UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Faciliter la conception d'un assistant conversationnel avec le clustering interactif

THÈSE

présentée et soutenue publiquement le 01 mai 2023

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Erwan SCHILD

Composition du jury

Présidents :	Dr. Pascal CUXAC (à demander)
Rapporteurs :	Dr. Thomas LAMPERT (attendre réponse) Dr. (avoir une femme !)
Examinateurs :	Dr. Adrien COULET (à demander) Dr. (avoir une femme !)
Invités :	Dr. Gautier DURANTIN (à demander) Dr. Mathieu POWALKA (à demander)
Encadrants :	Dr. Jean-Charles LAMIREL Dr. Florian MICONI

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

M i s e n p a g e a v e c l a c l a s s e t h e s u l .

Résumé

Le résumé.

Mots-clés: chat, chien, puces.

Abstract

In computational geometry many search problems and range queries can be solved by performing an iterative search for the same key in separate ordered lists. In Part I of this report we show that, if these ordered lists can be put in a one-to-one correspondence with the nodes of a graph of degree d so that the iterative search always proceeds along edges of that graph, then we can do much better than the obvious sequence of binary searches. Without expanding the storage by more than a constant factor, we can build a data-structure, called a fractional cascading structure, in which all original searches after the first can be carried out at only log d extra cost per search. Several results related to the dynamization of this structure are also presented. Part II gives numerous applications of this technique to geometric problems. Examples include intersecting a polygonal path with a line, slanted range search, orthogonal range search, computing locus functions, and others. Some results on the optimality of fractional cascading, and certain extensions of the technique for retrieving additional information are also included.

Keywords: cat, dog, flees.

Remerciements

Par la présente, je souhaites remercier :

☐ ma femme

☐ ma tortue

☐ ma famille

☐ mes amis

☐ mes collègies

☐ mes encadrants

☐ chatGPT

☐ ...

Je dédie cette thèse
à quelqu'un de bien.

Table des matières

Table des figures

Liste des tableaux

Introduction

1 Accroche sur l'essor des chatbots

☐ Des enjeux ou problèmes actuels

- Accessibilité à l'information : o Grosses bases documentaires, pas toujours ordonnées ;
- Relations client à distance o Besoin d'un accessibilité h24 ;

☐ Utilisation de plus en plus fréquente des chatbots

- Description succinte ;
- Cas d'usage usuels ;
- Tous les canaux d'utilisation ;
- Avantages et Dérives potentiels de l'utilisation (emploi, biais, pertinence, ergonomie, ...) ;

☐ Révolution techniques fréquentes (règles, classification, modèles)

- Moteurs de règles : o Basé sur la détecté de mots clés, o (+) facile à mettre en œuvre, o (-) peu robuste au langage naturel, o Paramétrage des réponses ;
- Paramétrage intentions-entités : o Classification d'intention et/ou détection d'entités, o (+) plus robuste au langage naturel, facile à paramétrer, réponses contrôlées, o (-) demande de l'entrainement, des données, ..., o Paramétrage des réponses ;
- Génération de réponse : o Réseau de neurones avec attention, o Transformers, o (+) plus robuste, o (-) plus complexe à mettre en œuvre, réponses non contrôlées, o Réponses non paramétrées ;
- Approche hybride : o Cumul des trois approches pour cumuler certains avantages suivant les besoins ;

☐ Besoins de données

- Collecte de données spécifiqus au domaine traité : o extraction de base de données (solution simple), o collecte manuelle (organisation complexe, biais de collecte), o scraping (pas toujours fiable) ;
- Nombreux biais : o Biais,, o Réglementation, o Compétences (NOTRE COEUR DU SUJET), o ...

2 Ennoncé de la problématique

☐ Le nerf de la guerre = les données

☐ Cependant, la collecte est difficile

☐ On va étudier l'organisation usuel des projets de création d'un chatbot et voir comment l'assister

Chapitre 1

État de l'art : concevons un jeu de données

Dans cette partie, nous allons faire un état des lieux des méthodes pour créer le premier jeu de données nécessaire à l'entrainement d'un assistant conversationnel. Cela comprend une description des acteurs du projet, un rappel de l'organisation usuelle en fonction de leur compétence, et une énumération des problèmes et solutions les plus communs. **Rappel des contraintes industrielles**

Sommaire

1.1 Rappel sur le fonctionnement usuel d'un chatbot

• Description du cas d'un chatbot "classique" modélisé à base d'intention et d'entités o On se concentre sur ces implémentations car on peut y controller les réponses (image de marque en jeu)

• Classification d'intention (règles, classification supervisée, ...)

• Extraction d'entités (règles, ner, ...)

• Mapping des réponses sur la base du couple $(intention, enites)$

• **CITATION**

1.2 Les étapes usuelles de conception d'un chatbot

Préambule : l'organisation peut bien entendu varier suivant les contextes, mais la description qui suit est représentative des organisation principales

1.2.1 Définition des acteurs

• Data scientistes : o Experts en IA o Peu de connaissance métier, i.e. peu de regarde critique sur la pertinence des résultats (autre que statistique)

• Expert métier : o Pas de connaissance en IA, i.e. nécessitent des formations o Connaissance métier forte, i.e. peuvent décrire la pertinence d'un résultat

• Chef de projet o Pas de connaissance en IA o Pas de connaissance métier o Connaissance du besoin (hypothèse non vérifiée car parfois ils ne savent pas ce qu'ils veulent dû à la méconnaissance des capacités de l'IA)

1.2.2 Cadrage du projet

• Objectifs : o Clarification du besoin, o Définition du périmètre couvert (i.e. les fonctionnalités et réponses à proposer),

• Livrable : un cahier des charges

1.2.3 Collecte des données

• Souvent pas de données à disposition : o En R&D, "80%" sur la recherche d'algo sur des données publiques, d'où le besoin de datascientists, o En entreprise, "80%" sur la gestion des données privées/spécifiques sur des algo connus, d'où le besoin d'experts métiers ;

• Risque de biais dans les données : o Biais d'échantillon : la collecte ne représente pas la réalité, o Biais de sélection : le trie de la collecte ne représente plus la réalité, o Biais de confirmation : on garde les données qui nous arrangent, o Biais de valeur : les données ne sont pas éthiquement représentatives, o Biais de contexte : les données d'un cas d'usage ne sont pas toujours réutilisables pour un autre cas d'usage (ex : différence entre les jargons des AV clients et celui des AV conseillers) ; o **A COMPLETER**

• Livrable : une collecte de données brutes

1.2.4 Modélisation d'une structure et Labellisation des données

• Le coeur "métier" de la création du projet ;

• Objectif : Définition d'une modélisation sur la base des besoins attendus restreints au périmètre à couvrir ;

• En théorie : o Intention : verbe d'actions, o Entités : informations complémentaires, personnes, date, lieux, montants, noms de produits, ... ;

• Complexité de la tâche : o Intention abstraite : définition difficile voir subjective, ... o Annotation difficile : différence entre théorie et pratique, données ambigues, ... o Plusieurs itérations car modélisation trop théorique / pas pratique o Besoins de beaucoup de formation (pour donner la compétence aux experts) et d'atelier (pour se mettre d'accord)

• Livrable : un jeu de données annotées

1.2.5 Entrainement et tests

• Le coeur "technique" de la création du projet ;
• Objectif : avoir un modèle qui soit adapté à son utilisation en production
• En théorie : o Split en train et tests o Entrainement et tests o Association des réponses
• Complexité de la tâche : o Modélisation précédente pas toujours adaptée : OK pour un métier, mais pas possible à entrainer à cause de déséquilibre, de manque de données, ... o Algorithme fixe mais données variables : savoir quelle modélisation est la plus adaptée est compliqué à deviner o Réponses pas toujours adaptées aux questions : décalage entre entrainement (modélisation théorique) et réponse (modélisation pratique)

1.2.6 Déploiement de la première version

• RAS
• Parfois la modélisation est décalée par rapport à l'utilisation en production o Comportement en moteur de recherche avec des questions courtes o Vocabulaire non maitrisé par les utilisateurs o problème d'ergo ou d'expérience utilisateur

1.2.7 Amélioration continue

• Vérification du comportement ;
• Ajustement du modèle ;
• Déploiement des versions suivantes.

1.3 Zoom sur la partie Modélisation et Labellisation de la bse d'apprentissage

1.3.1 Création « manuelle »

• Enchainement de plusieurs ateliers/cycles : o Définition d'une structure en atelier et Annotation des données o Premier conflit : La structure est trop théorique o Redéfinition et Ré-annotation o Second conflit : Les structure ou les données ne sont pas adaptées o Collecte complémentaire, Redéfinition et Ré-annotation
• Avantages : o Transmission progressive du savoir aux datascientist o Test des modélisations potentielles
• Inconvénients : o Nombreux ateliers o Nombreuses remises en questions / aller-retour de conception o L'avis initiale sur le périmètre à couvrir est flou quand cela concerne une centaine de demandes clients o Se base sur de la connaissance que les experts métiers n'ont pas o Comment les aider dans ce problème d'organisation ?

1.3.2 Création assistée par des regroupements non-supervisés

• Constat : o Pour des jeux de données à taille humaine (moins de 20.000 données), le premier tri est parfois "optimisé" manuellement sur la base des patterns commun (ordonnancement alphabétique)
• Solution : o Un clustering pourrait simplifier cette tâche ! o Rappel : grandes lignes du fonctionnement d'un algorithme de clustering ? o NB : une section ou une annexe détaillera les algorithmes de clustering les plus utilisés
• Avantages : o Regroupement automatique o Découverte de la structure

• Inconvénients : o Les résultats sont souvent peu pertinents o Similarité par entités, et pas par intentions o Nuances métiers non comprises o Plusieurs soucis si le jeu de données est déséquilibré ou spécifique o Absence d'un modèle de langue spécifique au contexte... o parfois besoin d'hyperparamètres compexes à déterminer

1.3.3 Conception assistée par des regroupements semi-supervisés

• Solution : o On peut envisager ainsi de corriger le clustering en y insérant des contraintes métiers [Lampert et al., 2018] o Méthodes semi-supervisée o NB : une section ou une annexe détaillera les algorithmes de clustering sous contraintes

• Interactions possibles avec le clustering (sur la base de proposition de l'humain) o Sur les données / sur le résultat : ajouts de contraintes sur les données, suppressions ou modifications manuelles de données, réorganisation manuelles des clusters, ... o Sur les paramètres : modifier les hyper-paramètres, modifier le nombre de clusters, modifier les embeddings, utiliser d'autres algorithmes, ... o Besoin de visualisation : vue des contraintes, de la représentation vectorielle, ...

• Avantage : o On a réglé les problèmes de pertinence en ajoutant des contraintes

• Inconvénients : o Choisir comment modéliser ces contraintes peut être complexe o Surtout énorme en ajoutant des contraintes o Choisir les contraintes pertinentes est une tâche difficile

1.3.4 Conception basée sur des méthodes d'apprentissage actif

• Solution : o On peut demander à la machine de définir les contraintes dont elle a besoin pour s'améliorer / confirmer son comportement o On peut séparer et cibler les tâches pour que le clustering se nourrissent des commentaires de l'expert et que l'expert corrige ce qui semble utile au clustering o Sous-entendu : Préférer la collaboration à la supériorité (que ce soit celle de la machine ou celle de l'expert) o NB : une section ou une annexe détaillera les interactions possibles entre homme et machine

• Interactions possibles avec le clustering (sur la base de propositions de la machine) o Sur les données / sur le résultat : proposition de suppression de données abhérrantes, proposition d'ajout de contraintes à des endroits stratégiques, ... o Sur les paramètres : réévaluation des paramètres, combiner plusieurs algorithmes et synthétiser le résultat, ...

• Avantage : o On a réglé les problèmes de pertinence et de coûts en ajoutant des contraintes

• Inconvénients / problème à résoudre : o Accepter de collaborer avec la machine (problème UX, ergo, accompagnement au changement) o Il faut prouver cette méthode

Chapitre 2

Proposition d'un Clustering Interactif

Sommaire

Transition / Récap : On a vu que :

• Le travail est principalement manuel : comment peut-on l'assister ?

• La définition de la structure de classes est un mélange de connaissances métiers et de regroupement manuel sur la base de patterns linguistiques commun : faisons du clustering !

• Mais le clustering est souvent peu pertinent pour un usage métier : intégrons-y des interactions avec la machine !

• Les intercations non guidées sont fastidieuses : ajoutons de l'active learning !

Nous proposons donc notre version d'intercative clustering.

NB : la démonstration de cette proposition sera démontrée dans la partie 3.

2.1 Description théorique de la méthode

Nous proposons de combiner les techniques vues précédemment :

• Clustering sous contraintes o Kmeans : Classique, Incontournable, Rapide, Efficace o Hiérarchique : Lent mais facile à implémenter o Spectral : Permet des topologies complexes o DBScan : Classique, Incontournable, Rapide, Efficace, Peu d'hyperparamètres o Affinity propagation : o Metric learning : Lent mais plus adapté au corpus o ...

• Echantillonnage de contraintes à annoter o Random ou Pseudo-random o Farhtest : Scinder les gros clusters o Closest : Redéfinir la position des frontières de clusters o ...

• Annotation de contraintes o MUST-LINK / CANNOT-LINK / SKIP o « Répondriez-vous de la même manière à ces deux demandes ? »

• Boucle itérative entre clustering, échantillonnage et annotation o Améliorer le résultat précédent o Autant de boucle que « nécessaire » o Avoir le clustering le plus efficace pour avoir de bon résultats o Avoir l'échantillonnage le plus efficace pour améliorer le plus efficacement o Avoir une annotation sans ambiguïté pour ne pas biaiser la construction itérative

• Analyses o Analyse de l'évolution de l'accord clustering->annotation o Analyse des patterns linguistiques pertinents o Analyse de la formation de clusters (taille, répartition, ...) • NB : Réutilisation de schéma etc

2.2 Espoirs de la méthode proposée (? ?ETAT DE L'ART ? ?)

• Moins de formations, d'ateliers, ...
• Se concentrer sur son domaine de compétence (i.e. pas de datascience pour les experts métiers)
• Permettre de trouver la base d'apprentissage
• Méthode réaliste / pas trop coûteuse• ...

2.3 Description technique et implémentation

• cognitivefactory.interactive-clustering : Gestion des données
• cognitivefactory.interactive-clustering : Gestion des contraintes + conflits
• cognitivefactory.interactive-clustering : Algorithmes de clustering
• cognitivefactory.interactive-clustering : Algorithmes de sampling
• cognitivefactory.interactive-clustering-gui : Interface d'annotation
• cognitivefactory.interactive-clustering-gui : Interface d'analyse
• NB : captures d'écrans pour donner un aperçu, puis redirection vers les annexes

2.4 Protocole d'utilisation : Mode d'emploi associé (? ?CONCLUSION ? ?)

• Collecte des données
• Itération de clustering > échantillonnage > annotation
• A chaque conflit : correction nécessaire
• A la fin d'un clustering : caractériser la pertinence métier avec FMC
• A chaque itération : voir l'évolution par rapport à la précédente
NB : la démonstration de cette proposition de protocole sera démontrée dans la partie 3.

Chapitre 3

Etude de la méthode

Sommaire

Transition / Récap : Cette méthode étant nouvelle, plusieurs études sont nécessaires :

1. Est-ce que la méthode marche/converge ?
2. Quelle est la meilleure implémentation ?
3. Quels sont les coûts à investir ?
4. Quel est l'impact d'une différence / erreur d'annotation ?
5. Comment interpréter les résultats et leur donner du sens ?
6. Quand s'arrêter ?

(à formuler sous la forme d'hypothèse de travail à vérifier ? et/ou formuler/rappeler ces hypothèse en début de chaque sous-partie)

==> Remarque en préambule : C'est compliqué de remettre en cause l'annotation manuelle avec des vérité terrain conçues manuellement... Certaines études ne sont donc pas faisables

3.1 Etude de viabilité : « est-ce que la méthode marche ? »

3.1.1 Etude de convergence vers une vérité terrain établie

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain o Jeu de données : Carte bancaire (500 questions) o Cf. EGC/IJDWM

• Résultats : o Les contraintes sont respectées o La vérité terrain est atteinte (La vmeasure atteint 100%) o L'ajout de contraintes permet vite d'être meilleur qu'un clustering simple

• Conclusion : la méthode est viable : on trouve bien vérité terrain o Avantage : Pas besoin de définir une structure, elle se découvre toute seule o Avantage : Besoin de peu de connaissance en IA (same ? oui ou non) o Inconvénient : Besoin de « beaucoup » de contraintes o ==> IL FAUT UNE OPTIMISATION (IV.B.)

• Discussion : balance entre annotations de contraintes et clustering o Si pas assez de contraintes : alors clustering pas assez pertinent ! o Si 100% de contraintes : alors résultat trop subjectif ! o Trouvons une juste milieu : supposons 80% pour la suite (annotation partielle)

3.2 Etude technique : « quelle est la meilleure implémentation ? »

3.2.1 Etude des paramètres optimaux

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain et étudier des itérations clés pour déterminer quelle implémentation est la plus efficace. o Jeu de données : Carte bancaire (500 questions) o Paramètres étudiés : prétraitement, vectorisation, sampling, clustering o Métrique : VMeasure du clustering obtenue avec la vérité terrain en fonction du nombre de contraintes annotées o Cf. EGC/IJDWM

• Résultats : o Tous les paramètres sont importants, surtout le sampling et le clustering o Meilleure implémentation trouvée

• Conclusion : o Il y a donc moyen d'optimiser la méthode pour annoter moins de contraintes

• Remarques : o Expliquer le choix du kmeans : rapide, efficace, itération rapide o Expliquer le choix du closest : favorise les MUST-LINK o Expliquer le non-choix du farthest : favorise CANNOT-LINK, mais n'aide pas

3.3 Etude des coûts : « quels sont les coûts à investir ? »

3.3.1 Etude du temps d'annotation

• Expérience : Faire annoter des contraintes par plusieurs annotateurs o Jeu de données : Carte bancaire (500 questions) ou Titre de journaux (? ? ?)

• Résultats : (**! !TODO ! !**)

• Conclusions : (**! !TODO ! !**)

• Discussions : Définition d'un batch moyen

3.3.2 Etude du temps de calcul

• Expérience : Estimer le temps de calcul de chaque algorithme o Jeu de données : Carte bancaire (1000 questions, artificiellement augmenté jusqu'à 1000 données) o Estimation des facteurs influents : nombre de données, nombre de contraintes, nombre de clusters, hyper-paramètres, effets aléatoires, . . .

• Résultats : o Estimation des temps de calcul et des facteurs influents

• Conclusions : o Certains algorithmes sont longs. . . o Définition d'une fonction d'estimation du temps de calcul pour chaque algo

• Discussion : balance entre performance et coût temporel o Les itérations doivent être fluides o On a plus à gagner à ajouter des contraintes qu'attendre un algo trop long o Un optimal serait d'annoter des batch d'une durée d'un algo pour ne pas avoir trop à attendre entre deux itérations

3.4 Etude des erreurs : « quel est l'impact d'une différence d'annotation ? »

Préambules

• Source potentielle des différences : o Différence de points de vue, o Erreurs d'inattention, o Données ambigües ;

• Identification des différences : o Relectures / revue d'annotation, o Détection des incohérences entre contraintes, o Adjudication / ajout de redondance ;

3.4.1 Etude de l'impact d'une erreur par simulation

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » en simulant un pourcentage d'erreur d'annotation o Jeu de données : Carte bancaire (500 questions), o Pourcentage d'erreur variable, o Prise en compte ou non de la résolution de conflits ;

• Résultats : o Peut gravement impacter le résultat, o N'est pas détectable sans redondance ;

• Conclusions : (**! !TODO ! !**)

• Discussions : importance de la fiabilité des annotations : o Besoin de savoir ce que l'on veut : Ajouter de l'adjudication en début de projet pour confronter rapidement les visions, o Besoin de limiter les erreurs : Ajout de redondance ou de stratégie de vérification (de nouvelles méthode de sélection) mais cela entraîne un surcoût, o Préférer passer l'annotation que de forcer une annotation ambiguë ;

3.4.2 Etude de ré-annotation d'un chatbot existant

• Expérience : Ré-annoter un AV existant pour identifier les inconvénients pratiques : o Jeu de données : Moyens de paiements (8000 questions de production), o Deux annotateurs avec des

règles « floues » (comme c'est le cas en initialisation) mais avec un « objectif » commun (arbre de dialogues & réponses « connues »), o Revues d'annotations à mi-projet ;

• Résultats : o 23% de désaccord, mais au moins 77% d'accord sans concertation ! o Après revue d'annotation : (???), o Expériences interrompue ;

• Conclusions : (**!!TODO!!**)

• Discussions : Quelques conseils : o En début de projet : Adjudication pour confirmer les visions, o Correction des incohérences le plus rapidement possible, o D'autres sélection pour ajouter de la redondance, o Mais cela représente un cout supplémentaire,

3.5 Etude métier : « comment interpréter les résultats et leur donner du sens ? »

Préambule de définition de la FMC ou référence à l'état de l'art / annexe :

3.5.1 Etude de la caractérisation du clustering par FMC

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain et étudier la FMC du clustering en cours : o Jeu de données : Carte bancaire (500 questions), o Métrique : VMeasure adaptée pour comparer la FMC d'un clustering et la FMC de la vérité terrain ;

• Résultats : (**!!TODO!!**)

• Conclusions : (**!!TODO!!**)

• Discussions : (**!!TODO!!**)

3.6 Etude d'arrêt : « quand le résultat est-il satisfaisant ? »

3.6.1 Etude d'un prérequis d'arrêt : la cohérence du clustering obtenu

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain et étudier si un classifier entrainé sur cette base est stable, i.e. s'il arrive à retrouver sa base d'apprentissage o Jeu de données : Carte bancaire (500 questions), o Remarque : si la base d'apprentissage n'est pas stable, le clustering doit encore être modifié. . .,

• Résultats : (**!!TODO!!**)

• Conclusions : (**!!TODO!!**)

• Discussions : (**!!TODO!!**)

3.6.2 Etude d'un critère d'arrêt : l'accord entre un batch d'annotation et le clustering précédent

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain et étudier l'évolution des accords/désaccords entre l'annotateur et le clustering o Jeu de données : Carte bancaire (500 questions), o Remarque : si l'accord est maximal, l'annotateur n'a plus de valeur ajoutée car le clustering n'est jamais modifiée,

• Résultats : (**!!TODO!!**)

• Conclusions : (**!!TODO!!**)

• Discussions : (**!!TODO!!**)

3.6.3 Etude d'un critère d'arrêt : la similitude entre deux clustering consécutifs

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain et étudier l'évolution des similitudes entre le clustering courant et le clustering précédent o Jeu de données : Carte bancaire (500 questions), o Remarque : si la similitude est forte, c'est que le clustering devient stable,

• Résultats : (**!!TODO!!**)
• Conclusions : (**!!TODO!!**)
• Discussions : (**!!TODO!!**)

3.6.4 Etude d'un critère d'arrêt : la similitude entre deux FMC de clustering consécutifs

• Expérience : Faire des itérations de « clustering > échantillonnage > annotation » jusqu'à convergence vers la vérité terrain et étudier l'évolution des similitudes entre la FMC du clustering courant et la FMC du clustering précédent o Jeu de données : Carte bancaire (500 questions) o Remarque : si la similitude est forte, c'est que le clustering devient stable

• Résultats : (**!!TODO!!**)
• Conclusions : (**!!TODO!!**)
• Discussions : (**!!TODO!!**)

3.7 Autres études à réaliser (!!PAS FAIT PAR MANQUE DE TEMPS!!)

3.7.1 Choix du nombre de clusters ==> problème de recherche complexe

o Piste de résolution : plusieurs clusterings + vote collaboratif ? algorithmes sans le nombre de clusters en hyper-paramètres

3.7.2 Impact d'un modèle de langage ==> nécessite de nombreuses données spécifiques au domaine

o Piste de résolution : script d'étude comparative déjà prêt, mais il manque les données opensources...

3.7.3 Paradigme d'annotation (intention vs dialogue) ==> problème d'UX + objectif métier

o Etude Ergo, sort de mon domaine d'expertise

3.7.4 (et plein d'autres que j'ajouterai au fur et à mesure de ma rédaction)

o

Conclusion

1 Rappel de la problématique ? ?

TODO

2 Avantage et limites de la méthodes ? ?

TODO

3 Ouverture ? ?

TODO

Annexe A

Annexe théorique

plein de texte très varié. Une autre page avec plein de texte très varié.

A.1 AAAAAA

Une autre page avec plein de texte très varié. Une autre page avec plein de texte très varié. Une autre page avec plein de texte très varié.

A.2 BBBBBB

Une autre page avec plein de texte très varié. Une autre page avec plein de texte très varié.

A.3 CCCCCCC

Une autre page avec plein de texte très varié.

Annexe B

Annexe technique

plein de texte très varié. Une autre page avec plein de texte très varié.

B.1 AAAAAA

Une autre page avec plein de texte très varié. Une autre page avec plein de texte très varié. Une autre page avec plein de texte très varié.

B.2 BBBBBB

Une autre page avec plein de texte très varié. Une autre page avec plein de texte très varié.

B.3 CCCCCCC

Une autre page avec plein de texte très varié.

Bibliographie

[Lampert et al., 2018] Lampert, T., Dao, T.-B.-H., Lafabregue, B., Serrette, N., Forestier, G., Cremilleux, B., Vrain, C., and Gancarski, P. (2018). Constrained distance based clustering for time-series : a comparative and experimental study. *Data Mining and Knowledge Discovery*, 32(6) :1663–1707.

Glossaire

Années 1980 une decénie de musiques chouettes. 4

Années 1990 une decénie de musiques discutables. 4

comics une bande dessinée à parution régulière. 4

SVM Support Vector Machine. 3