

HAI508I - Projet Long Intégrateur CMI

Recherche Bibliographique

De par l'analyse des échanges entre acteurs au sein d'une population, peut-on grâce à des méthodes d'intelligence artificielle déduire des marqueurs de l'interdisciplinarité ?

BOURRET MAXIME

COSSU ARNAUD

HADDAD GATIEN

SAID ADAM

Encadrés par : LAURENT Anne



Table des matières

1	L’interdisciplinarité	1
2	Les mails, une mine d’or pour l’information	2
2.1	Les mails au cours du temps	2
2.2	L’email mining	3
3	IA et langage naturel	4
4	Conclusion	6
5	Annexes	7
5.1	Citations	7
5.2	Références	7
5.3	Sources	7

1 L'interdisciplinarité

« La discipline est une catégorie organisationnelle au sein de la connaissance scientifique ; elle y institue la division et la spécialisation du travail et elle répond à la diversité des domaines que recouvrent les sciences » [1] c'est ce qu'a dit Edgar Morin dans le Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires (CIRET) de juin 1994. Le Directeur émérite du CNRS explique qu'il existe différents niveaux de relations et d'entrelacements entre les disciplines et que chacun possède une définition et une utilisation bien précise. On peut citer par exemple : la transdisciplinarité, l'interdisciplinarité, l'intradisciplinarité ou encore la pluridisciplinarité. Mais comment faire pour déceler de ces préfixes les différences parfois infimes permettant d'identifier les travaux scientifiques ?

Par exemple, quand on parle d'interdisciplinarité il s'agit de « la mise en relation d'au moins deux disciplines, en vue d'élaborer une représentation originale d'une notion, d'une situation, d'une problématique » [2] (Maingain, Dufour et Fourez, 2002) ou selon Nourrit, Aleveque, Laurent et Libourel dans JIMS n°11 (À paraître) [5] l'interdisciplinarité « C'est une stratégie d'interpellation, de confrontation et de conflit qui se résout lorsque des transferts de méthodes, de concepts et d'outils s'opèrent. » . Quant à la transdisciplinarité, il s'agirait plutôt de la « Mise en œuvre d'une méthode fondamentale commune à toutes les disciplines » (Angers et Bouchard, 1992, p. 69) mais selon l'Institut Canadien de Recherches Avancées (ICRA), [3] « le principe de la transdisciplinarité consiste en un refus d'aborder le monde et ses problèmes par les catégories que sont les disciplines. À partir des problèmes du monde réel, la transdisciplinarité tente de construire ses propres contenus et ses propres méthodes. Elle offre une nouvelle vision de la réalité, qui émerge de la confrontation des disciplines ». On constate donc que la définition de ces termes n'est pas universelle et que le sens peut drastiquement varier d'une personne à l'autre.

Il y a également un risque « d'hyperspécialisation » et de « chosification » de l'objet que l'on étudie dont on pourrait oublier le fait qu'il provient de la construction de disciplines et de liaisons qui seront alors négligées comme l'explique Morin. Il est donc primordial de s'accorder sur un corpus de définitions permettant de clarifier et de montrer les limites entre ses relations disciplinaires, c'est ce que tente d'exposer Nourrit, Aleveque, Laurent et Libourel [5] : « La principale difficulté est que le terme interdisciplinarité chez de nombreux chercheurs recouvre souvent les autres. Comme le note Klein, par exemple, « l'interdisciplinarité encyclopédique » chez Boden a la même définition que la multidisciplinarité. » JIMS n°11 (À paraître) .

Dans notre étude où nous allons étudier l'interdisciplinarité, cette dernière sera le résultat de la démarche par laquelle deux disciplines vont croiser leurs compétences et leurs savoir-faire pour permettre de comprendre une notion, de l'apprendre ou construire un apprentissage.

2 Les mails, une mine d'or pour l'information

2.1 Les mails au cours du temps

L'utilisation et la fréquence d'envoi des mails ont beaucoup évolué durant les dernières décennies. Étant à la base un simple projet liant 4 ordinateurs entre eux, les mails devinrent une innovation majeure du XXe siècle. Le premier E-mail fut envoyé en 1971, par Ray Tomlinson créant ainsi ARPANET. Bien évidemment, afin d'organiser notre boîte aux lettres virtuelle, un manager d'E-mail fut créé par Lawrence Roberts, seulement un an plus tard. Mais l'utilisation des mails ne restera pas longtemps si ludique et utile et sera vite entachée par les mails marketing. Le premier fut envoyé en 1978 et fut reconnu comme le premier SPAM. Ce mail fut envoyé à 400 personnes afin de promouvoir un nouveau produit, ce qui rapporta 13 millions (52.6M de dollars actuels) à son expéditeur comme l'explique l'article de Radhika Chugh "The Evolution of Email : A complete timeline" [4]. Avec de tels résultats, ce mouvement s'est vite démocratisé, jusqu'à le rendre banal. Vers les années 90, beaucoup de services de messagerie gratuits furent lancés comme Hotmail ou Internet Mail, ce qui eut pour conséquence d'augmenter considérablement le nombre d'utilisateurs, en rendant la messagerie électronique plus accessible.

Aujourd'hui, les mails sont très présents et depuis leur création, il y a eu énormément d'avancées et d'évolutions, comme la possibilité d'ajouter des pièces jointes, l'évolution de la taille acceptée de ces dernières. . . Actuellement, des robots, ou des algorithmes bien définis permettent d'affiner le triage de mails, nous proposent des formulations lors de l'écriture et peuvent même répondre à notre place. Nous passons donc énormément de temps à trier nos e-mails, à les lire et à les écrire. Les sujets sont divers, nous pouvons parler entre membres d'une même famille et avoir des discussions de l'ordre du privé, nous pouvons aussi utiliser ce moyen de communication pour notre travail, afin de rester en contact avec une équipe, prévoir des rendez-vous, organiser des réunions. De ce fait, beaucoup d'informations publiques et privées sont contenues dans ces mails, et il serait donc possible d'établir des liens d'affinités entre plusieurs personnes. Un outil pourrait par exemple nous suggérer des contacts de notre équipe lorsque nous sommes au travail, et inversement nous fournir le mail d'un de nos proches quand nous sommes chez nous. Il serait aussi possible de visualiser les centres d'intérêts de certains utilisateurs. Pour cela, il est important d'analyser cette source de données que sont les mails, cette dernière étant très importante il est préférable d'utiliser des systèmes d'automatisation ainsi que des intelligences artificielles afin d'extraire et de proposer des données tout d'abord triées, "nettoyées" et si possible analysées.

2.2 L'email mining

L'email mining est le processus d'étude et d'analyse d'un grand nombre de courriels, souvent voire obligatoirement aidé d'un logiciel qui peut repérer et découvrir des informations, des modèles, des mots clés et beaucoup d'autres éléments utiles dans les données de courriel. Il y a différentes étapes dans l'email mining, il faut extraire les données, les représenter, puis les catégoriser/analyser. Il existe énormément d'approches et de méthodes afin de traiter le contenu de ces mails. L'avantage de l'email mining sont les suivants : Analyser un grand nombre de données envoyées par des clients comme par exemple des commentaires, des retours . . . Cela peut aussi servir à créer des listes d'email marketing. L'email mining est aussi utilisé pour l'amélioration des stratégies marketing de certaines entreprises. Il est aussi possible d'utiliser ces données afin d'améliorer la détection de spams, la catégorisation des emails, et la mise en avant de mails importants. Beaucoup de logiciels d'Email mining existent déjà et sont disponibles comme DiscoveryOne, Lexalytics, Medallia ou encore TowerData. Comme dit plus haut, l'extraction et la classification de ces données se fait généralement grâce à une intelligence artificielle, qui peut trier les mails et leur affecter des labels. Les techniques associées au data mining général ne fonctionnent pas forcément sur les mails, car ces derniers ont une structure différente. Ils peuvent contenir du langage familier, voire grossier, ils sont généralement assez courts, et contiennent parfois seulement des images comme démontré dans l'article "Email mining : Tasks, common techniques, and tools" de Guanting Tang [6]. Une des méthodes de visualisation et l'approche de la structure sociale, qui représente les interactions par un graphe composé de nœuds. Dans cette représentation il faut définir un nombre de mails cible. Une fois ce nombre dépassé, cela créera un lien entre deux adresses mails. Il est par la suite possible d'ajouter des données supplémentaires comme la force de la relation ou les sujets récurrents. Pour la classification, une des méthodes est le classifieur avec règles de base. Avec une structure utilisant des conditions, il permet de trier de la manière suivante : Si le mail contient un tel mot, alors le catégoriser dans une certaine classe. Avec cette méthode, nous avons un taux de 8% de mauvaise classification.

3 IA et langage naturel

La manière dont les IA peuvent communiquer avec un système intelligent, qu’il soit humain ou également une IA, s’appelle le traitement du langage naturel, autrement appelé traitement automatique du langage, ou encore NLP (de l’anglais, Natural Language Processing). Il est à la base de toute communication et est obligatoire si l’on veut qu’une machine suive des instructions à partir d’une conversation en langage naturel, par exemple dans le domaine médical, dans l’armée, etc. . . Le NLP est divisé en deux parties : la première est la génération du langage naturel, elle consiste à construire et produire des phrases ayant un sens dans leur langue d’origine. Cela inclut construire un plan de la phrase, savoir l’adapter au contexte et choisir les mots appropriés pour exprimer une idée concrète. La seconde composante est la compréhension du langage naturel, qui nécessite l’analyse des différentes parties du langage pour en tirer une représentation exploitable par l’IA.

Au premier abord, cela peut sembler contre-intuitif si l’on compare avec notre apprentissage des langues, mais la compréhension du langage naturel est plus difficile pour une machine que la génération du langage naturel. Cette difficulté est notamment due aux ambiguïtés des phrases qui peuvent être fournies en entrée, pour les mots ayant une fonction différente selon le contexte (ex : “allée”, qui peut être verbe ou un nom), ou encore les références au sein des phrases ; par exemple, avec la phrase “La caisse ne rentre pas dans la voiture car elle est trop grande”. Ici, même cela peut nous sembler évident, une IA pourrait ne pas arriver à déterminer le sujet (la caisse ou la voiture) auquel “elle est trop grande” fait référence. De plus, selon le contexte, une phrase peut signifier plusieurs choses, autant que différentes phrases peuvent avoir la même signification.

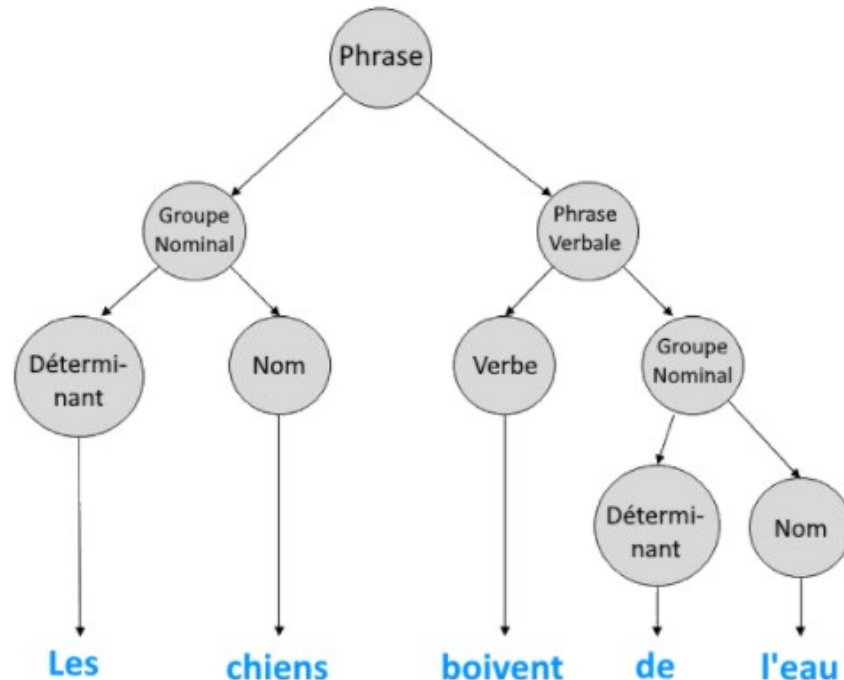
Pour arriver à créer une IA capable de NLP, il faut donc l’entraîner en suivant cinq grandes étapes :

Pour commencer, l’analyse du champ lexical. Cela requiert de savoir identifier et analyser la structure des mots, en divisant les paragraphes en phrases, et les phrases en plusieurs mots isolés. Dans un second temps, l’analyse syntaxique, qui consiste à analyser la grammaire d’une phrase, et réarranger ses mots pour y faire apparaître la relation entre les mots. Vient ensuite l’analyse sémantique, qui va venir essayer de construire des structures syntaxiques. Une structure syntaxique peut être représentée par un graphe dans lequel les mots sont ordonnés par rapport à leur sens, et plus ou moins connectés selon leur ressemblance entre eux. Il y a donc certaines phrases que l’analyseur syntaxique ne comprend pas et ne tient pas compte, comme les oxymores ou des phrases telles que “Une pluie sèche”. Après ça, c’est au tour de la mise en contexte (ou “intégration du discours”), nous savons qu’une phrase peut perdre son sens si elle est isolée de son contexte, c’est pour cela que l’étape de la mise en contexte récupère le sens des phrases précédentes pour déduire celui de la phrase actuelle ; et par induction, les phrases d’après. Enfin, la dernière phase est celle de l’analyse pragmatique. Cette étape consiste à reprendre les mots qui ont été entrés, et les ré-interpréter à l’aide du sens que nous leur avons donné précédemment. Cela requiert l’utilisation de synonymes, ou en d’autres termes, une connaissance des aspects du langage dans la vie courante. On peut par exemple transformer la phrase “Tom nous a quitté suite à un arrêt du cœur spontané” en “Tom est mort d’une crise cardiaque”.

Ces cinq étapes permettent donc à une Intelligence Artificielle de pouvoir comprendre notre langage. Il existe plusieurs types d’algorithmes qui suivent ce procédé, tels que des algorithmes de type Approche Descendante (ou Top-Down Parser en Anglais) ou de type Context-Free Grammar. Pour ce dernier, ce type d’algorithme analyse l’entrée et

construit un arbre d'analyse de la phrase en la coupant progressivement en les feuilles. Par exemple, la phrase “Les chiens boivent de l’eau” se décompose comme suit :

Résumé



La phrase "Les chiens boivent de l'eau" décomposée grâce à l'arbre d'analyse

L'arbre d'analyse décompose donc la phrase en parties structurées afin que l'IA qui va tenter de la lire puisse facilement la comprendre et la traiter. Pour que l'algorithme puisse construire cet arbre d'analyse, un ensemble de règles de réécriture doit être construit, décrivant quelles structures d'arbre sont légitimes ; pour notamment éviter que la phrase soit mélangée en “De l’eau boivent les chiens”. Ce type d'algorithme étant un des plus simples à implémenter, il est évidemment le plus largement utilisé dans de nombreux domaines nécessitant du NLP. La mails ayant une structure plutôt simple, le plus souvent écrits sans faute d'orthographe et assez courts, il est donc évident que ce type d'algorithmes pourrait aider à analyser un grand nombre de mails pour essayer d'en déduire le sens.

4 Conclusion

De par l'analyse de documents sur les sujets cités précédemment nous avons pu voir que l'interdisciplinarité permet une meilleure compréhension du sujet au sein d'un groupe de recherche et permet un meilleur avancement grâce à une communication de meilleure qualité. Elle est l'une des meilleures approches à adopter lors d'un projet collaboratif afin d'arriver au mieux à ses objectifs dans une équipe comportant des membres de domaines différents. Au sein d'une équipe de travail, on peut voir différents moyens de communication, dont les mails qui sont un des plus utilisés et donc là où l'on peut en apprendre le plus sur les faits et pensées des membres du groupe. Les mails représentent une grosse part de la communication dématérialisée. Ils s'étendent dans plusieurs domaines, leur contenu varie selon l'expéditeur et le destinataire, ils peuvent être à caractère publicitaire mais aussi personnels et professionnels. Depuis leur création, les mails sont un moyen pratique d'échange qui véhicule beaucoup d'informations qui peuvent être extraites de par leur analyse. L'email mining permet de créer les liens entre les utilisateurs, mieux comprendre les critères qui composent un "bon mail" et donc les réutiliser par la suite. Des compagnies commerciales peuvent par la suite se servir de ces critères pour optimiser leur mails de promotion, mais cette étude peut aussi favoriser l'analyse des comportements des personnes au sein d'une population. Cette dernière peut se faire à la main, mais serait fastidieuse, longue et peu fructueuse. Néanmoins il serait possible d'automatiser le tout à l'aide de programmes créés spécialement pour étudier les mails et leur contenu. Les progrès de l'IA permettent une meilleure approche du langage naturel par les machines. Bien que compliquée et pas encore optimale, la compréhension de notre langue par un ordinateur est possible à l'aide d'intelligences artificielles entraînées est possible. Grâce à ces avancées on peut extraire le sens d'une phrase automatiquement, en analysant le sens de cette dernière sur plusieurs points (syntaxiques, grammaticaux...) mais aussi en étudiant les phrases qui la précède afin d'optimiser la compréhension du sens exact. Pour conclure il serait possible de déduire les marqueurs de l'interdisciplinarité de par l'analyse des comportements au sein d'une population à l'aide d'une IA spécialisée dans l'analyse des échanges des membres de cette dernière. En récupérant un registre suffisamment grand de mails échangés au sein d'une équipe de travail, il serait possible d'utiliser une intelligence artificielle entraînée à comprendre le langage naturel afin d'analyser les communications du groupe. De par cette étude, on pourrait repérer les schémas répétitifs, les mots et phrases récurrents, tisser des liens entre les utilisateurs afin de faire ressortir les marqueurs de l'interdisciplinarité qui peuvent exister au sein de cette population.

5 Annexes

5.1 Citations

- [1] Edgar MORIN. *Sur l'interdisciplinarité. [Dans RENCONTRES TRANSDISCIPLINAIRES]*. 1994. URL : <https://ciret-transdisciplinarity.org/bulletin/b2c2.php>.
- [2] Maingain et Dufour FOUREZ. *Approches didactiques de l'interdisciplinarité*. 2003. DOI : <https://doi.org/10.7202/009499ar>. URL : <https://id.erudit.org/iderudit/009499ar>.
- [3] Laboratoire interdisciplinaire littérature et MATHÉMATIQUES. *Inter, trans, multi, pluri ou intradisciplinarité ?* URL : <https://www.usherbrooke.ca/litt-et-maths/fondements/inter-trans-multi-pluri-ou-intradisciplinarite>.
- [4] Radhika CHUGH. *he Evolution of Email : A complete timeline*. 30 juillet 2019. URL : <https://www.linkedin.com/pulse/evolution-email-complete-timeline-radhika-chugh/>.
- [5] Laurent et Libourel NOURRIT Aleveque. *JIMS*. A paraître.
- [6] Guanting TANG. *Email mining : Tasks, common techniques, and tools*. octobre 2013. URL : https://www.researchgate.net/publication/257482113_Email%5C_mining%5C_Tasks%5C_common%5C_techniques%5C_and%5C_tools.

5.2 Références

- Stephanie Overby "Artificial intelligence (AI) vs. natural language processing (NLP) : What are the differences ?", Février 2020 enterpriseproject.com/artificial-intelligence-ai-vs-natural-language-processing
- Priya Dialani "ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING TRANSFORM CHATBOTS" Août 2020 analyticsinsight.net
- Tutorials Point "Compiler Design - Top-Down Parser" tutorialspoint.com/compiler_design

5.3 Sources

Logo Faculté des Sciences : [En ligne] [Grand \(2392*2950\).png](https://sciences.edu.umontpellier.fr)
sciences.edu.umontpellier.fr

Logo Université Montpellier : [En ligne] [LOGO-original-RVB-WEB-1.png](https://umontpellier.fr)
umontpellier.fr

Schéma syntaxique d'une phrase : par Gatien Haddad p.5