التفكيرمثل: - اضطراب الشخصية الحدية - اضطراب ذهاني آخر - وكذلك اضطراب لذا فنصيحتنا لك الان بضروره الذهاب في .(MPD) المعروف أيضاً باسم اضطراب في الشخصية المتعددة (DID) الهوية فصامي اقرب وقت للفحص لدى طبيب نفسي لتحديد المرض الذي تعاني منه تحديدا ومن ثم تحديد العلاج ;;;;;; علاج الوسواس القهري علاج اضطراب الوسواس القهري يمكن أن يكون صعباً والعلاج قد لا يؤدي إلى الشفاء. قد تحتاج العلاج لبقية حياتك.ومع ذلك يمكن الوصول إلى درجة تسمح للمريض بالسيطرة على أفعالة الناتجة عن الإصابة والهواجس. العلاج النفسي : لاضطراب الوسواس القهري وهناك يمكن أن يكون فعال. ويشمل العلاج السلوكي المعرفي إعادة التدريب على (CBT) نوع من العلاج يسمى العلاج السلوكي المعرفي التحفيز .(ECT) أنماط التفكير الخاص بالمريض ببحيث تخول دون العودة لممارسة السلوكيات القهرية. العلاج بالصدمة الكهربائية المغناطيسي عبر الجمجمة. التحفيز العميق للمخ - العلاج قد يستغرق مكان في جلسات فردية أو عائلية أو مجموعة. بعض الأدوية النفسية يمكن أن تساعد في السيطرة على الهواجس والدوافع من الوسواس القهريمثل مضادات الاكتئاب والتي قد تكون مفيدة لأنها قد تساعد على زيادة مستويات السيروتونين والتي قد تكون غير موجودة عند الاصابة بالوسواس القهري. مضادات الاكتئاب التي تمت الموافقة عليها أنافرانيل) (clomepramine) لعلاج الوسواس القهري ما يلي: - كلوميبرامين (FDA) على وجه التحديد من قبل إدارة الغذاء والدواء (serteraline) سيرترالين - (paroxetine) باروكسيتين - (fluxetine) فلوكستين - (fluxamine) فلوفوكسامين - (anavranil على جانب أخر . . . التعاطي مع الناس يعتمد في المقام الأول على اقتناع الأنسان الداخلي و هناك بعض التقنيات و النصائح العامة التي تساعدك بشكل كبير في هذا الشأن أهمها : - الانطباع الأول يدوم حقا : ابتسم ..فقط ابتسم عندما تقابل شخص جديد ..الابتسامه العريضه كافيه لترك انطباع محبب فى نفس من تقابله لأول مره ..كما انها ترفع من روحك المعنويه وتشعرك أكثر بالثقه وتنشط جهازك المناعى . - كن دائما ايجابى: فكر دائما بشكل ايجابى ..هذايجعلك دائما تشعر بشعور رائع ويبعد عنك الشيخوخه ..حتى فى أوقاتك العصيبه المرض أو اى ظروف سيئه ابق دائما ايجابى وتذكر ما وهبك الله من نعم عديده وسريعا ستعود لأفضل حال . - رائحتك الطيبه تعزز ثقتك بنفسك : بالطبع رائحتك الطيبه تجعلك دائما تشعر بالانتعاش والثقه وبأنك فى أحسن حالاتك ..وكذلك تجعلك شخص جذاب بالنسبه للآخرين ..احرص على الاستحمام بانتظام وارتداء ملابس نظيفه .. اقتنى مجموعه عنايه بها العطر والشامبو ومزيل رائحه العرق ..حافظ على نفسك منتعشا بالعنايه بفمك وأسنانك وزياره طبيب الاسنان كل فتره . - الملابس الرسميه طريقك للنجاح : أن تبدو قويا ;واثقا ; معتدل القامه وترتدى بدله سوداء أو رماديه ;ربطه عنق وجوارب متناسقه مع لون البدله ;شعرك مصفف بعنايه ولا تنسى أن يكون حذائك لامعا وابتعد تماما عن اى نوع من الحلى ..كل هذا هو طريقك للنجاح فى عملك وكسب ثقه رؤساءك وبالطبع ستحظى بفرصه العمل اذا كانت هذه أول مقابله عمل لك. - كوّن علاقات طيبة و (تعلم فن الاتيكيت): كلنا نحتاج العلاقات الطيبه لنشعر بالسعاده ..أن تكون عطوفا مع المحيطين بك يظهر لهم كم تقدرهم وتهتم لأمرهم ويجعلك تشعر ايضا انك شخص أفضل..تدرب على أن تكون مهذب وتساعد الآخرين ..وستذهلك النتيجه . - كن دائما وسطيا : لأن خير الأمور الوسط ...ممارسه اى شئ باعتدال هى الأفضل ..اذا بالغت فى فعل أمر ما ستكون النتائج عكسيه ...تناول طعامك باعتدال ...مارس الرياضه باعتدال (30 دقيقه)...ادخل على النت باعتدال ..اعمل باعتدال . للمزيد من المعلومات المفيدة ننصحك بمراجعة هذه المقالات : للرجال فقط :15 نصيحه تجعلك دائما فى المقدمة :http dailymedicalinfo.com articles a-194 هل القلق من مواجهة الناس يعيقك عن مواصلة حياتك :http dailymedicalinfo.com articles a-871 مع تمنياتنا بحياة ناجحة و سعيدة و مليئة بالثقة

</QAanswer>
</QApair>

-Le modèle choisi pour réaliser notre travail, implique que les entrées ne soient pas très longues. Pour cela, nous avons séparé toutes les questions incluses dans une seule question.

Exemple : ’ما هو الصرع ما هو علاجه ما هي أعراضه و السؤال الأخير كيف يتم التعايش معه‘ ceci est une question d’un seul utilisateur qui contient 4 questions imbriquées, elle devient : ‘ ما هو الصرع’, ‘ ما هي أعراض الصرع‘ ,’ما هو علاج الصرع ‘ ,’كيف يتم التعايش مع الصرع‘. Et donc chaque question sera associée à sa propre réponse.

L’exemple suivant présente un utilisateur qui pose 4 questions dans un seul passage. Cette question a été séparer en 4 questions différentes avec leurs réponses appropriées.

<QApair QAID="324102" QArel="I">
<QAquestion>السؤال السلام عليكم ورحمة الله وبركاته تحية طيبة إلى كل القائمين على هذا الموقع المبارك وكل من يساعد المسلمين ويعاونهم وجزاكم الله خيرًا. أنا عندي أكثر من استفسار بخصوص تأثير رياضة رفع الأثقال وكمال الأجسام على المفاصل: 1- كثيراً ما أسمع عن تأثير رياضة كمال الأجسام على المفاصل وخطورتها بسبب رفع أوزان عالية هل ممارسة رياضة رفع الأثقال وبناء الأجسام لها أضرار على المفاصل في المدى البعيد 2- هل يوجد فرق بين استخدام الكيابل (الأجهزة) أو الأثقال الحرة (الأوزان) في تأثيرها على المفاصل في المدى البعيد وما هو الأفضل لممارس هذه الرياضة على المفاصل أن يستخدم الكيبل أو الوزن الحر 3- هل تمارين إطالة العضلات لها تأثير سلبي على المفاصل على المدى البعيد 4- في الدرجات الثابتة يوجد ما يسمى المقاومة وهي زيادة قوة وثقل الحركة سواء بطريقة يدوية أو مغناطيسية هل لهذه المقاومة تأثير سلبي على المفاصل على المدى البعيد أعتذر على الإطالة وأكرر شكري ودعائي.</QAquestion>
<QAanswer>الإجابــة بسم الله الرحمن الرحيم الأخ الفاضل محمد حفظه الله. السلام عليكم ورحمة الله وبركاته وبعد: رياضة كمال الأجسام علم فيزيائي يعتمد على التدرج في رفع الأحمال جنبًا إلى جنب مع التدرج في بناء القوة العضلية ويتم ذلك في نادٍ ومع مدرب محترف هذا إذا كنت محترفًا لتلك الرياضة أما إذا كان الموضوع هواية وتريد بناء كتلة عضلية في جسمك بشكل سريع من خلال الأمينو أسيد وبودرة البروتين فهذا أمر غاية في الخطورة ليس على الكتلة العضلية فحسب ولكن على الكلى والكبد فيما بعد. ورفع أوزان دون تدرج أو استعداد يؤدي إلى تمزق بعض الألياف العضلية وإلى آلام في المفاصل واستخدام الأجهزة دون لياقة بدنية ودون تحضير عضلة القلب لتلك الأحمال أمر غير مستحب والتمارين الرياضة تقوي العضلات والمفاصل خصوصًا مع تناول كمية كافية من الكالسيوم وفيتامين (د) ومن خلال التغذية التي تعتمد على البروتين الحيواني من لحوم وأسماك ودجاج وليس على الأمينو أسيد وبودرة البروتين ومن خلال التعرض للشمس ومن خلال تناول كبسولات فيتامين د الإسبوعية جرعة 50000 وحدة دولية. والتدريب على الدراجات الثابتة بقوة أقل من قوة العضلات يرهق العضلات ويؤدي إلى تمزق الألياف العضلية ولا يفيد اللياقة البدنية في شيء وعمومًا محور الأسئلة كلها يدور حول القوة العضلية مقابل التمارين المناسبة لتلك القوة فإذا كانت اللياقة البدنية مرتفعة ومتدرجة فيمكنك زيادة شدة التمارين البدنية على الدراجات الثابتة والمتحركة دون خوف أو قلق وإذا أقدمت على الصالات الرياضية بدون إحماء أو لياقة بدنية مرتفعة فلن تجني إلا الإرهاق والتعب مع البعد عن البروتين الصناعي والاعتماد على البروتين الحيواني والفواكه في الغذاء. وفقك الله لما فيه الخير.
</QAanswer>
</QApair>

4.1.2 Prétraitement automatique

Le prétraitement dans la langue arabe diffère de celui de la langue anglaise, le but de ce prétraitement est de normaliser et standardiser le format des mots pour qu'ils soient plus uniforme. Les étapes sont les suivantes :

a-Elimination des signes de ponctuation :

Parfois les paires de question-réponse contient ponctuations spéciales, c'est pour cette raison que nous avons éliminé tous les caractères spéciaux sauf «?» dans les questions.

b- Formater les mots :

-Un des défis de langue arabe c'est le nombre important des dialectes, ceci résulte parfois le changement des lettres et quelques chiffres, donc pour un format uniforme on doit changer les lettres ainsi que les chiffres arabes orientaux :

- 'أ, إ, آ' devient 'ا'

-'ة ' devient 'ه'

-‘٣’ devient ‘3’

مدينة ⟶ مدينه

أصل ⟶ اصل

- Il faut éliminer l’accentuation qui utilise des marques comme des guides phonétiques (التشكيل)

حَرَكَات ⟶ حركات

السَّمَاء ⟶ السماء

- Il faut éliminer tatweel dans certains mots arabes (التطويل)

العـــربية ⟶ العربية

4.2 Adaptation et traitement du dataset anglais

Le dataset anglais comme mentionné auparavant est composé de 6 catégories (produits bébé, produits de santé, accessoires téléphones, produit de bureau, produits de maison et cuisine et outils de maison et amélioration de l’habitat). Ces catégories étaient choisies car elles ne contiennent pas beaucoup de références des produits, exemple : XRT200, B00004W4UK. Après avoir effectué le prétraitement nécessaire décrit dans ce qui suit, nous avons appliquer les mêmes conditions sur la longueur sur les entrées. Le tableau -7- donne des détails sur la taille initiale et la taille après l’application de ces conditions pour chaque catégorie de produit :

Catégorie	**Nombre de paires initiale**	**Nouveau nombre de paires**
Bébé	28933	13126
Santé	80496	35580
Accessoires téléphones	85865	40583
Produit de bureau	43608	19845
Cuisine et outils de maison	184439	84922
Outils de maison et amélioration de l’habitat	101088	55944
Total	**524429**	**250000**

Tableau 7: Détails sur le dataset anglais

Nous avons ensuite consacré 80% de dataset pour l’entrainement et le reste pour le test comme décrit le tableau -8- :

Dataset	Entrainement (80%)	Test (20%)	Total (100%)
Amazon QA	180000	45000	225000

Tableau 8 : Détails d'utilisation de dataset anglais

4.3 Prétraitement automatique

Un corpus de données ou dataset est toujours traité et nettoyé avant de l'utiliser dans le Machine learning ou Deep learning, la qualité des données peut affecter les résultats de l'apprentissage du modèle donc cette étape et cruciale et très importante. Elle consiste à transformer un texte en format utilisable et adapté et standard.

Cette étape consiste à standardiser le format des mots c-à-d convertir les mots vers un format uniforme vu la nature du dataset :

a-Elimination des signes de ponctuation :

Parfois les paires de question-réponse contient ponctuations spéciales, c'est pour cette raison que nous avons éliminé tous les caractères spéciaux sauf «?» dans les questions. Nous n'avons pas éliminer les stops-words car les questions résultantes ne seront pas de grand sens.

b-Mise en minuscules

Puisque la machine est sensible à la case, ceci peut causer des problèmes par la suite, donc cette étape et aussi importante pour les langues latines seulement.

Exemple : 'Thank you' et 'thank you' sont considéré comme deux mots différents dans notre vocabulaire si on n'applique pas la mise en minuscule.

c-Elimination des abréviations :

Dans cette étape, on doit éliminer les abréviations pour que l'ordinateur peut reconnaitre les mots similaires et aussi vu la nature de dataset informel, chaque client écrit sa question ou réponse de sa façon, cette étape empêche que le même mot soit compté deux fois dans le vocabulaire :

can't ⟶ cannot

I'm ⟶ I am

Can u ⟶ can you

9'' ⟶ 9 inches

d-Elimination des URL et noms propres :

Nous avons supprimé tous les noms propres avec la bibliothèque Spacy avant la mise en minuscule, exemple : ‘Thank you Emily for your question.’, ‘Adam, the product you are looking for is no longer manifactured’. Ainsi, tous les liens vers des sites web ou bien vers d’autres produits dans le site d’Amazon ont été supprimés.

5. Modèle de génération de question proposé

Pour réaliser notre travail, nous avons adopté un modèle basé sur les réseaux de neurones. Il s’agit d’un encodeur-décodeur avec mécanisme d’attention. Tout modèle d’apprentissage automatique est basé sur un entrainement en utilisant un dataset, et pour garantir des meilleurs résultats nous avons appliqué le prétraitement nécessaire, qui est considéré comme une étape cruciale et importante, sur les datasets choisis auparavant. Dans ce qui suit, nous détaillerons les différents composants de notre générateur de questions.

5.1 Modèle Encodeur/décodeur avec mécanisme d’attention

Le modèle Coder-Décodeur avec mécanisme d’attention est le plus adapté dans les travaux connexes vus dans un chapitre précédent. Nous avons conçu deux modèles : un pour la langue arabe et l’autre pour la langue anglaise. La figure -19- représente l’architecture globale de l’encodeur décodeur avec un mécanisme d’attention :

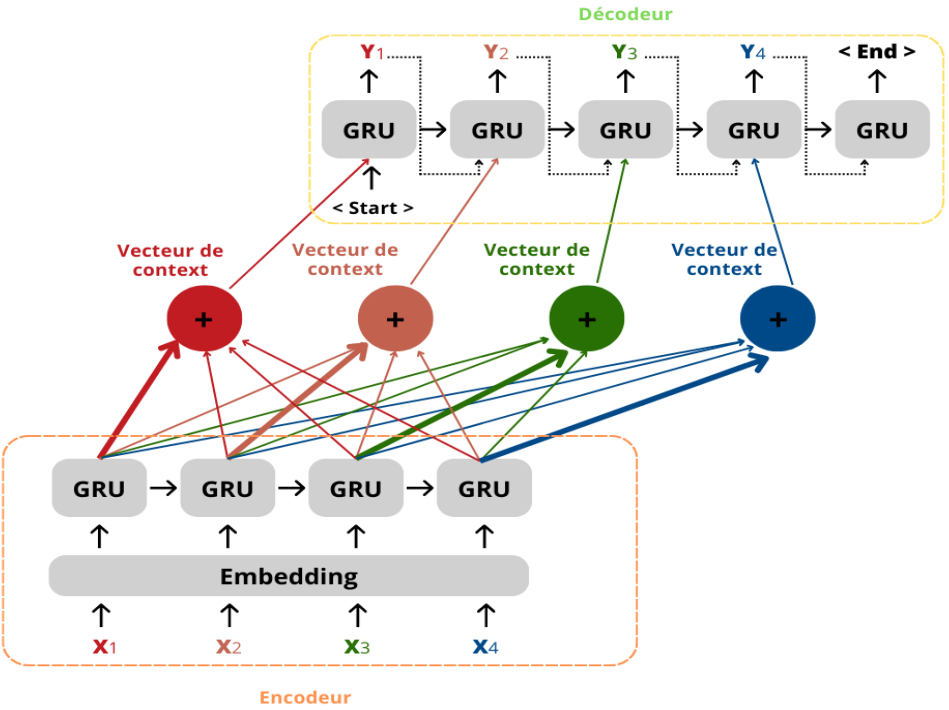


Figure 19: Architecture globale d'un encodeur décodeur avec mécanisme d'attention

Dans cette figure, les lignes fines contribuent moins à la prédiction des mots et les lignes épaisses contribuent le plus. Tous les états cachés de l'encodeur avec les états cachés du décodeur sont utilisés afin de générer le vecteur de contexte.

Tout modèle de l'apprentissage profond, nécessite la définition des hyper paramètres. Ils sont utilisés lors de l'entrainement et ne seront pas ajustés par le modèle.

5.2 Définition hyper paramètres

Fonction de perte (Loss function)

Le modèle d'un encodeur décodeur a pour but de minimiser la valeur de perte lors de l'entrainement, et donc minimiser l'erreur de prédiction. Si le mot généré s'écart trop de mot réel, la valeur de perte sera grande et vice versa. A l'aide de cette fonction, le modèle apprend à réduire l'erreur et donc les mots prédits seront plus proches aux mots réels.

Nous avons choisi 'Sparse Categorical Crossentropy', cette dernière mesure la distance entre le mot a généré et le mot réel (la probabilité de mot à générer par rapport à toutes les classes). Le

choix de cette fonction dépend du format des sorties [64]. La formule de la fonction de perte 'Sparse Categorical Crossentropy' est :

$$Loss = -\sum_{i=1}^{N} y_i \,.\log(\hat{y}_i) \qquad (12)$$

N = Nombre de classe (35 milles)

y_i= la classe réelle (le vrai mot)

$\hat{y}i$= la probabilité de Softmax pour la i éme classe par rapport à N (la probabilité du mot à générer par rapport aux mots restants du vocabulaire)

Fonction d'optimisation (Optimizer)

Les optimiseurs ont pour but de mettre à jour les paramètres d'un modèle dont ils convergent vers la valeur la plus optimale (d'où la valeur prédite est très proche à la vraie valeur). Cette fonction d'optimisation aide à savoir comment modifier les poids et le taux d'apprentissage du réseau de neurones pour réduire la fonction de perte [65].

Nous avons choisi 'ADAM', le Adaptive Moment Estimation pour ces avantages car il combine les meilleures propriétés des autres optimiseurs. Selon la littérature, ADAM est recommandé comme le meilleur optimiseur même avec ses paramètres par défauts.

Taux d'apprentissage

C'est un hyperparamètre très crucial et difficile à choisir, il contrôle le degré de modification de modèle par rapport à l'erreur à chaque fois que les poids sont ajustés et mis à jour [66]. Afin de déterminer le taux d'apprentissage approprié, nous avons tester différentes valeurs (0.005, 0.001, 0.003…) en changent à chaque fois la fonction d'optimisation. Avec chaque modification, nous avons entrainé le modèle pour suivre et observer les résultats de la fonction de perte pour déterminer les meilleurs paramètres. Après l'expérimentation, nous avons choisi la valeur de 0.001 avec l'optimizer ADAM.

Taille de lots (Batch size)

Représente le nombre d'échantillons d'apprentissage à traiter, avant la mise à jour des paramètres du modèle. Le modèle s'entraine au fur et à mesure sur ces lots pour mieux prédire et donc minimiser la fonction de perte [67]. Dans notre cas, nous avons choisi 64 pour les limites imposées par le matériel informatique. Un nombre plus grand des échantillons nécessite plus d'espace mémoire.

Taille de l'embedding

C'est la taille de l'espace vectoriel dans lequel nous avons représentés les mots. Généralement cette valeur varie entre [50 et 300], Nous avons choisi 256.

5.3 Encodeur

C'est la première partie du modèle, il est composé de la couche embedding et le générateur de contexte. Ce vecteur contient les informations nécessaires relatives aux entrées, qui va être par la suite utiliser par le décodeur comme le premier état caché afin de débuter la génération des questions.

- **Couche embedding**

Nous allons construire notre propre embedding, pour cela nous commençons par :

- Extraire le vocabulaire depuis le dataset (anglais et arabe).
- Attribution d'une valeur entière pour chaque mot et donc chaque phrase peut être converti à une séquence de numéro dont chaque numéro représente un mot dans cette dernière.

Dans un premier temps, nous avons essayé d'utiliser un embedding pré-entrainés pour arabe[10] et pour anglais [11]. Les embeddings [68] sont une technique qui transforme un texte brut vers un vecteur numérique, ce mappage sert à capturer les relations sémantiques et contextuels des mots dans un corpus, alors les embedding pré-entrainés sont appris dans une tâche, entrainés sur des larges datasets et enregistrés pour être utiliser par la suite pour résoudre une autre tâche similaire. Ils sont une forme d'apprentissage par transfert.

L'utilisation de ces embedding a causé la répétition des mots dans les questions générées après la 20^{éme} époques.

Pour résoudre le problème de répétition qui existe dans les modèles séquence à séquence, nous avons ajouté le mécanisme de coverage [69]. Ce mécanisme garde une trace du nombre de fois qu'un mot cible est généré et empêche de répéter les mêmes mots.

Mais le problème a persisté et n'a pas été résolu. Donc nous avons utilisé une couche embedding qui utilise les réseaux de neurones afin de créer les embedding pour les mots lors de l'entrainement du modèle.

[10] http://vectors.nlpl.eu/repository/20/31.zip

[11] http://vectors.nlpl.eu/repository/20/40.zip

- Nous commençons par une représentation de chaque mot en vecteur binaire (one-hot encoding), puis passer par un réseau de neurones qui permet une représentation réduite, c-à-d multiplier la représentation one-hot par une matrice de poids W (dont les poids sont initialisés aléatoirement et ajuster au fur et à mesure de l'entrainement). Cette méthode permet de garder les liens sémantiques entre les mots.

L'encodage one hot c'est quand chaque valeur est représentée sous la forme d'un vecteur binaire composé de toutes les valeurs nulles, à l'exception de l'indice dans le vocabulaire, qui est marqué d'un 1.

Nous avons eu un vocabulaire total entre le vocabulaire d'entrée et sortie, plus de 101000 mots uniques pour anglais et 52000 mots pour arabe. Chaqu'un de ces mots va subir le changement détailler dans l'exemple suivant pour passer de son format brut vers un vecteur numérique qui capture le sens sémantique.

Exemple : Supposons que nous avons un vocabulaire de 3 mots : ['rouge', 'vert', 'bleu'] chaque mot dans ce vocabulaire sera représenté par des 0 et 1 seulement dans son indice dans le vocabulaire. Alors, rouge = [1, 0, 0], vert = [0, 1, 0] et bleu = [0, 0, 1]

Voici un exemple pour montrer le travail de la couche embedding de l'encodeur :

$$(1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0) \times \begin{bmatrix} 0.5\ 0.6\ 0.32 \dots 0.41\ 0.9 \\ 0.12\ 0.2\ 0.88 \dots 0.45\ 0.67 \\ \vdots \\ 0.33\ 0.45\ 0.66 \dots 0.04\ 0.56 \\ 0.12\ 0.13\ 0.51 \dots 0.92\ 0.13 \end{bmatrix} = (0.5\ 0.6\ 0.32 \dots 0.41\ 0.9)$$

One-hot: 1 * V **Matrice: V * e** **Vecteur WE: 1*e**

Exemple de mappage de vecteur one-hot vers vecteur WE

Avec :

V : la taille de vocabulaire (53 milles)

e : taille de vecteur embedding = 256

- **Couche GRU**

Notre choix est porté sur les GRU car ils ont montré de meilleurs résultats par rapport aux cellules LSTM [70]. Les GRU sont rapide en termes de temps de calcul et sont plus simples et plus efficaces. Cette couche prend le texte d'entrée pour qu'il soit codé en un vecteur de contexte, et ce dernier sera utilisé par la suite par le décodeur.

Le mécanisme d'attention a permis de mettre en valeur toutes les informations dans les états cachés dans différent pas de temps. Alors Afin de générer un mot, nous devrons faire attention à chaque mot de la séquence d'entrée. Cette attention est exprimée par des poids.

L'encodeur génère un score avec la fonction Softmax pour chaque état caché. Et donc l'état caché qui aura besoin d'attention c'est celui avec le score le plus élevé.

Par la suite le vecteur de contexte est calculé en suivant cette formule :

$$VC = \sum_{i=1}^{N} w_i . h_i \qquad (13)$$

Avec :

w_i = le poids de l'état caché h_i

h_i = l'état caché a un instant i

5.4 Décodeur

Pour ce dernier, on distingue 2 couches : une couche GRU et une couche Softmax.

La couche GRU fonctionne de la même manière que celle de l'encodeur, sauf que le décodeur prend en entrée : le vecteur de contexte et le dernier état caché de l'encodeur comme état initial afin de générer le premier mot.

Par la suite chaque mot à un pas de temps *t* prends la sortie du nœud précédent à l'instant *t-1* et le vecteur de contexte générer par le mécanisme d'attention sans oublier l'embedding du i éme mot de la question.

Et la couche Softmax est utilisée pour prédire l'entier avec la plus grande probabilité et chaque entier représente un mot dans le vocabulaire.

6. Recherche par faisceaux (Beam search)

Comme l'objectif principal de notre travail est de générer plusieurs questions pour une seule requête, nous avons pensé à ajouter un algorithme qui utilise la recherche par faisceaux. Le but de cette méthode est de choisir les N-meilleures séquences, le décodage qui utilise le beam search considère les probabilités de toute la séquence générée au paravent. Cet algorithme a un hyperparamètre important, appelé 'beam-width' qui déterminer le nombre de séquence à générer. Donc vers la fin, nous allons avoir N-séquences (N = beam-width) avec les probabilités les plus élevées [71],[72]. La figure -20- détaille la recherche par faisceaux, avec un beam-width=2 :

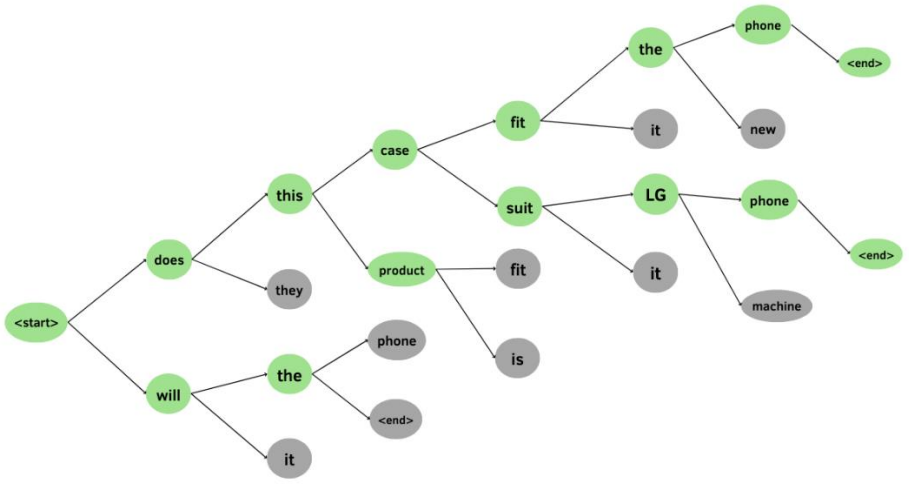


Figure 20: Exemple de beam search avec beam width=2

Dans ce cas la réponse était : yes this phone case fit all the LG phones.

Donc les deux questions générées avec le beam-search étaient :

- Does this case fit the phone?
- Does this case suit LG phone?

Nous allons détailler la recherche par faisceaux :

-1ére étape : pour commencer la génération, la première entrée sera <start> qui indique le début de la génération. Ensuite en utilisant la fonction Softmax du décodeur, on ne garde que les 2 premiers mots (depuis le dictionnaire qui contient 53K mots uniques) avec les probabilités les plus élevé (does et will).

-2éme étape : Afin de générer le 2éme mot, nous devrons prendre les 2 premiers mots générés comme entrées pour trouver les 2 premières séquences dans cette étape. Nous utilisons la couche Softmax pour trouver les 2 meilleures séquences depuis un vocabulaire de plus de 106K mots. Nous utilisons la probabilité conditionnelle afin de décider les mots de la deuxième position, prenant en compte les 2 premiers mots dans la première position. Pour trouver les 2 combinaisons avec les scores les plus élevés.

Une combinaison de mot peut être rejeté si une nouvelle combinaison aura un score de probabilité plus élevé. Dans cet exemple, les deux séquences finales étaient générées depuis le mot ‘does’ et le mot ‘will’ et tous ce que vient après a été rejeté.

-3^{éme} étape : Nous répétons ce processus jusqu'à ce que nous atteignions le mot <end> ou jusqu'à la fin de la séquence. Nous aurons à la fin deux séquences avec les plus hautes probabilités parmi toutes les combinaisons possibles. Ces séquences peuvent différer en longueur.

7. Exemples de questions générées

Les tableaux suivants 9 et 10 représentent des exemples des questions générées après l'entrainement de modèle pour la langue arabe et anglaise respectivement. Ces paires se figurent dans la partie test du dataset :

7.1 Arabe

La réponse	La vraie question	La question prédite
يحتوي الدوفاستون على الديدروجيسترون وهو يستخدم لعلاج اضطرابات الدوره الشهريه نزيف الرحم الغير طبيعي العقم والاجهاض المتكرر لا يجب التوقف عن تناول الدواء دون استشاره الطبيب لانه قد يؤدي الى اضطرابات في الدوره الشهريه	هل يمكنني التوقف عن تناول دوفاستون وهل لديه اعراض عند التوقف عنه	ما هو الدوفاستون ؟
يفضل مراجعه طبيب مختص كي يتم التشخيص الدقيق العلاج 1 حافظ على نظافه المنزل والمنطقه المحيطه 2 يجب ان يكون تركيزه 5%يوضع الكريم على المنطقه المصابه لمده 8 12ساعه يفضل قبل النوم ثم يغسل 3 تجنب تعرضه للوجه والعينين ملاحظه استخدم الدواء تحت اشراف الطبيب او الصيدلاني	ما هو علاج الجرب	ما هو علاج الجرب
لتهاب البروستات المزمن الجرثومي يجب معالجته على الرغم من صعوبه العلاج الاقدر على متابعه حالتك هو اختصاصي المسالك البوليه المتمرس لان العلاج يحتاج الى دوره متكامله من العلاجات سواء بالادويه او المساج او العلاج الطبيعي	انا اعاني من التهاب البرستاتا المزمن الجرثومي وتعالجت له كثيرا دون فائده هل له علاج نهائي وهل يتسبب في المستقبل بتضخم او سرطان البروستاتا	لدي التهاب فى المعده و لدي اعراض سرطان

اعاني من الم في المفاصل ولا استطيع المشي	انا امراه 49 سنه مريضه سكر واشعر بالم فى ركبتي اليمنى واليسرى ولا استطيع تحميل جسمى على دوكسه القدم الرجاء المساعده وما الاطعمه التى اكلها والتى ابعد عنها	عليك مراجعه جراح عظام لابد من الفحص السريري قد تحتاجين صوره اشعه للركبه والكاحل وعمل فحص وظائف الكلى وتخفيف الوزن ان وجد اخصائي التغذيه هو الاجدر بالاجابه على سؤالك حول الاطعمه التي تناسبك بعدان تذكري وزنك وطولك وحالتك الصحيه وطبيعه نشاطك اليومي
لدي التهاب في الانف هل هو التهاب الجيوب الانفيه	اصبت بالزكام الخميس بعد تناول الادويه تحسنت ولكني اصبحت الاثنين اشعر بالم يشبه الصداع في نصف وجهي الايسر وافراز ماده صفراء تمحيل للون الاحمر	هذا تاثير الالتهابات الجيوب الانفيه تحتاج الى تصوير طبقي للجيوب الانفيه

Tableau 9: Exemples de questions générées en arabe

Pour le décodage avec la recherche par faisceaux, nous pouvons générer plusieurs séquences en sortie pour chaque question, Nous montrons quelques exemples pour l'arabe :

- La réponse : يحتوي بيرياكتين على الريبافيرين وهو يستخدم كمضاد للفيروسات يستخدم لعلاج التهابات الجهاز التنفسي الفيروسيه في ويستخدم مع الفا انترفيرون لعلاج الالتهابات المزمنه والتهاب الكبد الفيروسي ج

- La vraie question : ما دواعى استعمال علاج بيرياكتين

-Les questions générées (avec beam 4) :

-ما هي تاثير دواء على المدى البعيد

-اريد ان استخدم دواء لعلاج حب الشباب و هل هو مفيد في الصيدليات

-ما هو تاثير دواء بيرياكتين مضر ام لا

-هل دواء له تاثير على المستوى البعيد

- La réponse : التهاب البروستات المزمن الجرثومي يجب معالجته على الرغم من صعوبه العلاج الاقدر على متابعه حالتك هو اختصاصي المسالك البوليه المتمرس لان العلاج يحتاج الى دوره متكامله من العلاجات سواء بالادويه او المساج او العلاج الطبيعي

- La vraie question : انا اعاني من التهاب البرستاتا المزمن الجرثومي وتعالجت له كثيرا دون فائده هل له علاج نهائي وهل يتسبب في المستقبل بتضخم او سرطان البروستاتا

-Les questions générées (avec beam 7) :

- حامل في منطقة و لدي اعراض سرطان

-جميع اعراض التهاب في المعدة و لدي اعراض سرطان

-قبل فترة احس بالم في المعدة و لدي اعراض سرطان

-عمري 19 يعاني من مرض سرطان الثدي

اعاني من الام في المعدة و لدي اعراض سرطان-

عندي التهاب في المعدة و لدي اعراض سرطان-

لدي التهاب في المعدة و لدي اعراض سرطان -

- La réponse : هل الاسنان اماميه ام خلفيه المفروض علاج الاعصاب اولا و من ثم التفكير بالخلع الا اذا كانت الاسنان معدومه

- La vraie question : ارجو عرض الطفل لاخصائي اسنان اطفال و ليس طبيب عام للمساعده و التخفيف من المشاكل المستقبليه ابني عمره خمس سنوات واسنانه متسوسات وقد قلعنا له ثلاث من الفك العلوي الان الم بين كل فتره واخري لا ادري ماهو الحل لهذا التسوس والالم

-Les questions générées (avec beam 10) :

عمري 25 سنه واعانى من ثلاث سنوات استخدم علاج حول الاسنان او علاج لهذه الحاله -

ضرس العقل واريد التخلص منها -

قمت بخلع عندها 17 سنه تقريبا لكن لا استطيع تنظيف عصب السن واريد التخلص منها -

لدي مشكله في اللثه واريد علاج لهذه الحاله -

عندي مشكله في اللثه واريد علاج لهذه الحاله -

اعانى من التسوس واريد علاج لهذه الحاله-

عملت تنظيف عصب الاسنان الاماميه فوق العينين واريد حل لهذه الحاله-

لدي مشكله في اللثه واريد علاج لهذه الحاله-

اعاني من التسوس عن السن واريد علاج لهذه الحاله-

توجد مشكله فى الاسنان الاماميه واريد حل لهذه الحاله-

7.2 Anglais

La réponse	La vraie question	La question prédite
there is no expiration date there is an assembled date on my of 2006	what does the date on the bottom of the box indicates? does it has an expiration date?	what is the expiration date on this product?
they do not rust but they are really really small	does it rust ?	does these rust ?
yes it works on any audio from your iphone and works on all iphones up until the	will it broadcast tunein radio from an iphone 4s?	will this work with an iphone 5s and ipad air?

iphone 5 because of the new port on that model		
it does not contain the only ingredients are cold liver oil natural lemon essence natural peppermint essence alpha tocopherol very good product	does this item contain soy if so from what ? does this contain ala ? if yes can you describe from what ?	does this contain any additives products ?
i do not see why not it is a very sturdy plastic i do not think i would put in in the microwave though	can i measure boiling water in it since its not glass ?	can this unit be used in the microwave ?
i measured them to be in the middle area 45 holes in the outer ring 42 holes	what is the size of the holes ?	what is the size of the ring ?

Tableau 10: Exemples de questions générées en anglais

Pour le décodage avec la recherche par faisceaux, nous pouvons générer plusieurs séquences en sortie pour chaque question, Nous montrons quelques exemples pour l'anglais :

- La réponse: they do not rust but they are really really small

- La vraie question : does it trust **?**

- Les questions générées (avec beam 4) :

- do these rust ?
- does the product rust ?
- i have read that they rust
- are these really truly stainless steel when not what finish ? and will they rust ?

- La réponse: i do not see why not it is a very sturdy plastic i do not think i would put in in the microwave though

- La vraie question: can i measure boiling water in it since its not glass ?

- Les questions générées (avec beam 7) :

- could this be mounted over a microwave ?
- can this unit be used in the microwave ?
- can this unit use for microwave ?
- is it microwaveable ?

- could this be used in the microwave ?
- are the containers microwave safe ?
- would this work in the microwave ?

- La réponse: the product works with any device that has micro usb

- La vraie question: will this adapter work with samsung galaxy ?

- Les questions générées (avec beam 10) :

- does it works with an iphone 6 ?
- does it work with android
- would this work with the samsung s5 ?
- how about for the iphone 6 ?
- is there any usb for microsoft surface tablet
- can i use it for the iphone 6 ?
- has anyone tried this for the iphone 6 ?
- can i use it for the iphone 6 ?
- this charger works with the samsung note 3 ?
- works with kindles ?

8. Conclusion

Dans ce chapitre, nous avons expliqué tout le travail fait sur les deux dataset afin de les utiliser avec la tâche de génération de question, nous avons proposé un modèle encodeur décodeur avec mécanisme d'attention afin de générer plusieurs exemples de questions pour la même entrée. L'entrainement des deux modèles était fait en parallèle sur deux datasets distincts pour Arabe et Anglais.

Dans le chapitre suivant, nous allons évaluer notre générateur de question avec une évaluation quantitative automatique avec des métriques telles que : BLEU, METEOR et ROUGE ; ainsi qu'une évaluation qualitative humaine et comparer nos résultats avec ceux de la littérature.

Chapitre IV : Expérimentation et Discussion

1. Introduction

Comme la langue arabe n'a pas eu sa part dans le domaine de TAL tandis que c'est une langue très riche, nous avons opté pour un modèle qui a été entrainé sur deux datasets distincts – arabe et anglais- afin d'évaluer sa performance. Le modèle avec le dataset anglais (Amazon QA) nous a permis de comparer le résultat final de notre modèle avec les résultats existants. Afin d'obtenir ces résultats nous avons fait plusieurs tentatives et expériences pour déterminer les meilleurs paramètres et leur impact sur le résultat final de l'évaluation quantitative automatique et qualitative humaine.

2. Détails d'implémentation

Chaque modèle d'apprentissage automatique à besoin de passer par les deux phases d'entrainement et de test. Nos machines personnelles n'ont pas servi pour effectuer ce travail, pour surmonter cette difficulté nous avons utilisé un serveur fournit par Google, appelé Google Colaboratory, nous avons exploité ce serveur avec 12G de RAM pour pouvoir entrainer et tester notre modèle. Parmi les problèmes rencontrés en utilisant Colab, c'est le temps limité d'utilisation de GPU (qui sert comme accélérateur graphique) de 12 heures et parfois moins, car nous utilisons une version gratuite de Colab. Pour résoudre ce problème nous étions obligées de créer plusieurs compte Google pour avoir accès à leur compte Colab et l'espace gratuit de 15G fournit par Google Drive et accélérer l'entrainement. Pour réaliser ce modèle, nous avons utilisé la version 2.3 de Tensorflow avec Keras 2.1. Ces versions antérieures non pas posées de problèmes lors de l'exécution.

- Premièrement, nous avons chargé le dataset pour que le modèle puisse l'utiliser. Puisque la machine ne peut pas utiliser des textes bruts quelques changements étaient nécessaires, on ajoute un espace entre le mot et la ponctuations ' ?', nous ajoutons '<start>' et '<end>' au début et la fin des phrases.

- Ensuite nous effectuons la tokenisation qui consiste à séparer les mots et la ponctuation d'une phrase, chaque élément de cette dernière – mot ou ponctuation- est appelé token. Chaque token est converti à la valeur de l'entier depuis un vocabulaire qui contient des mots uniques, qui lui correspond.

-La longueur de nos phrases est souvent incohérente et lorsque les GRU traitent les données par lots, nous devons assurer que les phrases sont de la même longueur. Alors nous ajoutons le 'padding' qui est l'ajout dans une valeur non informative à la séquence, Nous avons ajoutés la valeur '0' à droite des phrases pour que les phrases soient toutes de la même longueur.

- L'encodeur de son tour prends les phrases traitées et les faire passer par une couche embedding qui converti la séquence des entiers en des vecteurs des mot pour que l'encodeur puisse capter les liens et les informations sémantiques entre les différents éléments de la phrase. La couche GRU de l'encodeur code ces informations dans des vecteurs appelé état caché.

Notre modèle est une amélioration d'un modèle basique de séquence à séquence. Nous ajoutons le mécanisme d'attention pour adresser la limitation des modèles basiques qui ne peuvent pas retenir les informations des séquences longues.

-Bahdanau attention [73] ou bien l'attention additive c'est le type d'attention adopter par notre modèle. Ce type effectue une combinaison linéaire des états de l'encodeur et le décodeur. Pour créer la couche d'attention est composé de :

- **Score d'alignement** : Bahdanau attention utilise les scores d'alignement concaténé avec la fonction « tanh » pour évaluer si l'entrée de la position 'j'dans la vraie question s'accorde bien avec la sortie de la position 'i' dans la question prédite. Ce score est basé sur l'état caché du décodeur précédent s_{i-1} et l'état caché l'entrée de la phrase source h_j. Avec : e_{ij} = a (s_{i-1}, h_j).

- **Les poids d'attention** : On applique la fonction Softmax sur les scores d'alignement afin de les normalisé et ils peuvent être traiter comme des probabilités. Plus le poids d'attention de la séquence d'entrée est élevé, plus son influence sur la prédiction du mot cible sera élevée.

$$\alpha_{ij} = \frac{exp(e_{ij})}{\sum_{k=1}^{N} exp(e_{ik})} \qquad (14)$$

Où α_{ij} représentent les poids d'attention.

- **Le vecteur de contexte C_i**: ce vecteur est utilisé pour calculer la sortie du décodeur, c'est la somme pondérée des poids d'attention et les N états cachés de l'encodeur.

$$C_i = \sum_{j=1}^{N} \alpha_{ij}.h_j \qquad (15)$$

- Nous passons maintenait au décodeur, il dispose d'une couche embedding et une autre GRU. Le vecteur de contexte est ajouté avec l'état caché du décodeur précédent avec la sortie du décodeur au pas de temps t-1, sont utilisé pour prédire un mot. $s_i = f(s_{i-1}, C_i, y_{i-1})$ avec :

s_{i-1} :l'état caché précédent du décodeur

C_i : vecteur de contexte

y_{i-1} : la sortie de décodeur (mot prédit) précédente

- Le process de calcul de scores d'alignement, le poids d'attention et le vecteur de contexte se répète pour chaque instant de temps. Il s'arrête quand '<end>' est produit dans la sortie de décodeur ou quand il atteint la longueur maximale de la sortie définie précédemment.

3. Résultat

Afin d'évaluer la qualité des questions générés depuis le modèle après son entrainement, nous effectuons une évaluation automatique avec des métriques tels que : Bleu, Rouge et Meteor et une autre évaluation humaine faite par des experts de langue prenant en considération deux critères la correction grammaticale (grammaticalité et fluidité) et la pertinence.

La figure - 21- détaille le processus d'évaluation de notre modèle encodeur décodeur avec attention.

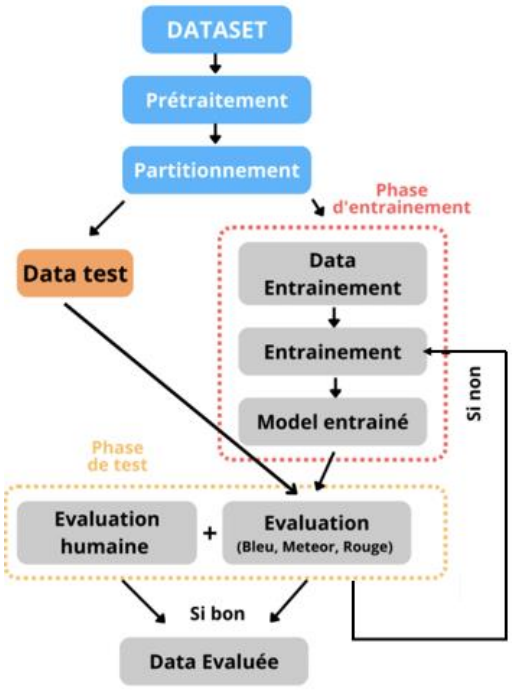


Figure 21: Processus d'évaluation

3.1 Evaluation automatique

Pour pouvoir évaluer la qualité des questions résultantes de notre modèle, nous utilisons des métriques utilisées dans les travaux de la littérature pour comparer les résultats. Notons que aucun travail est dédié à la langue arabe donc nous comparons la performance de notre modèle avec le dataset anglais communautaire de Sem-Eval2016 avec les résultats des travaux utilisant le dataset SQuAD.

Dataset	SQUAD					
	BLEU				Rouge-L	METEOR
Modèle	Bleu1	Bleu2	Bleu3	Bleu4		
Arabe	41.14	11.60	10.59	10.03	15.08	11.61
Anglais	42.60	25.60	15.58	9.48	23.65	15.63
NQG [44]	41	23.78	15.71	10.8	37.95	15.17
Yuan et al [45]	/	/	/	10.5	/	/
JointQA [46]	/	/	/	10.20	/	/
FOCUS [47]	/	/	/	19.98	48.23	22.26
ASs2s [48]	/	/	/	16.20	43.96	19.92
NQG++ [49]	43,11	29,13	21,39	16,31	/	/
CGC-QG [50]	46.58	30.90	22.82	17.55	44.53	21.24
Zhao et al [51]	43,47	28,23	20,4	15,32	43,91	19,29
Scialom et al. [52]	43.33	26.27	18.32	13.23	40.22	/
Lopez etal [54]	/	/	/	8.26	44.38	21.2
Klein et Nabi[56]	31.46	19.50	12.41	7.84	34.51	/

Tableau 11: Comparaison des résultats de notre modèle avec les travaux connexes

Les résultats obtenus sont raisonnables par rapport à ceux des travaux connexes, malgré l'utilisation d'un dataset informel qui n'est pas bien élaboré grammaticalement et qui n'est pas contrôlé, qui contient pleins de fusionnement des mots et parfois la suppression des lettres depuis les mots.

Pour l'anglais, nous avons eu de meilleurs résultats par rapport aux travaux ayant utilisé un transformer avec dataset 'SQuAD' (8.26 et 7.84).

Pour l'arabe les résultats sont prometteurs et peuvent être améliorés par la suite.

3.2 Evaluation automatique -Beam-

L'objectif principal de notre travail, est de générer plusieurs exemples de questions pour la requête introduite par l'utilisateur. Afin de réaliser cet objectif, nous avons utilisé le décodage avec la recherche par faisceaux (définie dans chapitre III-section 6). Nous avons choisi les valeurs 1, 4, 7 et 10 pour générer les séquences de sorties. Les résultats obtenus avec cette méthode sont dans les tableaux 12 et 13, pour arabe et anglais respectivement.

	Bleu 1	**Bleu 2**	**Bleu 3**	**Bleu 4**	**Meteor**	**Rouge**
Beam 1	41.49	11.61	10.60	10.02	11.60	14.97
Beam 4	37.51	8.40	7.45	6.93	8.33	11.42
Beam 7	36.87	7.91	6.98	6.47	7.76	10.76
Beam 10	36.33	7.52	6.62	6.12	7.39	10.31

Tableau 12: Résultats de l'évaluation automatique avec différentes valeurs de beam -Arabe-

	Bleu 1	**Bleu 2**	**Bleu 3**	**Bleu 4**	**Meteor**	**Rouge**
Beam 1	42.79	25.71	15.64	9.52	15.62	23.61
Beam 4	41.81	24.83	14.68	8.85	13.29	19.73

Beam 7	40.62	24.06	14.65	8.68	12.44	18.54
Beam 10	39.90	23.33	13.33	8.40	12.02	17.47

Tableau 13: Résultats de l'évaluation automatique avec différentes valeurs de beam -Anglais-

Nous remarquons pour les deux langues arabe et anglais, à chaque fois le nombre de questions générés augmente donc plus de nœuds sont exploités, les valeurs de toutes les métriques diminuent. Cela est dû au calcul des métriques avec cette méthode inclus le meilleur et le pire cas, Exemple avec Bleu1 : Beam1 = 41.49% et Beam 10=36.33% pour arabe et Beam1 = 42.79% et Beam 10=39.90% pour anglais.

Les résultats de Beam 1 sont très proches avec les résultats de Softmax, quand une seule question est générée, car la plus haute probabilité est prédite.

3.3 Evaluation humaine

L'évaluation automatique ne fournit aucune mesure sur la pertinence et la correction grammaticale. Pour cela, un échantillon de 100 paires des questions réponses choisi aléatoirement depuis le jeu de test complet pour anglais et arabe a été utilisé afin d'évaluer leur pertinence, fluidité et grammaticalité.

Ces mesures ont été choisi, pour qu'on puisse les comparer avec les résultats existants pour la langue anglaise, mais pour l'arabe aucun résultat n'est disponible.

Modèle	Correction syntaxique/5	Pertinence/5
Arabe	3.5	2
Anglais	4.5	3.12
NQG [44]	3.36	/
FOCUS [47]	4.14	4.13
Scialom et al [52]	4.31	3.59

NQG++ [49]	/	2.18

Tableau 14: 2 Comparaison des résultats de l'évaluation humaine par rapport à la correction syntaxique et la pertinence.

Malgré l'utilisation d'un dataset non formel, qui n'est pas bien élaboré ni grammaticalement ni syntaxiquement, il a dépassé le dataset SQuAD par rapport à la correction syntaxique avec une pertinence plus aumois raisonnables par rapport aux travaux existants.

Par contre, pour la langue arabe nous remarquons un score très bas en pertinence qui peut être justifié par la nature de dataset communautaire informel tandis que les autres travaux ont utilisé SQuAD. Car il arrive que les phrases dans nos datasets ne suivre pas ni la forme grammaticale ni encore syntaxique. Ainsi, parfois les paires ne contiennent pas directement une question sur le produit en anglais ou sur le cas médical en arabe donc pleins d'informations inutiles. Ce qui rend le processus de l'entrainement difficile et va effectuer largement les résultats dans la phase de test.

4. Conclusion

Malgré la nature des datasets communautaires informels qui ne sont pas correctes grammaticalement, ainsi ils ne sont pas réguliers comme les datasets utilisés pour la tâche de génération de question comme le dataset SQuAD, nous avons eu des résultats acceptables par rapport aux travaux de la littérature.

Notons que nous ne pouvons pas comparer les résultats de la langue arabe avec d'autres travaux, car il existe aucun travail dans ce cadre qui présente une évaluation quantitative avec (Bleu, Rouge, Meteor …) ou encore une évaluation qualitative humaine.

Conclusion Générale

L'intérêt de notre travail a été la réalisation d'un modèle de génération de questions pour la langue arabe à l'aide d'un système de réponses aux questions communautaires. Dans le but d'enrichir les travaux existants pour cette langue.

Nous avons focalisé sur l'utilisation des méthodes et des techniques de l'apprentissage profond afin d'en profiter de ces avantages.

D'après la synthèse faite dans l'état de l'art sur les travaux connexes les plus récents, nous avons opté pour l'utilisation d'un modèle séquence à séquence bien précisément Encodeur/Décodeur avec un mécanisme d'attention afin d'améliorer la performance du modèle.

Nous avons conçu un modèle appliqué sur deux datasets distincts, un pour la langue arabe et l'autre pour la langue anglaise pour pouvoir comparer les résultats d'évaluation avec d'autres modèles.

Nous avons effectué deux méthodes d'évaluation. Évaluation automatique par l'utilisation des mêmes métriques utilisées dans les travaux de la littérature (BLEU, ROUGE-L et METEOR), et une évaluation humaine prenant en considération les critères les plus utilisés (grammaticalité, fluidité et pertinence).

Ce travail nous a permis de poser des « baselines » pour la langue arabe.

Pour encourager la réutilisation, le dataset arabe que nous avons adapté est rendu disponible ici :(https://github.com/Amirarahma/QuestionGeneration/tree/main/Dataset?fbclid=IwAR2fwAJIvGM75Axq419h9UsFFN6-vhulRtvrBxePwlKRSoABdtL5TLnHT3w)

Toutefois, Nous avons rencontré des problèmes liés directement à la nature du dataset utilisé et le manque de ressources en langue arabe :

- Le dataset est dédié à un seul domaine qui est le domaine médical, donc les questions générées ne sont pas variées.
- Le dataset utilisé est un dataset communautaire, donc il n'est pas régulier et il n'est pas grammaticalement bien élaboré.

Afin de surmonter ces limites, nous proposons en perspectives :

- ✓ L'utilisation d'un dataset de taille plus grande et qui est dédié à plusieurs domaines. Il s'agit alors soit de :

- Concaténer des datasets et faire les prétraitements nécessaires afin d'obtenir un ensemble de données riche et qui satisfait les besoins du modèle.
- Construire directement un dataset approprié à partir de sites web pour QA communautaires.

✓ L'implémentation de modèle en utilisant les transformers et ceci pour les avantages qu'ils proposent et explorer dans les modèles de langages pré-entrainés (T5, Bert, ...).

✓ L'utilisation du mécanisme de coverage et mécanisme de copie sans les embeddings pré-entrainés.

✓ Comme les résultats de pertinence ont été basses puisque y'avait pas beaucoup de mots qui relient la phrase par sa questions (pas de chevauchement) alors que certaines questions générées sont correctes mais ne sont pas identiques à la vraie question. Nous avons pensé a changé le dataset pour que la phrase peut avoir plusieurs questions. Et donc améliorer les résultats de l'évaluation.

✓ Considération des entités nommées (noms propres) pour améliorer la qualité des questions générées.

Références Bibliographique

[1] « Qu'est-ce que l'intelligence artificielle ? » NetApp. https://www.netapp.com/fr/artificial-intelligence/what-is-artificial-intelligence/ (consulté le 24 août 2022).

[2] « Qu'est-ce que le machine learning ? » Oracle Cloud Infrastructure. https://www.oracle.com/dz/artificial-intelligence/machine-learning/what-is-machine-learning/ (consulté le 13 août 2022).

[3] L. Deng, « Deep learning : Methods and applications », Foundations and Trends® in Signal Processing, vol. 7, no 3-4, p. 197–387, 2014. Consulté le 25 août 2022. [En ligne]. Disponible : https://doi.org/10.1561/2000000039.

[4] D. Belhaouci. « Démystifier le Machine Learning, Partie 2 : les Réseaux de Neurones artificiels ». JuriPredis.
https://www.juripredis.com/fr/blog/id-19-demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels (consulté le 13 août 2022).

[5] M. Mouadil. « Introduction au Deep Learning : Les réseaux de neurones ». Meritis. https://meritis.fr/deep-learning/ (consulté le 7 sept. 2022).

[6] M. Nayak. « Introduction to the Architecture of Recurrent Neural Networks (RNNs) ». towards ai.
https://pub.towardsai.net/introduction-to-the-architecture-of-recurrent-neural-networks-rnns-a277 007984b7 (consulté le 25 août 2022).

[7] A. Kapathy. « The Unreasonable Effectiveness of Recurrent Neural Networks ». Andrej Kapathy blog. http://karpathy.github.io/2015/05/21/rnn-effectiveness/ (consulté le 25 août 2022).

[8] B. Miroslav et R. Viera, Machine Learning Approach to the Process of Question Generation.

[9] G. Singhal. « Introduction to LSTM Units in RNN ». pluralsight. https://www.pluralsight.com/guides/introduction-to-lstm-units-in-rnn (consulté le 25 août 2022).

[10] « Fonctions d’activation dans les réseaux de neurones ». Acervo Lima. https://fr.acervolima.com/fonctions-d-activation-dans-les-reseaux-de-neurones/ (consulté le 13 août 2022).

[11] P. Baheti. « Activation functions in neural networks [12 types & ; use cases] ». V7 - AI Data Platform for Computer Vision. https://www.v7labs.com/blog/neural-networks-activation-functions (consulté le 7 sept. 2022).

[12] S. Kostadinov. « Understanding Encoder-Decoder Sequence to Sequence Model ». Medium.https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-67 9e04af4346 (consulté le 26 août 2022).

[13] « BERT : Le "Transformer model" qui s’entraîne et qui représente ». Les Dieux Du Code. https://lesdieuxducode.com/blog/2019/4/bert--le-transformer-model-qui-sentraine-et-qui-represen te (consulté le 26 août 2022).

[14] A. Vaswani et al., « Attention Is All You Need », 31st Conference on Neural Information Processing Systems, p. 3, déc. 2017. Consulté le 26 août 2022. [En ligne]. Disponible : https://arxiv.org/pdf/1706.03762.pdf.

[15] T. Keldenich. « Le Mécanisme de l'Attention en Deep Learning - Comprendre rapidement
». Inside Machine Learning. https://inside-machinelearning.com/mecanisme-attention/ (consulté le 25 août 2022).

[16] Synced. « A Brief Overview of Attention Mechanism ». Medium. https://medium.com/syncedreview/a-brief-overview-of-attention-mechanism-13c578ba9129 (consulté le 25 août 2022).

[17] « The differences between classical arabic and modern standard arabic | live lingua ». Live Lingua.https://www.livelingua.com/arabic/the-differences-between-classical-arabic-and-modern-standard-arabic (consulté le 29 août 2022).

[18] B. Hammo, H. Abu-Salem et S. Lytinen, « QARAB », dans the ACL-02 workshop, Philadelphia, Pennsylvania, 11 juill. 2002. Morristown, NJ, USA : Association for Computational Linguistics, 2002. Consulté le 26 août 2022. [En ligne]. Disponible : https://doi.org/10.3115/1118637.1118644.

[19] J. H. Wolfe, « Automatic question generation from text - an aid to independent study », ACM SIGCUE Outlook, vol. 10, SI, p. 104–112, févr. 1976. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.1145/953026.803459.

[20] K. Papineni, S. Roukos, T. Ward et W.-J. Zhu, « BLEU », dans the 40th Annual Meeting, Philadelphia, Pennsylvania, 7–12 juill. 2002. Morristown, NJ, USA : Association for Computational Linguistics, 2001. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.3115/1073083.1073135.

[21] J. Briggs. « The Ultimate Performance Metric in NLP ». Medium. https://towardsdatascience.com/the-ultimate-performance-metric-in-nlp-111df6c64460 (consulté le 28 août 2022).

[22] A. Abhaya et L. Alon, « Proceedings of the Third Workshop on Statistical Machine Translation », Columbus, Ohio, USA. Columbus, Ohio : Association for Computational Linguistics, 2008, p. 115–118. Consulté le 28 août 2022. [En ligne]. Disponible : https://aclanthology.org/W08-0312.pdf.

[23] G. Kurdî, J. Leo, B. Parsia, U. Sattler et S. Al-Emari, « A systematic review of automatic question generation for educational purposes », International Journal of Artificial Intelligence in Education, vol. 30, no 1, p. 121–204, nov. 2019. Consulté le 28 août 2022. [En ligne]. Disponible
: https://doi.org/10.1007/s40593-019-00186-y.

[24] D. R. CH et S. K. Saha, « Automatic Multiple Choice Question Generation From Text : A Survey », IEEE Transactions on Learning Technologies, vol. 13, no 1, p. 14–25, janv. 2020. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.1109/tlt.2018.2889100.

[25] J. Amidei, P. Piwek et A. Willis, « Evaluation methodologies in Automatic Question Generation 2013-2018 », dans Proceedings of the 11th International Conference on Natural Language Generation, Tilburg University, The Netherlands. Stroudsburg, PA, USA : Association for Computational Linguistics, 2018. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/w18-6537.

[26] M. Divate et A. Salgaonkar, « Automatic Question Generation Approaches and Evaluation Techniques », Current Science, vol. 113, no 09, p. 1683, nov. 2017. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18520/cs/v113/i09/1683-1691.

[27] P. Rajpurkar, J. Zhang, K. Lopyrev et P. Liang, « SQuAD : 100,000+ Questions for Machine Comprehension of Text », dans Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas. Stroudsburg, PA, USA : Association for Computational Linguistics, 2016. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/d16-1264.

[28] P. Rajpurkar, R. Jia et P. Liang, « Know What You Don't Know : Unanswerable Questions for SQuAD », dans Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers), Melbourne, Australia. Stroudsburg, PA, USA : Association for Computational Linguistics, 2018. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/p18-2124.

[29] M. Joshi, E. Choi, D. Weld et L. Zettlemoyer, « TriviaQA : A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension », dans Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers), Vancouver, Canada. Stroudsburg, PA, USA : Association for Computational Linguistics, 2017. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/p17-1147.

[30] Y. Yang, W.-t. Yih et C. Meek, « WikiQA : A Challenge Dataset for Open-Domain Question Answering », dans Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. Stroudsburg, PA, USA : Association for Computational Linguistics, 2015. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/d15-1237.

[31] P. Bajaj, D. Campos, N. Craswell, L. Deng, J. Gao, X. Liu, R. Majumder, et al. « MS MARCO: A Human Generated MAchine Reading Comprehension Dataset ». arXiv, 31 octobre 2018. http://arxiv.org/abs/1611.09268.

[32] « Amazon review data ». http://jmcauley.ucsd.edu/data/amazon/index.html (consulté le 28 août 2022).

[33] G. Lai, Q. Xie, H. Liu, Y. Yang et E. Hovy, « RACE : Large-scale ReAding Comprehension Dataset From Examinations », dans Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark. Stroudsburg, PA, USA : Association for Computational Linguistics, 2017. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/d17-1082

[34] G. Chen, J. Yang, C. Hauff, and G.-J. Houben, « LearningQ: A Large-Scale Dataset for Educational Question Generation », ICWSM, vol. 12, no 01, Juin. 2018.

[35] A. Abujabal, R. Saha Roy, M. Yahya et G. Weikum, « ComQA: A Community-sourced Dataset for Complex Factoid Question Answering with Paraphrase Clusters », dans Proceedings of the 2019 Conference of the North, Minneapolis, Minnesota. Stroudsburg, PA, USA : Association for Computational Linguistics, 2019. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/n19-1027.

[36] M. Artetxe, S. Ruder et D. Yogatama, « On the cross-lingual transferability of monolingual representations », dans Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online. Stroudsburg, PA, USA : Association for Computational Linguistics, 2020. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/2020.acl-main.421.

[37] P. Lewis, B. Oguz, R. Rinott, S. Riedel et H. Schwenk, « MLQA : Evaluating cross-lingual extractive question answering », dans Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online. Stroudsburg, PA, USA : Association for Computational Linguistics, 2020. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/2020.acl-main.653.

[38] P. Nakov et al., « SemEval-2016 Task 3 : Community Question Answering », dans Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, California. Stroudsburg, PA, USA : Association for Computational Linguistics, 2016. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/s16-1083.

[39] O. Essam. « GitHub - Omarito2412/ASKFM : ASKFM : Arabic AskFM dataset, a knowledge base of Question/answers in Arabic ». GitHub. https://github.com/Omarito2412/ASKFM (consulté le 28 août 2022).

[40] A. Atef, B. Mattar, S. Sherif, E. Elrefai et M. Torki, « AQAD : 17,000+ Arabic Questions for Machine Comprehension of Text », dans 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), Antalya, Turkey, 2–5 nov. 2020. IEEE, 2020. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.1109/aiccsa50499.2020.9316526.

[41] W. S. Ismail et M. N. Homsi, « DAWQAS : A Dataset for Arabic Why Question Answering System », Procedia Computer Science, vol. 142, p. 123–131, 2018. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.1016/j.procs.2018.10.467.

[42] L. Ouahrani et D. Bennouar. « AR-ASAG An ARabic Dataset for Automatic Short Answer Grading Evaluation », dans Proceedings of the 12th Language Resources and Evaluation Conference, 2634-43. Marseille, France: European Language Resources Association, 2020. https://aclanthology.org/2020.lrec-1.321.

[43] H. Mozannar, E. Maamary, K. El Hajal et H. Hajj, « Neural Arabic Question Answering », dans Proceedings of the Fourth Arabic Natural Language Processing Workshop, Florence, Italy. Stroudsburg, PA, USA : Association for Computational Linguistics, 2019. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/w19-4612.

[44] X. Du, J. Shao et C. Cardie, « Learning to ask : Neural question generation for reading comprehension », dans Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers), Vancouver, Canada. Stroudsburg, PA, USA

: Association for Computational Linguistics, 2017. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/p17-1123.

[45] X. Yuan et al., « Machine comprehension by text-to-text neural question generation », dans Proceedings of the 2nd Workshop on Representation Learning for NLP, Vancouver, Canada. Stroudsburg, PA, USA : Association for Computational Linguistics, 2017. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/w17-2603.

[46] T. Wang, X. Yuan et A. Trischler, « A Joint Model for Question Answering and Question Generation », dans Learning to generate natural language workshop, ICML 2017. Consulté le 28 août 2022. [En ligne]. Disponible : https://arxiv.org/pdf/1706.01450.pdf.

[47] V. Harrison et M. Walker, « Neural generation of diverse questions using answer focus, contextual and linguistic features », dans Proceedings of the 11th International Conference on Natural Language Generation, Tilburg University, The Netherlands. Stroudsburg, PA, USA : Association for Computational Linguistics, 2018. Consulté le 28 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/w18-6536.

[48] Y. Kim, H. Lee, J. Shin et K. Jung, « Improving neural question generation using answer separation », Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, p. 6602–

6609, juill. 2019. Consulté le 29 août 2022. [En ligne]. Disponible : https://doi.org/10.1609/aaai.v33i01.33016602.

[59] W. Zhou, M. Zhang et Y. Wu, « Question-type driven question generation », dans Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China. Stroudsburg, PA, USA : Association for Computational Linguistics, 2019. Consulté le 11 sept. 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/d19-1622

[50] B. Liu et al., « Learning to generate questions by learning what not to generate », dans The World Wide Web Conference, San Francisco, CA, USA, 13–17 mai 2019. New York, New York, USA : ACM Press, 2019. Consulté le 29 août 2022. [En ligne]. Disponible : https://doi.org/10.1145/3308558.3313737.

[51] Y. Zhao, X. Ni, Y. Ding et Q. Ke, « Paragraph-level neural question generation with maxout pointer and gated self-attention networks », dans Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium. Stroudsburg, PA, USA : Association for Computational Linguistics, 2018. Consulté le 11 sept. 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/d18-1424.

[52] T. Scialom, B. Piwowarski et J. Staiano, « Self-Attention architectures for answer-agnostic neural question generation », dans Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy. Stroudsburg, PA, USA : Association for Computational Linguistics, 2019. Consulté le 29 août 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/p19-1604.

[53] K. Z. Bousmaha, N. H. Chergui, M. S. A. Mbarek et L. B. Hadrich, « AQG : Arabic question generator », Revue d'Intelligence Artificielle, vol. 34, no 6, p. 721–729, déc. 2020. Consulté le 29 août 2022. [En ligne]. Disponible : https://doi.org/10.18280/ria.340606.

[54] L. E. Lopez, D. K. Cruz, J. C. B. Cruz, et C. Cheng. « Simplifying Paragraph-level Question Generation via Transformer Language Models ». arXiv, 13 août 2021. http://arxiv.org/abs/2005.01107.

[55] K. Kettip, et A. Wangperawong, « Question Generation by Transformers ». arXiv, 14 septembre 2019. http://arxiv.org/abs/1909.05017.

[56] T. Klein, et M. Nabi, « Learning to Answer by Learning to Ask: Getting the Best of GPT-2 and BERT Worlds ». arXiv, 6 novembre 2019. [En ligne]. Disponible : http://arxiv.org/abs/1911.02365.

[57] S. K. Dwivedi, V. Singh, « Research and Reviews in Question Answering System ». Procedia Technology 10 (2013): p. 417-424. [En ligne]. Disponible : https://doi.org/10.1016/j.protcy.2013.12.378.

[58] I. Srba et M. Bielikova, « A comprehensive survey and classification of approaches for community question answering », ACM Transactions on the Web, vol. 10, no 3, p. 1–63, août 2016. Consulté le 29 août 2022. [En ligne]. Disponible : https://doi.org/10.1145/2934687.

[59] B. John, et J. Kurian. « Research Issues In Community Based Question Answering. » dans PACIS 2011 - 15th Pacific Asia Conference on Information Systems: Quality Research in Pacific, 29, 2011.

[60] « Amazon question/answer data ». https://jmcauley.ucsd.edu/data/amazon/qa/ (consulté le 5 sept. 2022).

[61] « Task 3 : Community question answering < ; semeval-2016 task 3 ». ALT Website – Arabic Language Technologies Group. https://alt.qcri.org/semeval2016/task3/ (consulté le 5 sept. 2022).

[62] P. Nakov, L. Màrquez, A. Moschitti et H. Mubarak, « Arabic community question answering », Natural Language Engineering, vol. 25, no 1, p. 5–41, déc. 2018. Consulté le 9 sept. 2022. [En ligne]. Disponible : https://doi.org/10.1017/s1351324918000426.

[63] M. Wan et J. McAuley, « Modeling ambiguity, subjectivity, and diverging viewpoints in opinion question answering systems », dans 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 déc. 2016. IEEE, 2016. Consulté le 5 sept. 2022. [En ligne]. Disponible : https://doi.org/10.1109/icdm.2016.0060.

[64] K. E. Koech. « Cross-Entropy loss function ». Medium.https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e (consulté le 5 sept. 2022).

[65] Musstafa. « Optimizers in deep learning ». Medium. https://medium.com/mlearning-ai/optimizers-in-deep-learning-7bf81fed78a0 (consulté le 5 sept. 2022).

[66] J. Brownlee. « Understand the impact of learning rate on neural network performance ». Machine Learning Mastery.https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neur al-networks/ (consulté le 5 sept. 2022).

[67] J. Brownlee. « Difference between a batch and an epoch in a neural network ». Machine Learning Mastery.https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/#:~:text=The%20batch%20size%20is%20a,passes%20through%20the%20training%20dataset (consulté le 5 sept. 2022).

[68] T. Mikolov, K. Chen, G. Corrado, et J. Dean. « Efficient Estimation of Word Representations in Vector Space ». arXiv, 6 septembre 2013. [En ligne]. Disponible : http://arxiv.org/abs/1301.3781.

[69] A. See, P. J. Liu et C. D. Manning, « Get to the point : Summarization with pointer-generator networks », dans *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, Vancouver, Canada. Stroudsburg, PA, USA : Association for Computational Linguistics, 2017. Consulté le 13 sept. 2022. [En ligne]. Disponible : https://doi.org/10.18653/v1/p17-1099

[70] S. Yang, X. Yu et Y. Zhou, « LSTM and GRU neural network performance comparison study : Taking yelp review dataset as an example », dans 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI), Shanghai, China, 12–14 juin 2020. IEEE, 2020. Consulté le 5 sept. 2022. [En ligne]. Disponible : https://doi.org/10.1109/iwecai50956.2020.00027.

[71] M. Payne. « What is beam search ? Explaining the beam search algorithm | width.ai ». Artificial Intelligence and Machine Learning Consulting Services | Width.ai. https://www.width.ai/post/what-is-beam-search (consulté le 5 sept. 2022).

[72] K. Doshi. « Foundations of NLP explained visually : Beam search, how it works ». Medium. https://towardsdatascience.com/foundations-of-nlp-explained-visually-beam-search-how-it-works-158 6b9849a24 (consulté le 5 sept. 2022).

[73] D. Bahdanau, C. Kyunghyun, et B. Yoshua, « Neural Machine Translation by Jointly Learning to Align and Translate », 2014. [En ligne]. Disponible : https://doi.org/10.48550/ARXIV.1409.0473

Annexes

Nous mettons ici des exemples (échantillons) des datasets mentionnées dans les tableaux (Tableau 1 et 2 dans chapitre II).

- **SQuAD**

• Id : 5733be284776f41900661182

• Title : University_of_Notre_Dame

• Context : Architecturally, the school has a Catholic character. Atop the Main Building's gold dome is a golden statue of the Virgin Mary. Immediately in front of the Main Building and facing it, is a copper statue of Christ with arms upraised with the legend "Venite Ad Me Omnes". Next to the Main Building is the Basilica of the Sacred Heart. Immediately behind the basilica is the Grotto, a Marian place of prayer and reflection. It is a replica of the grotto at Lourdes, France where the Virgin Mary reputedly appeared to Saint Bernadette Soubirous in 1858. At the end of the main drive (and in a direct line that connects through 3 statues and the Gold Dome), is a simple, modern stone statue of Mary.

• Question : To whom did the Virgin Mary allegedly appear in 1858 in Lourdes, France?

• Answers : { "text": ["Saint Bernadette Soubirous"], "answer start": [515] }

- **TriviaQA**

• Question: The Dodecanese Campaign of WWII that was an attempt by the Allied forces to capture islands in the Aegean Sea was the inspiration for which acclaimed 1961 commando film?.

• Answer: The Guns of Navarone.

• Excerpt: The Dodecanese Campaign of World War II was an attempt by Allied forces to capture the Italianheld Dodecanese islands in the Aegean Sea following the surrender of Italy in September 1943, and use them as bases against the German-controlled Balkans. The failed campaign, and in particular the Battle of Leros, inspired the 1957 novel The Guns of Navarone and the successful 1961 movie of the same name.

- **WikiQA**

• Question_id: Q0

• Question: HOW AFRICAN AMERICANS WERE IMMIGRATED TO THE US ?

• Document_title: African immigration to the United States.

• Answer: African immigration to the United States refers to immigrants to the United States who are or were nationals of Africa .

• Label_class: 0 (0).

- **MS Macro**

• Answers: ["Approximately $15,000 per year."]

• Passages: { "is_selected": [1, 0, 0, 0, 0, 0], "passage_text": ["The average Walgreens salary ranges from approximately $15,000 per year for Customer Service Associate / Cashier to $179,900 per year for District Manager. Average Walgreens hourly pay ranges from approximately $7.35 per hour for Laboratory Technician to $68.90 per hour for Pharmacy Manager. Salary information comes from 7,810 data points collected directly from employees, users, and jobs on Indeed.", "The average revenue in 2011 of a Starbuck Store was $1,078,000, up from $1,011,000 in 2010. The average ticket (total purchase) at domestic Starbuck stores in No … vember 2007 was reported at $6.36. In 2008, the average ticket was flat (0.0% change).",], }

• Query: walgreens store sales average.

• Query_id: 9,652.

• Query_type: numeric.

- **Amazon**

• "asin": "B000050B6Z",

• "questionType": "yes/no",

• "answerType": "Y",

• "answerTime": "Aug 8, 2014",

• "unixTime": 1407481200,

• "question": "Can you use this unit with GEL shaving cans?",

• "answer": "Yes. If the can fits in the machine it will dispense hot gel lather. I've been using my machine for both gel and traditional lather for over 10 years."

- **Race**

• Article: There is probably no field of human activity in which our values and lifestyles are shown more clearly and strongly than they are in the clothes that we choose to wear.

• Answer: D

• Question: Blue collar workers pay attention to their clothes because _ .

•Option: "they are concerned about the impression their clothes make on their superiors", "they know very clearly that people will judge them on the basis of their clothing", "they want to impress and influence others", "they don't want to be laughed at".

- **LearningQ**

Title: Do animals have language? - Michele Bishop

All animals communicate. Crabs wave their claws at each other to signal that they're healthy and ready to mate. Cuttlefish use pigmented skin cells called chromatophores to create patterns on their skin that act as camouflage or warnings to rivals. Honeybees perform complex dances to let other bees know the location and quality of a food source. All of these animals have impressive communication systems, but do they have language? To answer that question, we can look at four specific qualities that are often associated with language: discreteness, grammar, productivity, and displacement.

- **Com-QA**

• "cluster_id": "cluster-1754"

• “questions:” ["what years did cale yarborough win his cup championships?", "what years did cale yarborough win winston cup champs?"].

• “answers” : ["1976", "1978", "1977"].

● XQuAD

• id : 56beb4343aeaaa14008c925b

• context : لم يتخلى فريق بانثرز سوى عن 308 نقطة، ليحتل المركز السادس في الدوري، متصدراً في الوقت عينه دوري كرة القدم الأميركية في الاعتراضات مع 24 اعتراضاً ومتمتعاً ب أربعة اختيارات في قائمة برو بول. وقد تصدر المدافع المعترض كاوان شورت الفريق في الاستحواذ مع 11 استحواذاً، بينما أجبر الخصم أيضاً على ارتكاب ثلاثة أخطاء واسترجع اثنين. وقد أضاف زميله لاعب الخط ماريو أديسون 6 استحواذات ونصف. وقد ضم أيضاً المدافع الأخير المخضرم جاريد ألان، وهو لاعب كرة قدم اختير 5 مرات في قائمة برو بول Panthers خط وكان متصدر اللاعبين النشطين لدوري كرة القدم الأميركي في الاستحواذ مع 136 استحواذاً، إلى جانب المدافع الأخير كوني إيلي، الذي حصل على 5 استحواذات في 9 بدايات فقط. وخلفهم، تم أيضاً اختيار اثنين من أظهرة فريق Panthers ، الثلاثة الأساسيين للعب في برو بول: توماس ديفيس و لوك كوتشلي. قام ديفيز بتحصيل 5 استحواذ وأربعة مرات أجبر الخصم على ارتكاب أخطاء، وأربعة اعتراضات، في حين قاد كوتشي الفريق في العرقلات (118) وإجبار الخصم على ارتكاب خطأين، واعترضأربع تمريرات بنفسه. وضم ثانوي كارولاينا مؤمّن برو بول كورت كولمان، الذي تصدر الفريق بأعلى رقم في مسيرته بسبعة اعتراضات، في حين جمع أيضاً ما يصل إلى 88

عرقلة ومدافع الزاوية في قائمة برو بول جوش نورمان، الذي تحول إلى ركن إقفال خلال الموسم وقام بأربع اعتراضات، تم إرجاع اثنين منها للهدف.

• question : كم نقطة تخلى عنها دفاع البانثرز ؟

• answers : { "text": ["308"], "answer_start": [29] }.

- **MLQA**

'data': [{'title': ' 51 المنطقة ', 'paragraphs': [{'context': ' لأجزاء "xx" يستخدم نفس نظام تسمية المنطقة أخرى من مواقع التجارب في نيفادا.وتعتبر القاعدة المستطيلة الأصلية من 6 إلى 10 ميل هي الآن جزءا من ما يسمى ب "جروووم", وهى القاعدة المستطيلة من 23 إلى 25.3 ميل من المجال الجوي المحظور. وتتصل المنطقة بشبكة ، والطرق الممهدة المؤدية جنوبا إلى ميركري وغربا إلى مسطح يوكا. مما يؤدي شمال شرقNTS الطرق الداخلية البحيرة إلى طريق بحيرة جروووم الواسع الجيد الصيانة الذي يمر عبر تلال جمبلد. في السابق كان يؤدي إلى المناجم في حوض جروم، ولكنه تحسن منذ إغلاقها. الطريق المتعرج يمر عبر نقاط التفتيش الأمنية، ولكن المنطقة المحظورة حول قاعدة تمتد أكثر إلى الشرق.بعد خروجه من المنطقة المحظورة، طريق بحيرة جروووم ينحدر شرقا ، ويمر على مداخل الطرق الترابية لعدة مزارع مواشي، قبل أن يلتقيا مع الطريق الرئيسيTikaboo إلى سهل وادي أي نوع من الطرق ' :'question'}] ,'qas' .', ، على "الطريق السريع خارج الأرض"، إلى الجنوب من راشيل375 يؤدي إلى مزارع المواشي؟ ', 'answers': [{'text': ' الطرق الترابية ', 'answer start': 680}].

- **CQA-MD**

<QApair QAID="43211">

<QAquestion> اعاني من الم في البطن والظهر مع افرازات آخر دورة في 2014 10 21 في موعد الدورة التاني 2014 11 21 نزل نقطة دم وبعدها لم ينزل شي والالم لا يتوقف وقمت بفحص الشريط المنزلي مرتين والنتيجة سلبية والم راسي لا يتوقفة واحيانا يصيبني دوار </QAquestion>

<QAanswer> لا يجب الاعتماد على الفحص المنزلي انصح بمراجعة طبيب و عمل فحص حساس للحمل </QAanswer>

</QApair>

- **AskFm**

• question: شيخ انا اسفه ف السؤال ده بس في ثوابت ف الدين بتضيع مني هو بجد كان في أية ف القرآن وفي معزه اكلتها وحكمها لسه موجود لحد دلوقتي ف حد الزنا !!!!!!

• answer: هو حديث ضعيف، وعموما لو معزة كلت المصحف كله فالقرآن ثابت محفوظ. لا تعطي أذنك للجهلة الحمقى فتخسري دينك على حاجات تافهة

- **AQAD**

"title": " ,"متصفح ويب

"paragraphs": [{"qas":[{ "question": "ماذا يسمي الناس عادة متصفح الويب ؟ ", "id": 1, "answers": [{ "text": " المتصفح ", "answer_start": 0}], "is_impossible": false}, {"question": " ما

الإنترنت", " النظام الأساسي الذي يستخدمه المتصفح؟ ", "id":2, "answers": [{"text":
"answer_start": 96}], "is_impossible": false}.

- **DAWQAS**

• QID: 1

• Question: لماذا الرجال أكثر وزناً بعد الزواج ؟

• Answer 1 : "كشفت دراسة علمية عن أنّ الرجال المتزوّجين يكتسبون الوزن، مقارنة بأقرانهم، غير المتربطين نيويورك « عاطفياً، وقالت: إنّ مؤشر كتلة أجساد الرجال يزداد، بعد الزواج، وعقب إنجاب أول طفل، وفقاً لصحيفة إلى أن الرجال العازبين يحافظون على لياقتهم ونحافتهم، وذلك » باث « الأميركية. وأشار الباحثون في جامعة » بوست سعياً منهم من أجل اجتذاب الفتيات نحوهم، ولكنّهم يبدأون في تناول الأطعمة الأقل صحية، والأكثر دسامة بعد الزواج. وتفسِّر رئيسة الدراسة، جوانا سيردا، أن من يخوضون علاقات، يتناولون الطعام بدرجة أكبر، بسبب الالتزامات الاجتماعية التي قد تنشأ بعد الزواج، موضحة أن الرجال الذين ما زالوا في مرحلة البحث عن شريكة للحياة، يبذلون مجهوداً مضاعفاً من أجل الحفاظ على لياقتهم، عكس الأفراد المتزوجين، وهو ما يؤدي إلى ارتفاع مؤشر كتلة الجسم عند المتزوجين. وقامت سيردا في دراستها، بتحليل بيانات الطول والوزن على 8.729 رجلاً، والتي جمعتها من عام 1999 وحتى 2013 ، فاكتشفت أن مؤشر كتلة جسم الرجل ينخفضقبل وبعد الطلاق مباشرة، بينما لم تتأثر بحمل زوجته. وأضافت أن هناك نحو 7 من كل 10 رجال متزوجين في بريطانيا) 65 %(، يعانون من زيادة في الوزن أو السمنة. " .

- **AR-ASAG**

• Question: عرف مصطلح الجريمة الإلكترونية

• Answer: هي كل سلوك غير قانوني يتم باستخدام الأجهزة الإلكترونية(الهاتف، الكمبيوتر، الانترنت)ينتج عنه حصول المجرم على فوائد مادية أو معنوية مع تحميل الضحية خسارة وغالبا ما يكون هدف هذه الجرائم هو القرصنة هي سلوك غير أخلاقي يتم [من أجل سرقة أو إتلاف المعلومات وتكون عادة الانترنت أداة لها أو مسرحا لها 3.0]1
هي كل سلوك غير أخلاقي [عن طريق وسائل الكترونية يهدف الى عائدات مادية و يسبب اضرارا للضحية 5.0]2
يتم بواسطة الاجهزة الالكترونية ينتج عنها حصول المجرم على فوائد مادية أو معنوية مع تحصيل الضحية خسارة مقابلة، هدفها القرصنة من اجل سرقة او اتلاف المعلومات

- **ARCD**

'data': [{'title': ' جمال خاشقجي ', 'paragraphs': [{'context': ' جمال أحمد حمزة خاشقجي) 13 أكتوبر
، المدينة المنورة - 2 أكتوبر 2018 (، صحفي وإعلامي سعودي، رأس عدّة مناصب لعدد من الصحف في1958
من هو - ' :'question'} ,'qas': [{ .' السعودية، وتقلّد منصب مستشار، كما أنّه مدير عام قناة العرب الإخبارية سابقًا
صحفي وإعلامي', ' :'text'} ,'answers': [{ ' جمال أحمد حمزة خاشقجي؟ ', 'id': '969331847966'
'answer_start': 73}]}, {'question': ' - ال متى ولد جمال أحمد حمزة خاشقجي وتوفي؟ ', 'id':
حمزة خاشقجي) 13 أكتوبر 1958 ، المدينة المنورة - 2 أكتوبر ' :'text'} ,'answers': [{ '115150665555'
2018)،', 'answer_start': 10}]}, {'question': ' - ال في أي مدينة ولد جمال أحمد حمزة خاشقجي؟ ', 'id':

'74212080718', 'answers': [{'text': ' المدينة المنورة ', 'answer_start': 39}]}].

- **Arabic SQuAD**

'data': [{'title': 'ASCII', 'paragraphs': [{'context': ' يعتمد ASCII على أساس ا على الأبجدية الإنجليزية ، ويقوم بترميز 128 حرف ا محدد ا في أعداد صحيحة من سبعة أجزاء كما هو موضح في مخطط ASCII . الأحرف المشفرة هي الأرقام من 0 إلى 9 ، والأحرف الصغيرة ا إلى ز ، والأحرف الكبيرة A إلى Z ، وهي الأحرف الأبجدية الإنجليزية ا على اليمين ورموز الترقيم الأساسية ، ورموز التحكم التي نشأت مع أجهزة تيليتيبي ، ومساحة . على سبيل المثال ، سيصبح الحرف الصغير ج 1101010 والعشري 106 . تتضمن ASCII تعريفات ل 128 حرف ا 33 حرف ا تحكم ا غير الطباعة العديد منها الآن قديمة تؤثر على كيفية معالجة النص والمساحة و 95 حرف ا قابلا للطباعة ، بما في ذلك المساحة التي ي عتبر رسم ا غير مرئي 223 .', 'qas': [{'question': ' ما هو ASCII على أساس ؟ ', 'id': '570bce516b8089140040fa42', 'answers': [{'text': ' الأبجدية الإنجليزية ', 'answer_start': 23}]}, {'question': ' كم شخصيات محددة موجودة في كود ASCII؟', 'id': '570bce516b8089140040fa43', 'answers': [{'text': ' 128 حرف ا محدد ', 'answer_start': 58}]}, {'question': ' كم عدد أحرف التحكم غير حرف ا تحكم ا غير الطباعة 33' ,' الطباعة؟ ', 'id': '570bce516b8089140040fa44', 'answers': [{'text': ' 'answer_start': 405}]}, {'question': 'كم شخصيات قابلة للطباعة ؟ ', 'id': '570bce516b8089140040fa45', 'answers': [{'text': ' 95 حرف ا قابلا للطباعة ', 'answer_start': 494}]}, {'question': 'ما هو الفضاء المعروف أيضا باسم ماذا' ؟ ', 'id': '570bce516b8089140040fa46', 'answers': [{'text': ' 223 رسم ا غير مرئي ', 'answer_start': 550