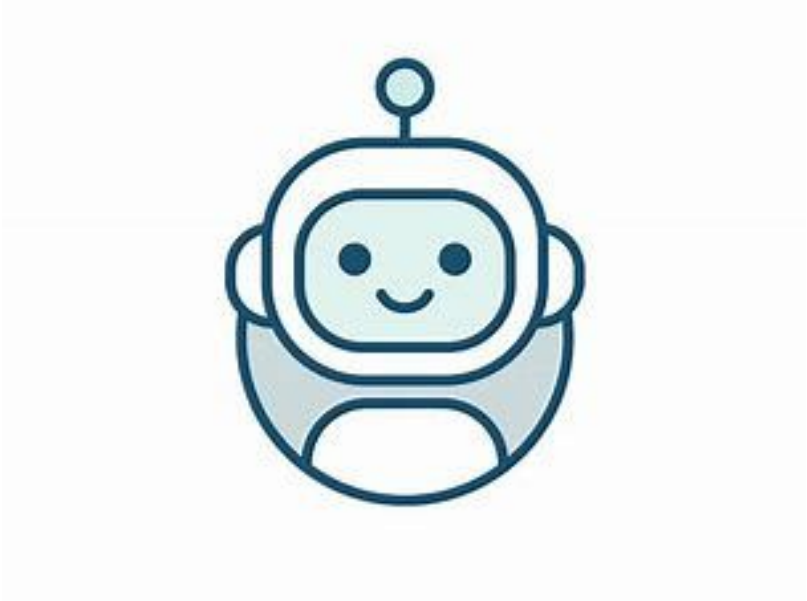
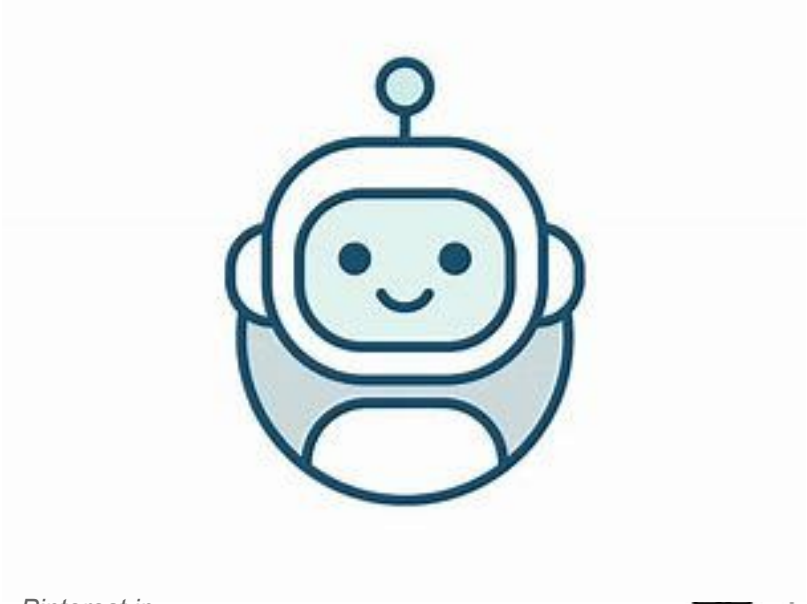
RVP1



Pinterest.jp

PAR - 102 : Chatbot

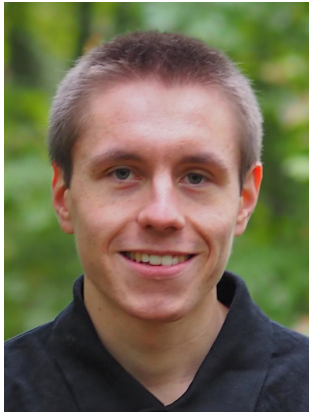
IA © Freepik



Pinterest.jp

Présentation générale: L’équipe

- Qui sommes nous?
  - Matthias Personnaz
  - Clément Yvernes



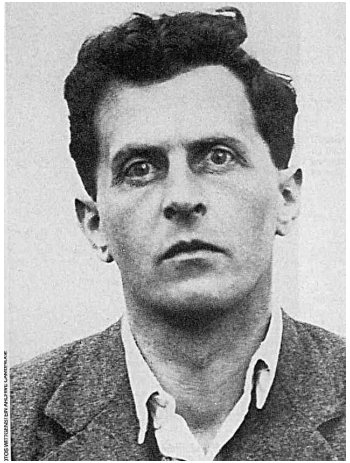
Présentation générale: Le projet

chatbot : « un programme informatique conçu pour tenir une conversation avec un être humain »

Cambridge dictionary

“ La signification d'un mot est son usage dans le langage.”

Ludwig Wittgenstein
Recherches philosophiques



Présentation générale: L’équipe pédagogique

- Tuteurs:
  - Alexandre Saidi
  - Philippe Michel

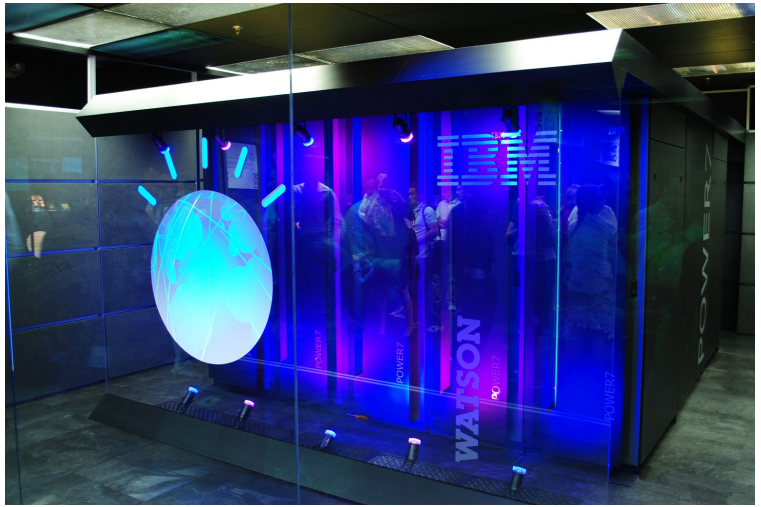
- Conseiller en communication:
  - Denis Mazuyer

Présentation générale: Contexte & motivations

Une (très) brève histoire du NLP…

NLP: Natural Language Processing (Traitement du Langage Naturel)

- 1950: conceptualisation du test de Turing
- 1957: travaux de Noam Chmosky sur la classification des grammaires formelles
- depuis les années 70-80: rationalisme (règles formelles)
- depuis 2010: progrès spectaculaires en NLP dû à l'augmentation de puissance
- ex.: Word2Vec, IBM Watson, GPT-3



Ordinateur IBM Watson (source: Wikimedia Commons)

Présentation générale: Contexte & motivations

- Motivation générale: besoin d'information pour les élèves de l'ECL
- Diversification des bassins de recrutement ces dernières années
- Grand nombre de double-diplômes entrants et sortants, échanges, etc.
- Nombreux documents et ressources

  → **Intérêt (besoin ?) de simplifier leur consultation**

Présentation générale: Enjeux

- Enjeux classiques de la recherche
- Création de connaissance et d’objet technique
- Réplications de résultats issus de la recherche
- Meilleure compréhension du NLP en contexte spécifique (milieu universitaire)

Objectifs et évaluations

Objectifs et évaluations

Objectif: Concevoir un chatbot permettant de renseigner sur:

- Le règlement de scolarité
- Les nombreux cursus offerts
- L'articulation entre les deux

… en se basant sur les documents officiels existants.

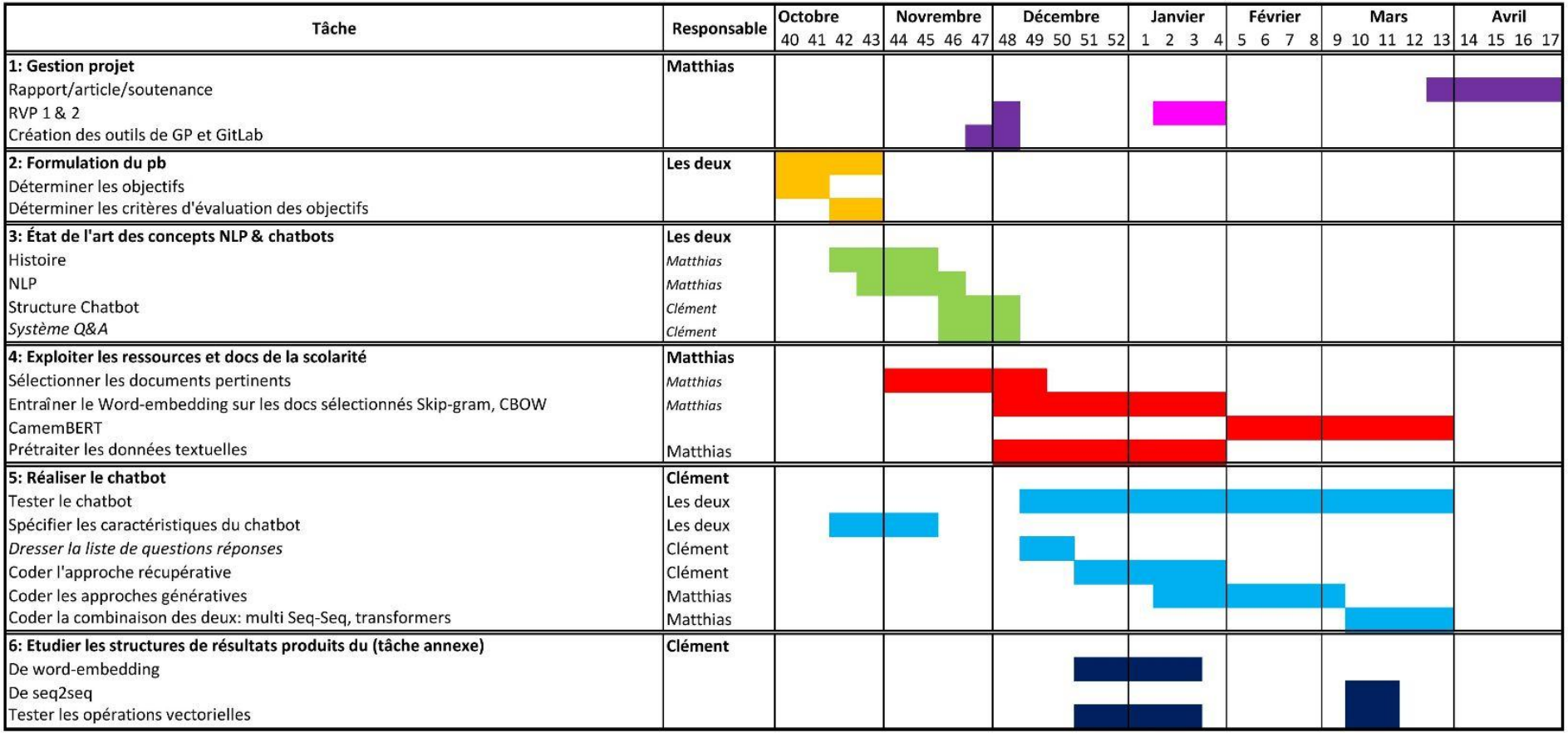
Objectifs et évaluations

Éléments d’évaluation des résultats:

- Correction de la langue (fautes de grammaire, syntaxe, accords)
  - Vérifier sur un ensemble de phrase générées la qualité de la langue.

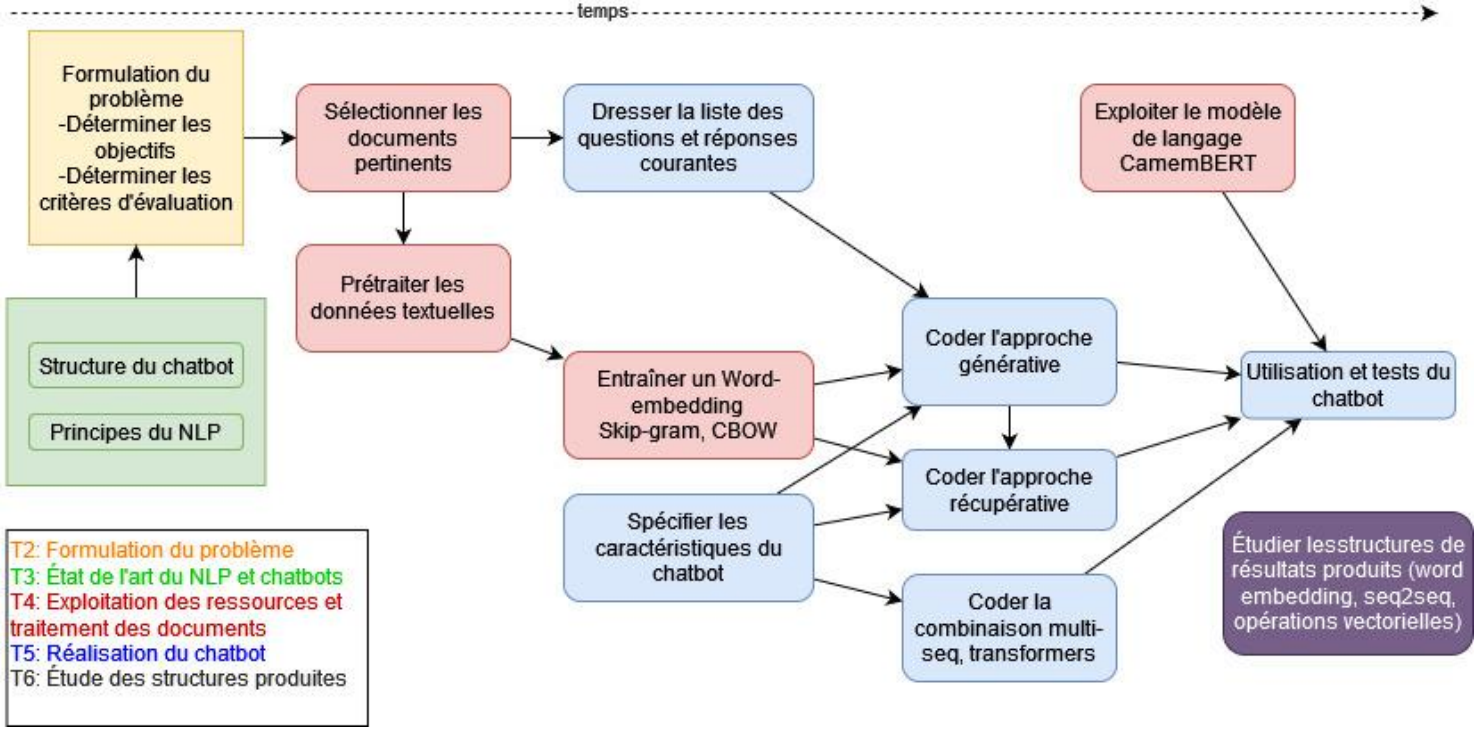
- Evaluation de la pertinence des réponses proposées (précision + justesse + exhaustivité)
  - Sur une cinquantaine de questions, on vérifie à la main la cohérence des réponses

Diagramme GANTT prévisionnel



Tâche	Responsable	Octobre 40 41 42 43	Novrembre 44 45 46 47	Décembre 48 49 50 51 52	Janvier 1 2 3 4	Février 5 6 7 8	Mars 9 10 11 12 13	Avril 14 15 16 17
1: Gestion projet	**Matthias**							
Rapport/article/soutenance								
RVP 1 & 2								
Création des outils de GP et GitLab								
2: Formulation du pb	**Les deux**							
Déterminer les objectifs								
Déterminer les critères d'évaluation des objectifs								
3: État de l'art des concepts NLP & chatbots	**Les deux**							
Histoire	*Matthias*							
NLP	*Matthias*							
Structure Chatbot	*Clément*							
Système Q&A	*Clément*							
4: Exploiter les ressources et docs de la scolarité	**Matthias**							
Sélectionner les documents pertinents	*Matthias*							
Entraîner le Word-embedding sur les docs sélectionnés Skip-gram, CBOW	*Matthias*							
CamemBERT								
Prétraiter les données textuelles	Matthias							
5: Réaliser le chatbot	**Clément**							
Tester le chatbot	Les deux							
Spécifier les caractéristiques du chatbot	Les deux							
Dresser la liste de questions réponses	Clément							
Coder l'approche récupérative	Clément							
Coder les approches génératives	Matthias							
Coder la combinaison des deux: multi Seq-Seq, transformers	Matthias							
6: Etudier les structures de résultats produits du (tâche annexe)	**Clément**							
De word-embedding								
De seq2seq								
Tester les opérations vectorielles								

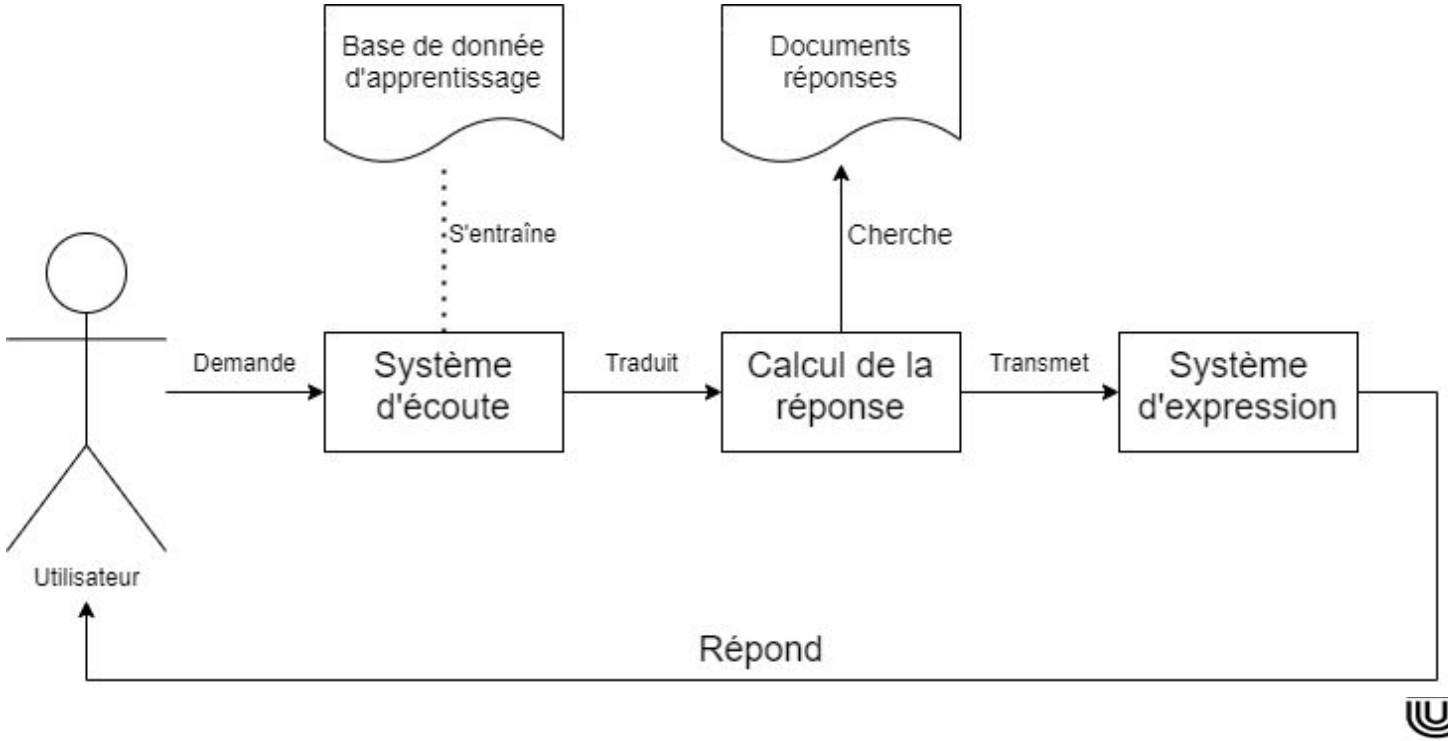
Diagramme PERT



- Les diagrammes et prévisions sont conditionnés cependant par les résultats.

Tâche 3 : Etat de l'art sur le NLP et les Chatbots

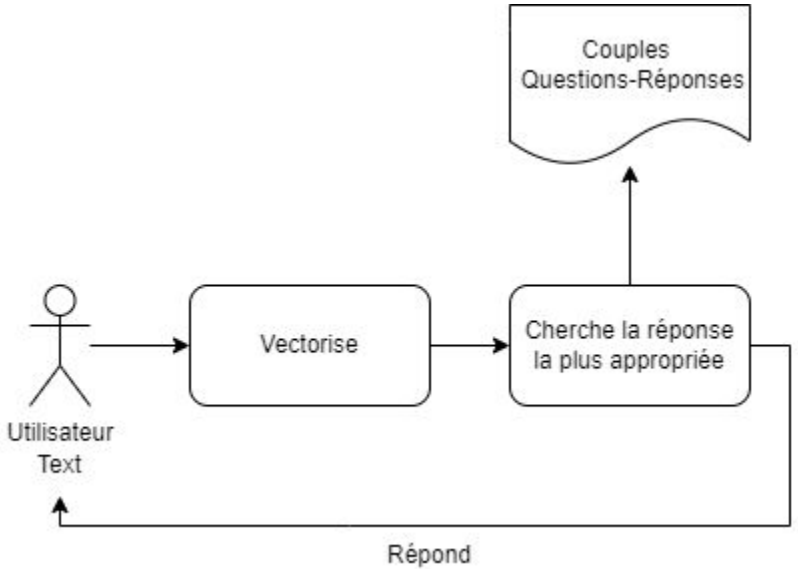
Structure générale d'un chatbot



Tâche 3 : Etat de l'art sur le NLP et les Chatbots

Une approche pour un Chatbot Q&A: l'approche récupérative

- Utilise un ensemble de couples questions-réponses
- Vectorise la question de l'utilisateur
- Renvoie la réponse préétablie dont la question associée est la plus similaire à celle de l'utilisateur

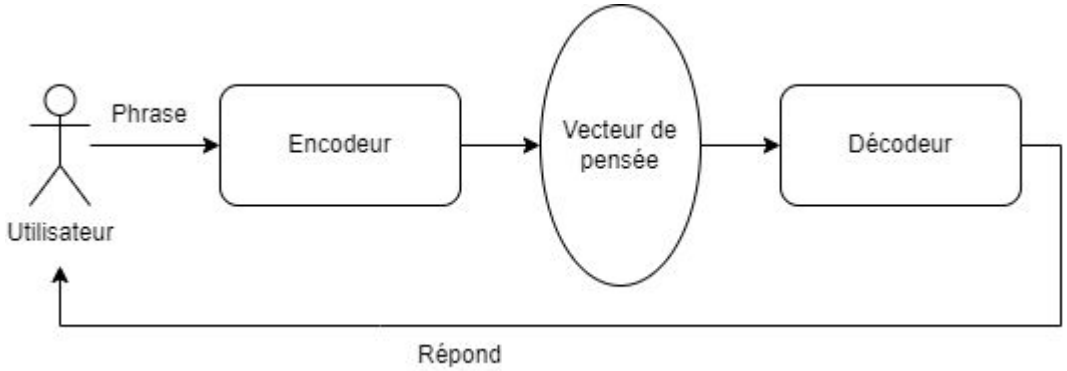


Tâche 3 : Etat de l’art sur le NLP et les Chatbots

Une approche pour un Chatbot Q&A: l’approche générative

Le modèle Seq2Seq:

- Transforme directement les questions de l’utilisateur en réponses
- Composé de 3 parties
- Intègre le contexte dans la phrase



Améliorations:

- Prise en compte de l’attention
- Transformers

Gestion du budget

- Frais du personnel : Coût du travail des encadrants (tuteurs + conseillers) : 40€/heure-homme x 30 semaines x (4heures-tuteurs + 1heure-conseiller communication) / semaine = 6000€

- Frais utilisation d'un ordinateur pour le machine learning

- Pas de coût lié à l'utilisation à l'utilisation de matériel spécifique ni d'occupation des bâtiments du LIRIS

Résultats partiels

T4: Sélection et exploitation des documents de scolarité

Documents présents sur https://campus.ec-lyon.fr
2 formes:

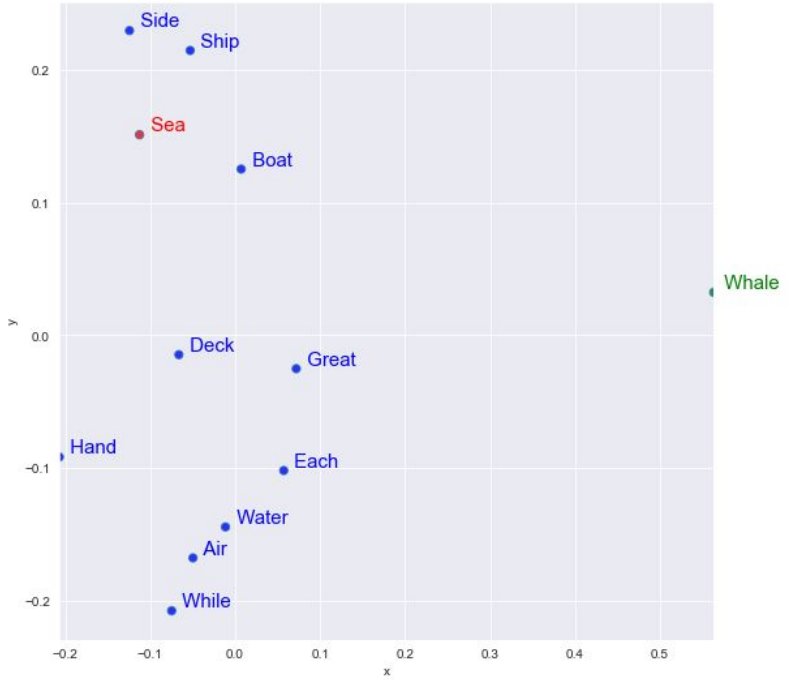
- texte brut format HTML
- documents PDF

Utilisation des bibliothèques PDFminer et gensim, sklearn pour tokéniser le texte de manière automatique.
Difficultés:

- Encodage des PDF + style de rédaction non verbal des docs
- Documentations obselètes / récents changements de syntaxe
- Puissance de calcul limitée & gestionnaire de paquets Conda pour le moins capricieux

Résultats partiels

T4: Sélection et exploitation des documents de scolarité



Premier test de word embedding par l'implémentation Word2Vec:

Moby Dick

Oeuvre de 115 000 mots, langue anglaise, 1928
Ci-contre: approche skip-gram, fenêtre de taille 5, sur 300 dimensions.

Conclusion

- La recherche biblio et les premiers word embeddings ont demandé énormément de travail
- La cadence de production de code s’accélère maintenant fortement.
- Les tâches sont maintenant claires et nous savons quelles directions explorer.
- Le résumé détaillé de l’état de l’art vous sera communiqué rapidement

Merci de votre attention