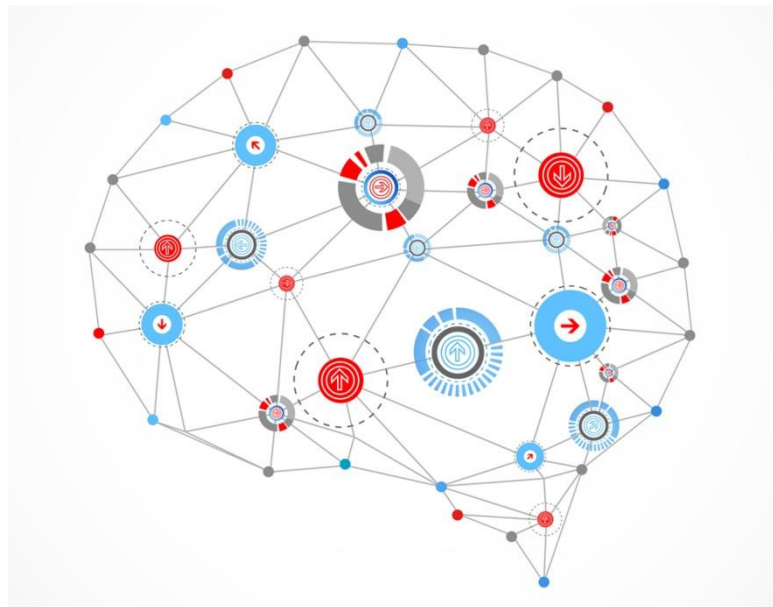


PROJET D'OPTION

Deep Learning et traitement de langues naturelles et Text Mining

Damien Douteaux – Vincent Hocquemiller – Louis Redonnet



Introduction

Contexte du sujet

Objectifs

Livrables

Gestion de projet

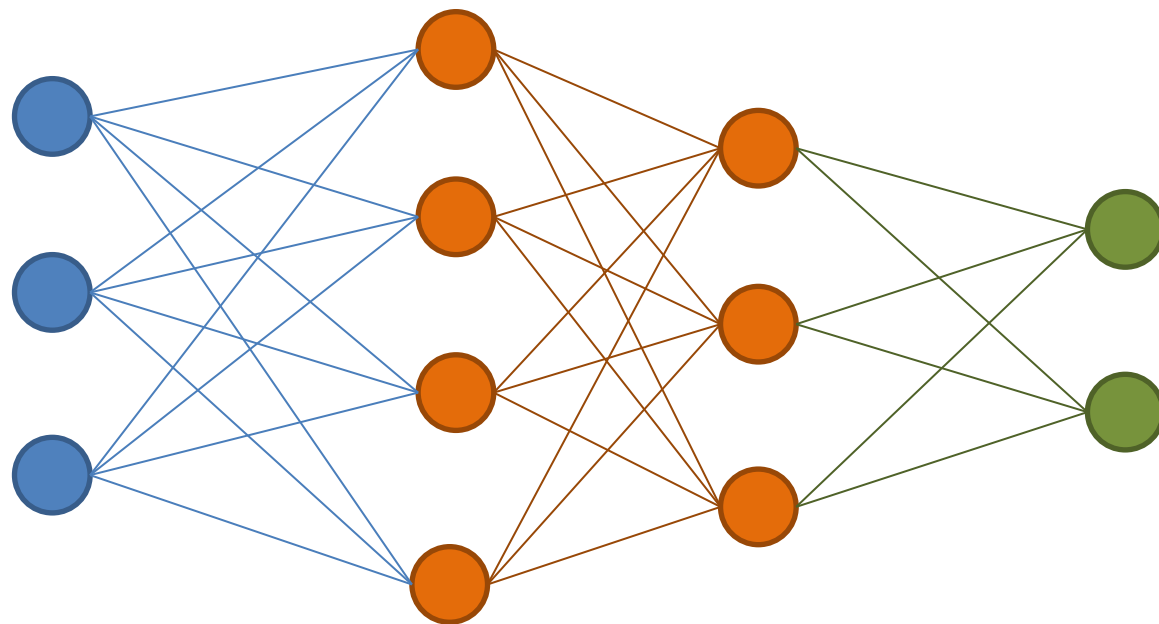
Conclusion

- ⊙ Introduction
- ⊙ Contexte du sujet
- ⊙ Objectifs
- ⊙ Livrables
- ⊙ Planification et gestion de projet

Deep Learning et traitement de langues naturelles et Text Mining

- ⊙ Méthode récente.
- ⊙ Couches de neurones (unités de traitements).
- ⊙ Application pour :
 - Reconnaissance faciale/vocale
 - Vision par ordinateur
 - Traitement automatisé du langage

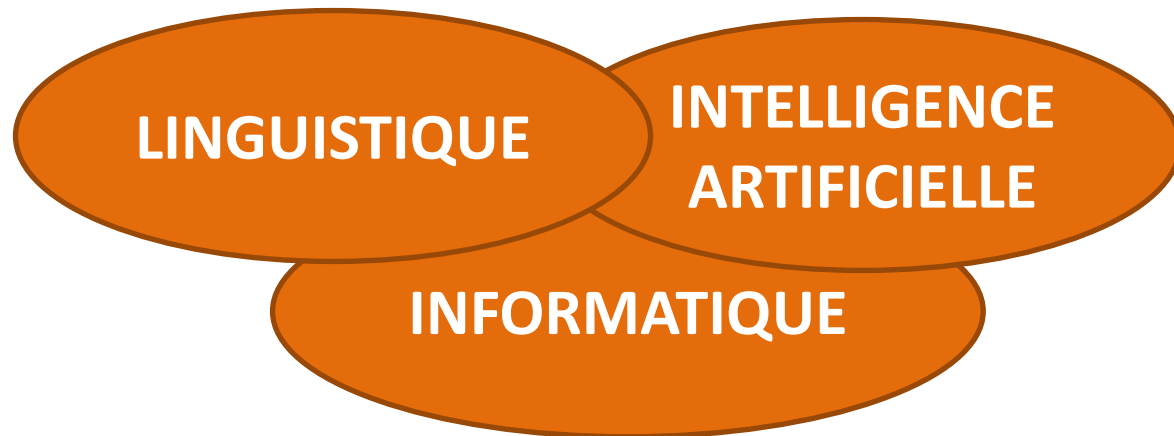
Deep Learning et traitement de langues naturelles et Text Mining



Entrée → Couches cachées → Sortie

Deep Learning et **traitement de langues naturelles** et Text Mining

Le traitement (automatique) des langues naturelles est l'exploitation du langage humain par les outils informatiques.



Deep Learning et traitement de langues naturelles et **Text Mining**

- ⊙ Recenser, structurer des données textuelles.
- ⊙ Approche globale et grossière du texte (sans s'attarder sur le sens).
- ⊙ Historiquement différents modèles :
 - Bag of words ;
 - N-Gram.
- ⊙ Applications dans de nombreux domaines :
 - Page ranking ;
 - Filtrage des communications ;
 - Intelligence économique (détection de sujets clés).

- ⦿ Fournir un état de l'art précis sur le contexte.
- ⦿ (Ré)utilisation d'un réseau de neurones pour appliquer les conclusions de nos recherches sur notre base de données.

- ⊙ Cibler les champs d'application du Deep Learning.
- ⊙ Déterminer les principaux objectifs du Text Mining.
- ⊙ Définir les objectifs finaux du projet.

- ⊙ Cibler une ou deux applications.
- ⊙ Exploiter une base de données :
 - À partir d'un réseau pré-entraîné,
 - En entraînant manuellement le réseau.
- ⊙ Travail en collaboration avec un PAr (pour la BDD).

- ⦿ Recherche de sources « fiables ».
- ⦿ Examen des documents et extraction des concepts.
- ⦿ Synthèse des documents.
- ⦿ Détermination des objectifs pour la suite du projet.

RÉCUPÉRATION BASE DE DONNÉES

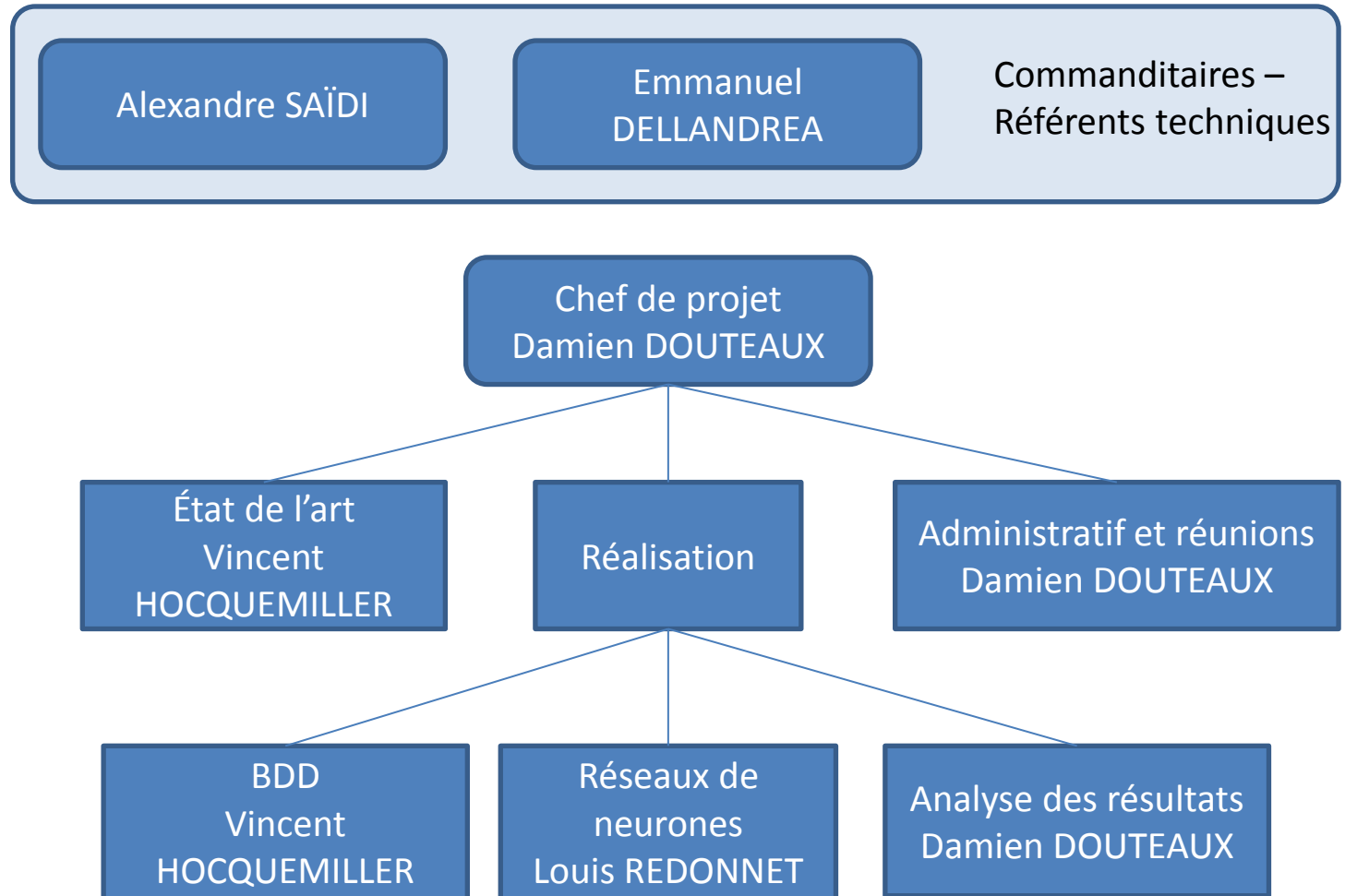
- ⊙ Base existante (en lien avec le PAr).
- ⊙ Préparation des données (nettoyage, analyse).

RÉCUPÉRATION D'UN RÉSEAU PRÉ-ENTRAÎNÉ

- ⊙ Déterminer les « bonnes » sources.
- ⊙ Analyse du réseau.
- ⊙ Étude des méthodes pour le faire évoluer.

ÉVALUATION MODÈLE

- ⊙ Utilisation du réseau avec X-validation.
- ⊙ Comparaison de plusieurs modèles (si possible).
- ⊙ Mise en perspective avec la littérature.
- ⊙ Pertinence de la méthode





- ⦿ Grande importance de l'état de l'art
- ⦿ Un certains nombres de documents déjà récupérés, mais à analyser.
- ⦿ Originalité dans l'utilisation du réseau de neurones
- ⦿ Objectifs « sur mesure » en fonction des centres d'intérêts et des données disponibles.