



HAL
open science

Dispositif d'apprentissage automatique collaboratif pour la pratique du débat

Imen Ben Sassi, Hani Guenoune, Alexandre Bazin, Marianne Huchard,
Mathieu Lafourcade, Jean Sallantin

► **To cite this version:**

Imen Ben Sassi, Hani Guenoune, Alexandre Bazin, Marianne Huchard, Mathieu Lafourcade, et al..
Dispositif d'apprentissage automatique collaboratif pour la pratique du débat. JADT 2024 - 17th International Conference on Statistical Analysis of Textual Data, Jun 2024, Bruxelles, Belgique. lirmm-04637347

HAL Id: lirmm-04637347

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-04637347v1>

Submitted on 5 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dispositif d'apprentissage automatique collaboratif pour la pratique du débat

Imen Ben Sassi, Hani Guenoune, Alexandre Bazin, Marianne Huchard,
Mathieu Lafourcade, Jean Sallantin

LIRMM, Université de Montpellier, CNRS, Montpellier, France

imen.ben-sassi@lirmm.fr, prenom.nom@lirmm.fr

Abstract

The AREN-DIA project (ARGumentation and Digital - Didactics & Artificial Intelligence) aims to raise awareness of debate practices within both educational contexts and civil society discussions. The project is realized through the creation and testing of a debate platform. This platform allows for engaging in structured debates based on text, renewing the traditional approach to argumentative exchanges. It incorporates collaborative Natural Language Processing (NLP) technology to enhance the efficiency of the debate process. Our paper focuses on the challenges of the AI axis of the project, specifically, how to design a reinforcement mechanism that encourages users to participate in improving the AI system producing a structured representation of debate discourse. Thus, an automatic procedure complements the debate by suggesting key terms that synthesize the statements made. This indexing serves as the starting point for the machine's analysis and support of the debate. It is subject to completion by users who are invited to validate, invalidate, or supplement these key terms. To resolve semantic ambiguity, we employ an enrichment step for the terms to prepare them for the knowledge extraction operation based on Formal Concept Analysis (FCA). The knowledge, expressed in the form of implications, is utilized to update the relationships in the knowledge base that are leveraged during these processes.

Keywords: Knowledge extraction, Formal concept analysis, Thematic indexing, Semantic augmentation.

Résumé

Le projet AREN-DIA (ARGumentation Et Numérique - Didactique & Intelligence Artificielle) vise à sensibiliser à la pratique du débat dans le cadre de l'éducation et au sein de la société civile. Le projet se matérialise à travers la création et la mise à l'épreuve d'une plateforme de débat. Cette dernière offre la possibilité d'engager des débats structurés à partir d'un texte, renouvelant l'approche traditionnelle des échanges argumentatifs. Elle intègre une technologie collaborative de Traitement Automatique du Langage afin d'augmenter l'efficacité du processus de débat. Notre article se consacrera aux enjeux de l'axe IA du projet, à savoir : comment concevoir un mécanisme de renforcement incitant les utilisateurs à participer à l'amélioration du système d'IA produisant une représentation structurée des propos d'un débat ? Ainsi, une procédure automatique vient compléter le débat en suggérant des termes-clés synthétisant les propos tenus. Cette indexation est le point de départ de l'analyse et de l'accompagnement du débat par la machine. Elle est soumise à une interaction avec les utilisateurs, qui seront invités à valider, invalider ou compléter ces termes-clés. Afin de lever l'ambiguïté sémantique, nous avons recours à une étape d'enrichissement des termes pour les préparer à l'opération d'extraction de connaissances basée sur l'analyse formelle de concepts (AFC). Ces connaissances, sous forme d'implications, sont utilisées pour mettre à jour les relations dans la base de connaissances exploitée.

Mots clés : Extraction de connaissances, Analyse formelle de concepts, Indexation thématique, Augmentation sémantique.

1. Introduction

Dans le cadre de l'intégration progressive de l'intelligence artificielle (IA) dans notre vie quotidienne, le projet AREN-DIA (ARgumentation Et Numérique - Didactique & Intelligence Artificielle)¹ vise à sensibiliser les élèves à la pratique du débat dans le cadre de leur éducation à la citoyenneté. La plateforme de débats en ligne ainsi conçue est également disponible à la société civile, offrant la possibilité d'engager des débats structurés à partir d'un texte, renouvelant l'approche traditionnelle des échanges argumentatifs². Elle présente la particularité d'intégrer une technologie collaborative de Traitement Automatique du Langage en vue d'augmenter l'efficacité du processus de débat. Les résultats des études menées au sein des lycées révèlent que l'utilisation judicieuse de ce logiciel, insérée dans un dispositif didactique approprié, conduit à améliorer les compétences argumentatives des élèves (Bächtold et al., 2023). Dans cette perspective, AREN-DIA se déploie selon un axe didactique et un axe IA (objet de cet article) qui s'intéresse à la manière de concevoir un mécanisme de renforcement incitant les utilisateurs à participer à l'amélioration du système d'IA produisant une représentation structurée des propos d'un débat.

Dans la Section 2, nous abordons le fonctionnement de la plateforme AREN ainsi que celui des procédures automatiques assurant l'accompagnement des débats à savoir l'indexation thématique des propos, la désambiguïsation et l'augmentation sémantique des termes d'indexation, l'extraction de connaissances à l'aide de l'analyse formelle de concepts (AFC) et l'enrichissement de la base de connaissances. Nous nous penchons, dans la Section 3, sur les défis spécifiques rencontrés dans AREN, notamment l'interaction humain-machine et l'apprentissage automatique soutenu par le *crowdsourcing*.

2. Fonctionnement de la plateforme AREN

Le projet AREN-DIA se matérialise par une application web qui offre un espace de débat, réunissant un ensemble d'utilisateurs. Le débat porte sur un texte support publié en amont sur la plateforme. Les utilisateurs interviennent à travers des commentaires exprimant une opinion, une argumentation ou un avis sur une sélection d'un segment du texte support ou un commentaire préalablement publié, créant ainsi des embranchements dans l'arbre général du débat. Outre l'intervention des débattants, une procédure automatique (c.f. Section 2.1) vient compléter chaque commentaire en suggérant des termes-clés synthétisant les propos tenus. Cette opération d'indexation représente le point de départ de l'analyse et de l'accompagnement du débat par la machine. Les utilisateurs peuvent valider, invalider ces suggestions de termes-clés et même les compléter par ceux qu'ils estiment manquants. Afin de lever l'ambiguïté sémantique résultant de la polysémie des termes proposés, nous avons recours à une étape d'enrichissement sémantique des termes (c.f. Section 2.2) pour les préparer à l'opération d'extraction de connaissances (c.f. Section 2.3) basée sur l'AFC. Ces connaissances seront utilisées sous forme d'implications pour mettre à jour les relations dans la base de connaissances exploitée lors de ces processus (c.f. Section 2.4). Nous présentons

¹ Ce projet est financé par l'Agence Nationale de la Recherche : ANR-22-FRAN-0001

² La plateforme est accessible via le lien suivant : <https://portail-aren.lirmm.fr/aren2023>

dans le reste de cette section une IA d'accompagnement de débats, KeepTalk (Knowledge Extraction for Enhanced online Public Talks and Argumentative Learning Know-how), dont un des objectifs est d'extraire des associations nouvelles entre termes à partir des listes de termes-clés des commentaires d'un débat.

2.1. Indexation des propos

Les divers arguments des participants au débat sont contenus dans des textes bruts et non-structurés. L'indexation thématique des commentaires (la sélection du texte, sa reformulation et l'argumentation) des débattants, a pour objectif d'associer ces données textuelles à une représentation structurée permettant de synthétiser les propos par des ensembles de termes-clés, référencés dans des bases de connaissances et pouvant servir de point d'entrée à une procédure automatisée. Les termes extraits peuvent désigner des concepts évoqués dans le texte ou des unités lexicales dont la saillance au sein du commentaire est jugée importante. Cette étape d'extraction de mots-clés s'appuie sur des connaissances externes issues du réseau lexico-sémantique *JeuxDeMots* (*JDM*) (Lafourcade et Le Brun, 2023), et est réalisée par le service *IDÉFIX*³.

2.1.1. *JeuxDeMots* : un réseau lexico-sémantique

JDM est un réseau lexico-sémantique sous forme de graphe orienté. Les nœuds du graphe représentent les termes, tandis que les arcs désignent des relations typées, pondérées et potentiellement annotées entre ces termes. Le graphe représente la polysémie des mots en explicitant des raffinements sémantiques hiérarchisés, où un sens spécifique est affilié au sens général du terme (via une relation de type *r_raft* de poids positif). Basé sur une série de notions, principes et outils originaux, par exemple : la notion de raffinement, la palette des types de relations sémantiques - des liens entre un type de relation et son inverse (*r_isa* et *r_hypo*, par exemple) - l'outil contributif *Diko*, etc. *JDM* est conçu pour une utilisation humaine, mais en premier lieu comme support de connaissances pour des processus d'intelligence artificielle (analyse sémantique de texte, raisonnement, assistance à la prise de décision, résumé automatique, etc.).

2.1.2. *Bellérophon* : raffinement lexical

*Bellérophon*⁴ est un service fourni par la plateforme *JDM*, qui à partir d'une liste de termes, permet de sélectionner les sens probables des termes polysémiques de la liste. Par exemple, à partir de la liste des termes d'indexation qui suit (*•* est un séparateur, *>* introduit un raffinement) :

outil conceptuel • être utile • outil>moyen d'action • MLC • économie locale • moyen d'action • monnaie locale • crise commerciale • monnaie locale complémentaire et citoyenne • outil • financier • Sol-violette • économie • monnaie locale complémentaire • monnaie • outil financier • local

Nous obtenons le résultat d'activation des raffinements ci-dessous :

³ L'outil *IDÉFIX* est accessible via le lien suivant : https://www.jeuxdemots.org/intern_extract.php

⁴ L'outil *Bellérophon* est accessible via le lien suivant : https://www.jeuxdemots.org/intern_desamb.php

outil>*moyen d'action* / 480
monnaie>*argent* / 403
MLC>*monnaie locale complémentaire* / 260
économie>*activité économique* / 234
local>*propre à un lieu* / 216
financier>*finance* / 202
monnaie>*unité monétaire* / 93

local>*bâtiment* / 0
économie>*gestion* / 0
local>*médecine* / 0
économie>*science* / 0
économie>*épargne* / 0
économie>*organisation* / 0
MLC>*Mouvement de libération du Congo* / 0
MLC>*métro léger de Charleroi* / 0

financier>*homme d'affaires* / -23
financier>*gâteau* / -25
financier>*théâtre* / -25
monnaie>*rendu* / -25
outil>*pénis* / -25
monnaie>*pièce de monnaie* / -25
outil>*objet matériel* / -121
outil>*poisson* / -136

Chaque raffinement est associé à une valeur numérique d'activation. Plus la valeur est élevée, plus l'activation est forte, et à l'inverse une activation négative (inhibition) correspond à un rejet du raffinement. Un poids de 0 indique une non-activation et peut-être un manque dans la base de connaissances, qui est interprété dans le contexte d'un monde clos comme un rejet.

2.1.3. IDÉFIX : indexation thématique

IDÉFIX est une sur-couche du réseau *JDM* fondée sur des réseaux de neurones permettant de sélectionner des concepts pertinents pour un texte fourni en entrée. Cette sélection se fait de manière abductive et locale au commentaire, par imitation des exemples déjà appris des interactions précédentes avec l'utilisateur (validation, invalidation et proposition de termes-clés). *IDÉFIX* permet de calculer des termes pertinents à partir d'un texte. Par exemple, à partir du texte qui suit : “*les monnaies locales ont des contreparties en Euro. le principe est de faire circuler la monnaie locale pour faire fonctionner les entreprises locales.*” (Débat sur les monnaies locales - AREN⁵). Nous obtenons les termes suivants :

contreparties • *monnaie* • *entreprises locales* • *économie circulaire* • *échange* •
monnaie alternative • *échange local* • *devise*>*monnaie* • *commerce de proximité* •
monnaie complémentaire

Pour aboutir à son résultat, *IDÉFIX* calcule et combine trois ensembles de termes pondérés.

⁵ <https://portail-aren.lirmm.fr/aren2023/debates/6>

Le premier ensemble, les **termes extraits**, est l'extraction directe des termes du texte avec lemmatisation et identification des termes composés. Dans l'exemple précédent, nous obtenons :

monnaies locales • monnaie locale • avoir de la monnaie • avoir la monnaie de • contrepartie • contreparties • Euro • euro • principe • faire circuler • une locale • faire de • faire fonctionner • les entreprises • entreprises locales • entreprise locale • monnaie • avoir la monnaie • locale • circuler

Ce premier ensemble de termes extraits est l'objet d'une désambiguïisation lexicale via le service *Belléophon*. Les sens sélectionnés pour les termes polysémiques sont ajoutés à cet ensemble, à savoir :

monnaie>unité monétaire ; monnaie>argent ; Euro>monnaie

Le second ensemble, les **termes associés**, est l'ensemble de toutes les associations (relation *r_associated* dans *JDM*) liées positivement aux termes extraits. Dans l'exemple précédent, nous obtenons (le caractère | sépare le terme de sa pondération) :

monnaie | 10 • économie circulaire | 4.486 • produits locaux | 3.686 • traditions locales | 3.447 • culture locale | 2.506 • dialecte local | 2.435 • événements locaux | 2.435 • échange | 2.258 • artisanat local | 2.068 • monnaie alternative | 1.868 • échange local | 1.274 • devise>monnaie | 1.168 • commerce de proximité | 1.099 • histoire locale | 0.974 • monnaie complémentaire | 0.916 • consommation responsable | 0.6 • développement durable | 0.556 • circuit court | 0.534 • sucre>monnaie | 0.484 • échange de biens | 0.432 • billet | 0.423 • échange équitable | 0.378 • monnaies locales complémentaires | 0.273 • monnaie citoyenne | 0.24 • échange de biens et services | 0.232 • Monnaie complémentaire | 0.232 • centime | 0.231 • économie locale | 0.219 • échange de services | 0.216 • monnaie sociale | 0.216

Le troisième ensemble, les **termes activés**, correspond aux activations de sortie dans un réseau neuronal construit dans *JDM* avec comme points d'entrée les termes extraits. Il s'agit d'un réseau neuronal discret (RND) qui associe à un ensemble de termes d'entrée un ensemble de termes de sortie (contrairement au réseau *JDM* qui associe des paires de termes). L'élément de base est un "neurone" qui est un nœud faisant la "glue" entre les deux ensembles (c.f. Figure 1). Ce RND est qualifié de discret car aucun plongement lexical n'est présent. Le fonctionnement du RND est similaire au mécanisme expliqué ci-dessus pour *Belléophon* à quelques différences près :

- Les neurones ne déclenchent que de façon probabiliste en fonction de leur niveau d'excitation et du nombre d'entrées activées (au moins la moitié) ;
- Les nœuds de sorties activés constituent la réponse calculée ;
- Tout nœud de sortie peut être un nœud d'entrée d'un neurone ;
- Le calcul se réalise en une seule passe (aucune itération) ;
- Seules les parties du RND connectées aux entrées activées sont concernées par le calcul, qui est donc strictement localiste malgré la taille conséquente du RND.

Dans l'exemple précédent, nous obtenons après seuillage :

faire circuler • contreparties • monnaie • économie circulaire • entreprises locales • monnaie alternative • économie circulaire et monnaies locales • monnaies locales parallèles • participation citoyenne et monnaies locales • monnaies locales et résilientes • monnaies locales et durables • valorisation des monnaies locales • réseau de monnaies locales • échange non-monétaire et monnaies locales • monnaies locales

et équitables • valorisation des monnaies locales dans le tourisme • stabilité des monnaies locales

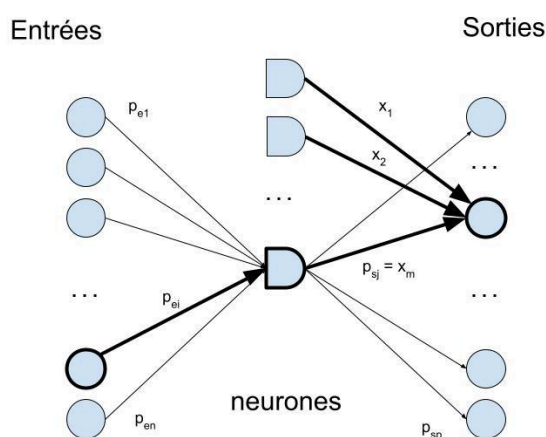


Figure 1. Illustration du réseau neuronal discret (RND) de JDM utilisé par IDÉFIX. Chaque neurone connecte un ensemble de termes + neurones d'entrée à un ensemble de termes + neurones de sortie.

Le RND permet de calculer des termes pertinents par abduction (imitation par l'exemple) non-linéaire à partir de ce qui a déjà été appris (c'est-à-dire des exemples déjà connus qui sont dans le RND). Les termes associés, eux, sont ceux qui sont sémantiquement reliés dans JDM avec un effet d'émergence linéaire.

La combinaison des trois ensembles de termes (extraits, associés, activés), une fois les poids normés, se fait de manière linéaire dans la proportion 1/5, 2/5 et 3/5. Il est tout à fait possible que la liste des termes activés soit vide dans le cas où aucun neurone n'a pu être déclenché. Cela arrive quand des sujets nouveaux sont abordés, et la liste des termes associés fait donc office de filet de sécurité (mais avec parfois des risques de confusion surtout concernant des éléments textuels à la sémantique implicite).

IDÉFIX est le service ainsi sollicité pour calculer les termes et concepts supposément pertinents d'une contribution à un débat faite sur la plateforme AREN. C'est ce résultat qui va être l'objet d'un traitement d'enrichissement.

2.2. Désambiguïsation et augmentation sémantiques

Afin d'assurer une représentativité des propos des utilisateurs, nous procédons à l'enrichissement des ensembles de mots-clés produits à l'étape précédente. En premier lieu, nous cherchons à assurer une couverture sémantique suffisante en nous occupant des éventuels phénomènes d'ambiguïté lexicale⁶ engendrés par la polysémie des termes-clés. Ceci revient à séparer les termes semblables en apparence, mais dont les sens sont différents, en identifiant les raffinements sémantiques adéquats dans le réseau JDM. L'enrichissement des termes de description par leurs termes synonymes ou hyperonymes pertinents, permet, à l'inverse de la désambiguïsation, de regrouper les termes différents en apparence, dont les sens sont (quasi-)semblables. Dans le cas de la restriction aux termes du débat, l'ajout d'un synonyme dans les termes-clés d'un propos ne se fait que si ce synonyme existe déjà comme mot-clé

⁶ Ambiguïté traitée via le service *Bellérophon*.

d'un autre propos du débat, ceci afin d'éviter une dérive liée à des cas de synonymie foisonnante. Dans ce cas, pour l'exemple précédent, les termes *fric* et *thune* ne seront pas ajoutés aux termes d'indexation comme synonymes du terme *monnaie*, vu qu'ils ne sont pas utilisés pour indexer d'autres propos du même débat.

L'intérêt de cet enrichissement est d'identifier, indépendamment des diverses variantes lexicales, les occurrences d'un sens donné, ceci afin d'éviter un éparpillement des éléments de sens (relations sémantiques) pris en compte par l'algorithme, pour un terme donné.

2.3. Extraction de connaissances

L'extraction de connaissances à partir de l'indexation des commentaires utilise l'AFC, un cadre mathématique basé sur la théorie des treillis, théorie permettant la représentation de l'information contenue dans des données sous forme algébrique ou logique (Ganter et Wille, 1999). L'AFC part de données sous la forme d'un contexte formel, c'est-à-dire d'un triplet (O, A, R) où $R \subseteq O \times A$ est une relation binaire entre des objets O et les attributs A qui les décrivent. Dans AREN, les objets sont les commentaires du débat et les attributs sont les mots-clés (ou termes) proposés par le débattant ou ajoutés lors de la phase d'indexation. Un terme est en relation avec un commentaire s'il l'indexe. Un contexte formel donne lieu à deux opérateurs de dérivation, tous deux notés \cdot' et définis tels que :

$$\cdot' : 2^A \rightarrow 2^O$$

$$\text{Pour } I \subseteq A, I' = \{o \in O \mid \forall a \in I, (o, a) \in R\}$$

$$\cdot' : 2^O \rightarrow 2^A$$

$$\text{Pour } E \subseteq O, E' = \{a \in A \mid \forall o \in E, (o, a) \in R\}$$

Les compositions \cdot'' de ces opérateurs forment des opérateurs de fermeture. Par exemple, dans le Tableau 1, la fermeture d'*échange local* est {échange local, crise financière}.

Nous cherchons à extraire des régularités dans la co-occurrence des mots-clés dans l'indexation des commentaires. L'AFC offre différentes possibilités de représentation de ces régularités : implications, règles d'association, treillis de concepts ou relations causales (Bazin et al., 2022). Nous avons choisi de commencer par explorer l'utilisation des implications, règles de la forme $A \rightarrow B$ où A et B sont des ensembles de termes. Une telle règle est valide si et seulement si tous les commentaires indexés par les termes de A sont aussi indexés par les termes de B . Ainsi, dans l'exemple de la Table 1, les implications {monnaie alternative} \rightarrow {devise>monnaie} et {échange local} \rightarrow {crise financière} sont valides.

	monnaie alternative	échange local	devise>monnaie	moyen d'action	crise financière
c_1	×		×		
c_2			×	×	×
c_3		×			×
c_4		×	×	×	×

Tableau 1. Exemple de relation binaire entre quatre commentaires (c_i) et cinq termes

Afin de réduire le nombre de règles à présenter aux débattants, notre attention se focalise spécifiquement sur les règles réduites et exactes. En d'autres termes, nous considérons seulement les implications de la forme $\{a\} \rightarrow B$ telles que a est un terme dit irréductible, ou pour le dire de manière simplifiée qui n'est pas équivalent à une intersection d'autres termes. Plus précisément, l'ensemble $\{a\}'$ des objets qu'il décrit n'est pas égal à l'intersection des ensembles d'objets décrits par d'autres attributs (Liquière, 2021), c'est-à-dire qu'il n'existe pas d'ensemble d'attributs X tel que $\{a\}' = \cap \{x\}'$, $x \in X$. Dans l'exemple précédent, *moyen d'action* est équivalent à l'intersection de *devise>monnaie* et *crise financière*.

2.4. Enrichissement de la base de connaissances

Les implications obtenues avec l'AFC sont utilisées pour mettre à jour les relations dans la base de connaissances exploitée lors de ce processus. Depuis une implication de la forme $\{a\} \rightarrow \{b, c, d, \dots\}$ nous ajoutons donc dans la base de connaissances des relations $a \rightarrow x$ avec $x \in \{b, c, d, \dots\}$. Dans l'exemple précédent, la mise à jour de la base de connaissances *JDM* se fait par l'ajout du terme *devise>monnaie* associé à *monnaie alternative* et du terme *crise financière* associé à *échange local*. Ces modifications améliorent globalement la composante associative des calculs ultérieurs des indexations des propos.

3. Etude expérimentale

3.1 Mesures d'évaluation des règles

Dans le but d'étudier l'impact de l'augmentation sémantique sur la qualité des règles, nous avons calculé le support et la nouveauté des règles.

3.1.1. Support

Le support d'un ensemble d'attributs ou termes T est le nombre d'objets (ou de commentaires) décrits par T divisé par le nombre total d'objets. Il peut être défini par : $Supp(r) = p(T^p_r, T^c_r) / C$; où T^p_r et T^c_r sont respectivement les termes de la prémisse et de la conclusion de la règle r et C sont les commentaires.

3.1.2. Nouveauté

Une règle est considérée nouvelle si sa prémisse et sa conclusion ne sont pas statistiquement indépendantes (Lavrac et al., 1999). La nouveauté d'une règle est définie par : $Nov(r) = p(T^p_r T^c_r) - p(T^p_r) p(T^c_r)$; où r est une règle, T^p_r et T^c_r sont respectivement les termes de la prémisse et de la conclusion de la règle r .

3.2 Jeux de données et configurations

Nous procédons à l'évaluation de notre approche à l'aide des données issues d'un débat sur la plateforme AREN concernant les monnaies locales⁷ qui a eu lieu de mars 2020 à mai 2023 entre 8 débattants ayant partagé 48 arguments. Nous comparons les résultats de trois variantes

⁷ <https://portail-aren.lirmm.fr/aren2023/debates/6>

de notre approche KT (KeepTalk) pour mesurer l'effet de l'augmentation sémantique sur la qualité des résultats de l'AFC. Les détails de nos méthodes sont énumérés ci-dessous :

- KT : Les implications sont calculées à partir du contexte d'extraction initial, défini par la relation binaire entre les reformulations des débattants et les termes-clés qui indexent leurs sélections, reformulations et argumentations.
- KT^\dagger : Le contexte d'extraction est enrichi par les synonymes des termes qui définissent les attributs pour générer les implications. On s'intéresse aux termes synonymes qui sont déjà utilisés lors de l'indexation initiale (R^\bullet) et aussi à ceux qui ne le sont pas (R°), donc, à de nouveaux termes n'apparaissant pas dans le débat.
- KT^\ddagger : Identique à la configuration précédente avec des hyperonymes au lieu de synonymes.

3.3 Résultats

Nous rapportons dans le Tableau 2, le nombre d'attributs, irréductibles et implications dans les configurations se limitant, ou pas, aux termes-clés du débat.

Configuration	KT	KT^\dagger		KT^\ddagger	
		DL°	DL^\bullet	DL°	DL^\bullet
Avec restriction aux termes-clés du débat (R^\bullet)					
# Attributs	413	413	413	413	413
# Irréductibles	52	54	50	67	64
# Implications	25	37	29	57	55
Avg Support	0.05478	0.08754	0.05697	0.12128	0.11818
Avg Nouveauté	0.04858	0.06795	0.05004	0.06746	0.06510
Sans restriction aux termes-clés du débat (R°)					
# Attributs	413	2448	1144	772	619
# Irréductibles	52	86	47	73	59
# Implications	25	56	25	54	46
Avg Support	0.05478	0.07414	0.063478	0.15539	0.12051
Avg Nouveauté	0.04858	0.061824	0.053534	0.07834	0.06318

Tableau 2. Comparaison des résultats de l'analyse formelle de concepts : KT avec le contexte initial ; KT^\dagger avec le contexte augmenté avec les synonymes ; KT^\ddagger avec le contexte augmenté avec les hyperonymes. L'augmentation est faite avec et sans désambiguïsation lexicale (DL^\bullet / DL°)

Nous observons une augmentation du nombre d'attributs dans le cas où l'ajout de termes n'est pas restreint à ceux du débat, ce qui est conforme à l'intuition. Dans tous les cas, l'utilisation de la désambiguïsation lexicale réduit le nombre d'irréductibles et d'implications. Nous constatons que la nouveauté est globalement très faible, ce qui indique que l'on trouve peu d'associations n'existant pas dans la base de connaissances. Ceci est positif du point de vue de la complétude de la base. Notre approche parvient de façon très satisfaisante à extraire des signaux faibles des débats. On constate par ailleurs que l'enrichissement sémantique a

amélioré, dans tous les cas, les résultats de l’AFC. La configuration la plus favorable est celle avec une augmentation avec synonymes ou hyperonymes, et sans l’étape de la DL. La pire configuration, qui a les plus mauvais résultats, est l’approche sans enrichissement du contexte initial d’extraction.

4. Conclusion et perspectives

La procédure de description thématique s’inscrit dans une démarche collaborative, itérative et incrémentale. Les ensembles de termes indexant chaque commentaire sont co-construits d’une part, par la procédure automatisée *IDÉFIX*, et de l’autre, par une supervision et complétion par les utilisateurs des termes extraits par *IDÉFIX*. Cette supervision est permise en donnant à l’utilisateur la possibilité de proposer, valider ou invalider des termes de l’ensemble proposé par l’IA accompagnant le débat. Ce retour est pris en compte lors des itérations de descriptions thématiques ultérieures, menant à une indexation de meilleure qualité, ce qui a été vérifié dans les faits. L’approche coopérative a été pensée de manière à inciter l’utilisateur à proposer des termes-clés complétant ses propos. Ce fonctionnement pourra servir d’outil de relance ou d’incitation à l’élaboration de points de vue, sous une forme autre que celle du commentaire textuel brut.

En perspective, il serait pertinent d’envisager l’intégration d’une IA générative à des fins d’animation ou de synthèse du débat. Cet agent produirait de nouveaux propos au débat à partir du contenu structuré issu des procédures d’indexation des commentaires des utilisateurs. Concernant l’extraction des connaissances, en perspective, le projet explorera d’autres représentations des régularités : autres implications, règles d’association et relations causales ; ce qui permettra d’ajouter dans la base de connaissances des informations sur des types de relations plus précises (autres que les associations d’idées).

Bibliographie

- Bächtold M., Pallarès G., De Checchi K. et Munier V. (2023). Combining debates and reflective activities to develop students’ argumentation on socio-scientific issues. *Journal of Research in Science Teaching* 60 (4), 761–806.
- Bazin A., Couceiro M., Devignes M.-D. et Napoli A. (2022). Steps towards causal formal concept analysis. *International Journal of Approximate Reasoning* 142, 338–348.
- Ganter B. et Wille R. (1999). Formal concept analysis: mathematical foundations. *Springer. Springer 1999, ISBN 978-3-540-62771-5*, pp. I-X, 1-284.
- Lafourcade M. et Le Brun N. (2023). Apport du jeu pour la construction de connaissances : le projet JeuxDeMots. *Technologie et innovation* 8 (*Le jeu pour innover*).
- Lavrac, N., Flach, P. A., and Zupan, B. (1999). Rule evaluation measures : A unifying view. *International Workshop on Inductive Logic Programming, ILP ’99*, page 174–185, Berlin, Heidelberg. Springer-Verlag.
- Liquière, M. (2021). Utilisation des irréductibles d’un treillis de concepts pour la sélection de motifs. Technical report, LIRMM, Université de Montpellier.