



GRCE

SIFED

2024

# Symposium International Francophone sur l'Écrit et le Document

sponsorisé par :



organisé durant les journées scientifiques de Nantes Université



Colloque N°3 - Niveau 3 - Salle KL  
La Cité des Congrès, 5 Rue de Valmy, 44000 Nantes

# Programme

**8h30 : Accueil & café**

**8h50-9h : Mot d'accueil**

**9h-10h : Session Orale #1 : Analyse de l'écriture**

Orale 1 : *Xie Yejing, Nantes Université, LS2N*

↔ Stroke-Level Graph Labeling with Edge-weighted Graph Attention Network for Handwritten Mathematical Expression Recognition

Orale 2 : *Mowlavi Seginus, ENS Paris-Saclay*

↔ Reconnaissance stochastique de caractères par exploitation de la redondance des formes

Orale 3 : *Imbert Florent, Université de Rennes, INSA Rennes, IRISA*

↔ Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors

**10h-10h15 : Assemblée Générale du GRCE**

**10h15-10h30 : Pause**

**10h30-11h30 : Session Poster**

↔ Rendez-vous à la salle des posters à proximité des sponsors de cette évènement : LUMINESS, TEKLIA, YOOZ

**11h30-12h30 : Session Orale #2 : Documents Anciens**

Orale 4 : *Constum Thomas, Université de Rouen, LITIS*

↔ Extraction d'informations de bout en bout dans des documents manuscrits : vers une compréhension des actes de mariage de Paris de 1880 à 1940

Orale 5 : *Giovannangeli Vincent, CNRS, SPHere, PHILIUMM*

↔ L'édition des brouillons de Gottfried Wilhelm Leibniz - Approche numérique

Orale 6 : *Girdhar Nancy, La Rochelle Université, L3i*

↔ STRAS: A Semantic Textual-Cues Leveraged Rule-Based Approach for Article Separation in Historical Newspapers

**12h30-14h : Déjeuner**

**14h-15h : Session Invitée**

*Dufour Richard, Nantes Université, LS2N / LIA*

↔ Grands modèles de langue pour le traitement de documents textuels dans domaine médical

**15h-16h : Session Orale #3 : Extraction d'information**

Orale 7 *Tarride Solène, TEKLIA*

↔ Extraction d'informations structurées dans des documents historiques

Orale 8 *Rayar Frédéric, Université de Tours, LIFAT*

↔ Extraction automatique de données à partir des formulaires médicaux

Orale 9 *Delaunay Julien, La Rochelle Université, L3i*

↔ CoastTerm: a corpus for Multidisciplinary Term Extraction in Coastal Scientific Literature

**16h-16h30 : Prix + clôture**

## Sessions Orales :

### Orale 1

**Titre :** Stroke-Level Graph Labeling with Edge-weighted Graph Attention Network for Handwritten Mathematical Expression Recognition

**Auteur(s) :** Yejing Xie<sup>1</sup>, Harold Mouchère<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> Nantes Université, LS2N

**Résumé :** Handwritten Mathematical Expression Recognition (HMER) algorithms with deep learning approaches have developed rapidly in recent years, most algorithms are dependent on heavy pre-training and also complex network structures. Existing architectures are build on encoder-decoder from on-line or off-line inputs to produce  $\text{\LaTeX}$  markup strings, or stroke-level graphs to generate symbol-level graphs. They all remain on a latent space, which is not directly related to the input data: the strokes. Using the Stroke Label Graph Modelization allows a direct connection between the input data and the output labels. In this research, we proposed a novel stroke-level graph labeling method with edge-weighted graph attention network (EGAT). This lightweight model doesn't rely on any pre-training, abandons the laborious process of encoder-decoder, is totally end-to-end, directly accomplishes stroke-to-stroke feature extraction, and produces strokes and relations classification. Experiments show that our proposed EGAT algorithm can effectively fuse the node features as well as the weighted edge features, and predict the node and edge attributes simultaneously.

## Orale 2

**Titre :** Reconnaissance stochastique de caractères par exploitation de la redondance des formes

**Auteur(s) :** Seginus Mowlavi<sup>1</sup>, Diego Belzarena<sup>2</sup>, Natalia Bottaioli<sup>1</sup>, Antoine Tadros<sup>1</sup>, Marina Gardella<sup>3</sup>, Gabriele Facciolo<sup>1</sup>, Gregory Randall<sup>2</sup>, Jean-Michel Morel<sup>4</sup>

**Affiliation(s) :**

<sup>1</sup> ENS Paris-Saclay

<sup>2</sup> Universidad de la República, Uruguay

<sup>3</sup> IMPA, Rio de Janeiro, Brazil

<sup>4</sup> City University of Hong Kong

**Résumé :** La transformation numérique qui s’amorce dans les archives et les bibliothèques de part le monde se traduit par une numérisation massive de documents, sur laquelle se développe de plus en plus la reconnaissance automatique de texte (ATR) afin de convertir les pages scannées en texte numérique, quels qu’en soient le contenu ou la période. Dans cet environnement, l’ATR se justifie déjà par des cas d’usage ne nécessitant pas forcément des taux d’erreurs très faibles, tels que la recherche par mots clés dans les bases documentaires. Cependant, d’autres applications restent hors de portée, comme la transcription intégrale du texte, sans parler de vectorisation complète des documents ou de production de fac-similés. En effet, la technologie actuelle est encore loin d’atteindre des taux de performance humains et ainsi de répondre aux critères drastiques de fiabilité et de généralisabilité imposés par les bibliothèques et archives. En contrepois à ces exigences élevées, les documents considérés présentent la caractéristique d’une redondance exceptionnelle des formes que sont les caractères, notamment grâce à un volume et une densité moyenne de texte considérables. Les particularités de ce contexte suggèrent la nécessité de dépasser les techniques à base de réseaux de neurones qui, par leur capacité à apprendre des descripteurs flous adaptés à des formes variées mais de domaine bien délimité, fournissent une base minimale de performance sans exploiter la précision des formes ni garantir leurs taux d’erreurs à tous les domaines. Nous proposons ainsi une méthode de post-traitement dont le principe est, partant d’un résultat initial d’ATR avec un taux d’erreur typique des systèmes neuronaux actuels, d’établir des modèles stochastiques de chaque caractère selon sa police et sa fonte, puis d’intégrer ces modèles à un processus itératif de reconnaissance bayésienne. Les promesses de cette méthode sans apprentissage profond, qui ne fait intrinsèquement intervenir aucune modélisation de langage, sont d’une part l’interprétabilité directe des résultats fournis, d’autre part la généralisabilité avec la capacité à s’adapter en “zero-shot” à des nouveaux domaines de documents ou corpus. Il s’agit d’un travail en cours, dont nous présenterons les résultats préliminaires. Nous présenterons notamment l’influence sur ces résultats de la modélisation stochastique, avec un aperçu des techniques de traitement d’image nécessaires à l’établissement de ces modèles.

## Orale 3

**Titre :** Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors

**Auteur(s) :** Florent Imbert<sup>1</sup>, Eric Anquetil<sup>1</sup>, Yann Soullard<sup>2</sup>, Romain Tavenard<sup>2</sup>

**Affiliation(s) :**

<sup>1</sup> IRISA, Université de Rennes, INSA Rennes, Rennes, France

<sup>2</sup> IRISA, Université Rennes 2, Rennes, France

**Résumé :** The use of digital pens for online handwriting trajectory reconstruction is a prevalent method for human-computer interaction. In this study, we focus on a digital pen equipped with sensors where we aim at reconstructing the online handwriting trajectory. This pen enables writing on any surface and preserving the digital trace of handwriting. This type of pen could be used as an aid to learning to write in classroom. In this paper, we propose a new approach learning to finely reconstruct the touching trajectories while precisely analyzing the hovering part in order to position the next touching trace correctly. This relies on a Mixture-Of-Experts (MOE) approach. The first expert is dedicated for the pencil touch, and is named touching expert model. The second one is dedicated for the hovering pen trajectory, and is named hovering expert model. We improve on the learning of each of these experts based on additional context or specific examples. In addition we introduce a novel public benchmark dataset, to enable future research and comparisons in the field of handwriting reconstruction. The results demonstrates a significant enhancement compared to its primary competitors

## Orale 4

**Titre :** Extraction d'informations de bout en bout dans des documents manuscrits : vers une compréhension des actes de mariage de Paris de 1880 à 1940

**Auteur(s) :** Thomas Constum<sup>1</sup>, Lucas Preel<sup>1</sup>, Théo Larcher<sup>2</sup>, Thierry Paquet<sup>1</sup>, Pierrick Tranquez<sup>1</sup>, Sandra Bree<sup>3</sup>

**Affiliation(s) :**

<sup>1</sup> LITIS, Université de Rouen

<sup>2</sup> LIRMM, Université de Montpellier

<sup>3</sup> LARHRA, CNRS

**Résumé :** Nous présentons une architecture entièrement de bout en bout conçue pour effectuer simultanément la reconnaissance de texte manuscrit et l'extraction d'informations sur des images de pages, éliminant le besoin d'une segmentation explicite. Notre objectif avec cette architecture est d'établir dans le cadre du projet EXO-POPP une base de données complète comprenant 300 000 actes de mariage de Paris et de ses banlieues, couvrant les années 1880 à 1940, qui sont conservés dans plus de 130 000 scans de doubles pages. Chaque acte de mariage peut comprendre jusqu'à 118 types d'informations distincts qui nécessitent une extraction à partir du texte brut. Dans cet présentation, nous mettons en avant les capacités d'extraction d'informations atteintes par notre architecture en utilisant le nouveau jeu de données M-POPP, qui sera accessible à la communauté de recherche. Ce jeu de données fournit des annotations pour la reconnaissance de texte en pleine page et l'extraction d'informations à travers des documents manuscrits et imprimés. Nous évaluons et comparons également comment différentes stratégies d'encodage pour les entités nommées dans le texte affectent la performance de la reconnaissance conjointe de texte manuscrit et de l'extraction d'informations, à partir de pages entières.

**Lien(s) :** ICDAR 2024

## Orale 5

**Titre :** L'édition des brouillons de Gottfried Wilhelm Leibniz - Approche numérique

**Auteur(s) :** Vincent Giovannangeli<sup>1</sup>, David Rabouin<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> SPHere, PHILIUMM, Paris 7

**Résumé :** We launched an initiative for an open source handwritten text recognition (HTR) model for Leibniz script. One of the main goals of the Philiumm project is to train this model on the HTR platform eScriptorium. This model would be trained with the ground truth provided by the Berlin team and other transcriptions from the volume 4 of the serie 6 (Reihe VI, Band IV in German) that we plan to add to the new model. Our work led us to a close collaboration with the team of the Leibniz-Edition Berlin. Harald Siebert and his team already trained a model of automatic transcription on the Transkribus platform of OCR and HTR. The problem of this former model is that we can't reuse it because Transkribus doesn't comply with the FAIR guidelines. We also organize a session of the seminar Editing Leibniz, to reflect upon the issues of complex layout of the manuscripts, and also mathematical equations and diagrams drawn by Leibniz. These particular objects are not recognized by the HTR models, that is why we also study the possibility to segment the pages, separating the text areas from the equations. For that we have annotated few pages and fine tune an existing U-Net based segmentation model for our documents. Preliminary results are encouraging.

**Lien(s) :** <https://eman-archives.org/philiumm/>

## Orale 6

**Titre :** STRAS: A Semantic Textual-Cues Leveraged Rule-Based Approach for Article Separation in Historical Newspapers

**Auteur(s) :** Nancy Girdhar<sup>1</sup>, Mickaël Coustaty<sup>1</sup>, Antoine Doucet<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

**Résumé :** The digitization of historical documents is a critical task for preserving cultural heritage and making vast amounts of information accessible to the wider public. One of the challenges in this process is separating individual articles from old newspaper images, which is significant for text analysis and information retrieval. In this work, we present a novel approach, Semantic Textual-cues leveraged Rule-based approach for Article Separation (STRAS) in historical newspapers. The presented approach, STRAS, involves utilizing textual information by extracting text region embeddings using scanned input images and their corresponding PAGE format files. The text regions with similar contextual embeddings are then categorized and articles are separated based on a defined rule set. The presented approach is tested on French and Finnish newspapers of the 19th and early 20th centuries. Besides this, novel metrics are introduced specifically for the article separation task: article error rate (AER), article coverage score (ACS), and proper predicted article (PPA). Our study evaluates the performance of various models, including skip-gram (sgSTRAS), continuous-bag-of-words (cbowSTRAS), FastText (ftSTRAS), and pre-trained SpaCy model (preSTRAS), and the results show that the sgSTRAS model achieves the highest mean ACS scores of 0.8343 and 0.8611 on the French and Finnish datasets, respectively, outperforming all other models. Our findings demonstrate that the semantic textual features contain valuable information, and the selection of an appropriate embedding method significantly influences the overall performance of the proposed approach to segment articles. To the best of our knowledge, this is the first study that applies a semantic textual similarity rule-based approach for article separation in historical newspapers, filling a gap in the existing literature and opening up new avenues for further research in this area.

**Lien(s) :** [https://dl.acm.org/doi/10.1007/978-981-99-8085-7\\_8](https://dl.acm.org/doi/10.1007/978-981-99-8085-7_8)

# Présentation invitée

**Titre :** Grands modèles de langue pour le traitement de documents textuels dans domaine médical

**Auteur(s) :** Richard Dufour<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> LS2N, TALN, Nantes Université

**Résumé :** Ces dernières années, les grands modèles de langue (LLM) pré-entraînés ont permis d'atteindre un très haut niveau de performance et de polyvalence sur un large éventail de tâches de traitement du langage (TAL). Alors qu'initialement les modèles étaient entraînés afin de pouvoir s'appliquer le plus largement possible, différents travaux ont montré que, dans le cadre de domaines de spécialité, une adaptation au domaine permettait d'améliorer leurs performances. Dans ce contexte, le domaine médical suscite maintenant un intérêt grandissant de part les enjeux sociétaux et les difficultés qui lui sont associées. Cette présentation sera l'occasion de présenter deux modèles de langue, DrBERT et BioMistral, pré-entraînés sur des objectifs différents (prédictif et génératif) dans le domaine de la santé. Les études associées à ces modèles permettront de mettre en lumière les capacités de ces modèles (performances, capacités multilingues, type d'apprentissage...), leurs limites, et les enjeux futurs.

# Orale 7

**Titre :** Extraction d'informations structurées dans des documents historiques

**Auteur(s) :** Solène Tarride<sup>1</sup>, Mélodie Boillet<sup>1</sup>, Christopher Kermorvant<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> TEKLIA

**Résumé :** Dans cette présentation, TEKLIA partagera son expérience pratique de l'utilisation de modèles de vision encodeur-décodeur (DAN - Document Attention Network) pour l'extraction d'informations structurées à partir d'images de documents historiques. Les différentes étapes nécessaires pour ce type d'extraction seront abordées, mettant en lumière les problématiques rencontrées à chaque étape de la chaîne de traitement. Dans un premier temps, nous aborderons les stratégies d'annotation et de représentation des documents. Ensuite, les stratégies d'entraînement nécessaires à la convergence de tels modèles seront discutées, en détaillant les ressources de calcul et les quantités de données nécessaires à leur optimisation. Nous discuterons également la fiabilité des modèles et leur capacité à localiser les informations pertinentes grâce au mécanisme d'attention intégré. Enfin, nous discuterons des méthodes d'évaluation des modèles end-to-end et mettrons en évidence les limites de telles architectures, notamment les hallucinations et les difficultés liées aux mises en page complexes. Ces différentes étapes seront illustrées par trois cas d'usage concrets réalisés sur des formulaires (Fiches individuelles de soldats - Service Historique de la Défense), des tableaux (Listes nominatives du recensement - Archives Départementales ; Listes de prisonniers - CICR) et des registres (Registres de mariage - Esposalles).

## Orale 8

**Titre :** Extraction automatique de données à partir des formulaires médicaux

**Auteur(s) :** Frédéric Rayar<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> LIFAT, Université de Tours

**Résumé :** Dans le cadre de l'APR IR OPTIMEDIAS, qui vise à fédérer des acteurs académiques et industriels pour valoriser des données de santé par l'Intelligence Artificielle (IA) afin de développer des algorithmes de prédiction du diagnostic, du pronostic et d'aide à la décision thérapeutique, 3 axes d'études ont été identifiés : les troubles du neurodéveloppement (TND), parmi lesquels figurent les troubles du spectre de l'autisme (TSA), la Sclérose Latérale Amyotrophique (SLA) et la Pneumopathie Acquisée sous Ventilation Mécanique (PAVM). Pour l'axe lié aux TND, nous travaillons avec la Plateforme de Coordination et d'Orientation (PCO) 37. « La PCO est un dispositif chargé de mettre en œuvre un parcours d'interventions précoces et de diagnostic pour les enfants entre 0 et 6 ans et 6 et 12 ans pour lesquels l'entourage et/ou le médecin constate un écart inhabituel de développement. ». Concernant les enfants de moins de 7 ans, il s'agit de répondre à une problématique de repérer pour intervenir précocement via la détection de signes d'un développement inhabituel. Ce parcours d'intervention, généralement initié par le médecin généraliste, fait par la suite intervenir de nombreux médecins spécialistes (psychologues, orthophonistes, psychomotriciens, ergothérapeutes, etc.) qui génèrent tout un ensemble de documents et formulaires hétérogènes (type d'écriture, format, qualité, etc.) qui sont actuellement traités manuellement, empêchant une exploitation efficace par des solutions d'IA. Parmi l'ensemble de ces documents, le premier - appelé le formulaire d'adressage - est renseigné par le médecin généraliste puis transmis à la plateforme. Il contient une grille de repérage qui permet de porter l'attention sur des points clés du développement global de l'enfant et ainsi repérer d'éventuels écarts inhabituels de développement. En particulier, ce formulaire est composé de plusieurs pages : celles avec des indicateurs quel que soit l'âge de l'enfant, celles présentant des signes d'alerte en fonction de l'âge et celle contenant les recommandations d'adressage vers la plateforme. Nous présentons ici des travaux préliminaires pour permettre le traitement automatique des formulaires d'adressage afin d'en extraire les informations pertinentes. Ces formulaires contenant des données de santé - à caractère personnel et sensibles donc - ils font l'objet d'une protection particulière par les textes (règlement européen sur la protection des données personnelles, de la santé publique, etc.). Par conséquent, nous ne disposons à l'heure actuelle que de très peu de ces formulaires, anonymisés, pour travailler : une approche d'apprentissage profond (deep learning) n'est donc pas envisageable. Ainsi, c'est une chaîne de traitement plus « classique » qui a été mise en place et implémentée dans une solution logicielle intuitive et ergonomique pour être utilisée par des non-experts. Des étapes de recalage d'image, de template matching, de reconnaissance optique de caractère et de reconnaissance optique de marques sont ainsi exploitées. Les premiers résultats, les difficultés rencontrées et les perspectives de ces travaux préliminaires seront présentés et discutés.

## Orale 9

**Titre :** CoastTerm: a corpus for Multidisciplinary Term Extraction in Coastal Scientific Literature

**Auteur(s) :** Julien Delaunay<sup>1,2</sup>, Hanh Thi Hong Tran<sup>1,3</sup>, Carlos-Emiliano González-Gallardo<sup>1</sup>, Georgeta Bordea<sup>1</sup>, Mathilde Ducos<sup>1</sup>, Nicolas Sidere<sup>1</sup>, Antoine Doucet<sup>1</sup>, Senja Pollak<sup>3</sup>, Olivier De Viron<sup>2</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

<sup>2</sup> LIENSs, Université La Rochelle, La Rochelle,

<sup>3</sup> Jozef Stefan Institute, Ljubljana, Slovenia

**Résumé :** The growing impact of climate change on coastal areas, particularly active but fragile regions, necessitates collaboration among diverse stakeholders to formulate effective environmental protection policies. Furthermore, the vast and diverse body of scientific literature studying these intricate coastal ecosystems presents a challenge for researchers to keep pace with relevant findings outside their specific fields of expertise. Thus, the ability to extract pertinent information assumes critical importance in facilitating interdisciplinary studies of coastal areas. In this paper, we introduce a novel multi-disciplinary corpus comprising 2,491 sentences from 410 scientific abstracts concerning coastal areas, for the Automatic Term Extraction (ATE) and Classification (ATC) tasks. Inspired by the ARDI framework, our focus lies in extracting informative data concerning the functioning of coastal systems, and the identified classes aim to illuminate the distinct roles of entities within the complex systems described in the abstracts. We also propose fine-tuned monolingual and multilingual term extractors as benchmarks. The evaluation demonstrates consistent results, achieving an F1 score of approximately 80%. These findings are promising and signify an initial step towards the development of a specialized Knowledge Base dedicated to coastal areas.

**Lien(s) :** 27th International Conference on Text, Speech and Dialogue

# Session Poster :

## Poster 1

**Titre :** Toward accessible comics for blind and low vision readers

**Auteur(s) :** Christophe Rigaud<sup>1,2</sup>, Jean-Christophe Burie<sup>1</sup>, Samuel Petit<sup>2</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

<sup>2</sup> ComixAI, La Rochelle

**Résumé :** A large amount of comics could be made accessible for blind and low vision readers using recent advances of text recognition combined with image description and speech synthesis. This work explores how to fine-tune recent large language model with content-related information in order to produce an accurate text description of the full story ready to be forwarded to off-the-shelve speech synthesis tools. The textual description is intended to mirror the image content by providing a complete scene understanding with character names, dialogues and interaction in natural language and following a proper reading order without repeating too much information from panel-to-panel. We propose to use existing computer vision and optical character recognition techniques to extract information from the comic strip images, such as panels, characters, text, reading order and the association of bubbles and characters. Then, we use them as contextual information to prompt several large language models to infer character's names and generate different level of context-aware description for panels (overall scene description with character's appearance, posture, mood and dialogues), pages summary, albums script. We believe that such enriched content description can be then easily used to produce audiobook and eBook with various human-like voices for characters, narration and playing sound effects.

## Poster 2

**Titre :** Vers la détection de contrefaçon de documents : Exploitation de l'analyse visuelle pour une détection améliorée

**Auteur(s) :** Gardella Marina<sup>1</sup>, Antoine Tadros<sup>1</sup>, Seginus Mowlavi<sup>1</sup>, Natalia Bottaioli<sup>1</sup>, Jean-Michel Morel<sup>2</sup>, Gabriele Facciolo<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> ENS Paris-Saclay

<sup>2</sup> City University of Hong Kong

**Résumé :** La détection de falsification de documents joue un rôle crucial dans la sauvegarde de l'intégrité de divers secteurs, y compris la finance, l'application de la loi et la sécurité nationale. La capacité à distinguer entre les documents authentiques et contrefaits est essentielle pour garantir la confiance dans les transactions, maintenir la conformité réglementaire et prévenir les activités frauduleuses. Malgré son importance, le domaine de la détection de la contrefaçon de documents reste sous-développé, notamment en comparaison avec les progrès réalisés dans la détection de la contrefaçon d'images. Notre enquête examine en détail les documents contrefaits, découvrant des indices subtils mais cruciaux pouvant être exploités pour la détection de la contrefaçon. Grâce à un examen méticuleux, nous avons observé des caractéristiques distinctes au sein des documents contrefaits qui fournissent des informations précieuses sur le processus de contrefaçon. Notamment, les caractères insérés dans ces documents présentent souvent des propriétés uniques : ils manquent du bruit typique présent dans les caractères authentiques et ne présentent pas d'aberrations chromatiques. De plus, une caractéristique courante des caractères insérés est l'utilisation de polices de caractères non correspondantes, une déviation subtile souvent imperceptible à l'œil nu mais détectable grâce à des outils spécifiques. Ces conclusions servent de repères essentiels dans notre quête pour développer des algorithmes robustes de détection de contrefaçon. En exploitant ces anomalies visuelles, nous visons à construire un cadre complet capable de discerner avec précision un texte authentique d'un texte contrefait. En intégrant ces informations dans des systèmes de détection de contrefaçon en temps réel, nous aspirons à fournir aux autorités et aux organisations les outils nécessaires pour lutter efficacement contre la fraude. La capacité à identifier les documents contrefaits, même lorsqu'ils sont méticuleusement fabriqués, offre une immense promesse pour protéger l'intégrité des transactions et des processus cruciaux.

## Poster 3

**Titre :** IntuiSketch : Interprétation des tracés semi-structurés pour la conception d'un système tutoriel intelligent orienté stylet

**Auteur(s) :** Islam Barchouch<sup>1</sup>, Eric Anquetil<sup>1</sup>, Nathalie Girard<sup>1</sup>, Omar Krichen<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> IRISA, Université de Rennes, INSA Rennes, CNRS, Rennes

**Résumé :** Ce travail présente une avancée significative dans la conception d'IntuiSketch, un système tutoriel intelligent pour l'aide à l'apprentissage par le dessin, en utilisant des tablettes orientées stylet (1). L'application de ce système cible spécifiquement la production de croquis d'anatomie pour aider les étudiants dans les filières de Santé (médecine, kiné, ...). L'objectif principal de ce travail de recherche est double. Pour la reconnaissance, le système doit être capable d'interpréter à la volée l'élaboration des croquis manuscrits semi-structurés. Pour cela, le système s'appuie sur un moteur de reconnaissance reposant sur une représentation bi-dimensionnelle du document i.e. des règles de grammaire visuelle CD-CMG (2). Pour aider l'étudiant dans ses apprentissages, le système doit produire des feedbacks de correction ou de guidage selon ce que l'étudiant a produit (en mode étudiant) et en rapport avec ce qui est attendu. Ces feedbacks sont produits par un tuteur intelligent basé sur un graphe de connaissance lié de manière cohérente aux propriétés des croquis d'anatomie produits par l'enseignant (en mode auteur) (3). Le travail présenté dans ce résumé vise à rendre plus souple la reconnaissance des tracés manuscrits d'anatomie afin d'introduire davantage de tolérance et de différer les feedbacks à des points d'étapes clés de l'élaboration du dessin. Pour ce faire, nous avons étendu le langage des grammaires visuelles CD-CMG en introduisant deux catégories de contraintes dans les règles d'interprétation CD-CMG. Cette extension catégorise les contraintes de forme, de positionnement et de relations géométriques des règles CD-CMG en deux catégories distinctes : les contraintes primaires, essentielles à une interprétation de base minimale des tracés manuscrits d'anatomie et qui doivent être validées pour garantir une reconnaissance du dessin produit, et les contraintes secondaires, qui permettent d'apprécier ensuite la qualité de la réalisation des étudiants par le tuteur intelligent. Cela lui permet de produire des feedbacks en temps réel et/ou différés, qualitatifs et précis, en fonction des propriétés du dessin. Cette nouvelle approche permet à IntuiSketch de reconnaître efficacement les dessins d'anatomie, même dans les cas où les contraintes secondaires ne sont pas entièrement satisfaites. De cette manière, le système est capable de fournir à la fois des feedbacks en temps réel lorsque les contraintes primaires ne sont pas respectées, ce qui permet à l'étudiant de corriger immédiatement et de continuer son travail ; Mais aussi, des feedbacks différés à des points d'étapes clés du dessin lorsque les contraintes secondaires ne sont pas respectées, évitant ainsi d'interrompre l'étudiant dans sa production. Ces feedbacks sont ainsi adaptés aux réalisations individuelles de chaque apprenant, ce qui favorise l'apprentissage. Ce travail présente donc deux contributions majeures. D'une part, le système peut analyser et reconnaître des dessins produits à main levée avec plus de tolérance. D'autre part, il évite d'interrompre excessivement l'étudiant à chaque tracé fait en reportant les feedbacks

correspondants à des erreurs mineurs à des points d'étapes clés du dessin. Cette double fonctionnalité permet une expérience d'apprentissage plus fluide grâce à des feedbacks bien équilibrés pour assurer une progression optimale de l'apprenant. Les évolutions de notre tuteur intelligent ont été conçues en partant de deux exercices spécifiques à la réalisation des croquis d'anatomie, portant respectivement sur le rachis et l'œil. Ces exercices ont été choisis pour leur représentativité des structures différentes d'anatomie et les défis spécifiques qu'ils posent en termes de reconnaissance des détails de dessin. Les expériences seront menées avec des étudiants de la Faculté de médecine afin d'évaluer les performances du tuteur intelligent IntuiSketch. L'objectif est d'analyser à la fois l'efficacité du système à reconnaître plus fluidement les tracés d'anatomie et à détecter les erreurs, ainsi que l'acceptabilité de ce double niveau de feedbacks : feedbacks immédiats pour les erreurs majeures et feedbacks différés pour les erreurs mineures. Cette approche méthodologique permettra d'évaluer en profondeur la pertinence et l'utilité du système IntuiSketch dans le contexte spécifique de l'apprentissage de l'anatomie par le dessin.

**Référence** : BARCHOUC, Islam, KRICHEN, Omar, ANQUETIL, Eric, et al. IntuiSketch, a pen-based tutoring system for anatomy sketch learning. In : International Graphonomics Conference. Cham : Springer Nature Switzerland, 2023. p. 61-74.2.

MACÉ, Sébastien et ANQUETIL, Eric. Eager interpretation of on-line hand-drawn structured documents: The dali methodology. Pattern Recognition, 2009, vol. 42, no 12, p. 3202-3214.3.

MITROVIC, Antonija. Fifteen years of constraint-based tutors: what we have achieved and where we are going. User modeling and user-adapted interaction, 2012, vol. 22, p. 39-72.

## Poster 4

**Titre :** External language models and rejection capabilities for text recognition in difficult conditions

**Auteur(s) :** Florent Meyer<sup>1,2</sup>

**Affiliation(s) :**

<sup>1</sup> IRISA, Université de Rennes, INSA Rennes, CNRS, Rennes

<sup>2</sup> ANTAI

**Résumé :** Vehicles license plate recognition is an OCR application for which the tremendous volume of contraventions recorded accounts for the need of a very high performance in the reading. Unfortunately, additionally to common image quality-related difficulties emerging from weather conditions or camera settings, the distribution of the characters on the plates is constantly evolving due to vehicles being put in and out of service which makes it harder for systems to preserve peak performance over time. A lever towards obtaining a predictor capable of adapting nearly daily is to first cancel the bias – being some kind of a “language model” – in the plates in order to better reinject a useful, up-to-date bias later on without fully retraining the model. The format which this external knowledge should take remains to be defined. To this end, work is currently being conducted on revealing where this undesired bias lies in a neural network and how it manifests, as well as on how it could be removed. Datasets on both French-only and multi-country license plates – thus having different syntaxes – have been built. ResNet-like models with diverse output layers have thus been explored since the end of 2023, as well as SOTA TrOCR variants as those have an explicit language decoder built in. The main idea behind our tests being to compare the performance between 1) models trained and tested on images taken the same date range and 2) models trained on some date range and tested on later dates. The experiments done so far tend to show that ResNet-like models do not hold any language model while Transformer-like ones do. We wish to present these preliminary results and the complementary ones to come.

## Poster 5

**Titre :** Reconnaissance de formules mathématiques pour le référencement de documents techniques

**Auteur(s) :** François Wieckowiak<sup>1,2</sup>, Véronique Eglin<sup>1</sup>, Tony Bonnet<sup>2</sup>

**Affiliation(s) :**

<sup>1</sup> LIRIS

<sup>2</sup> Luminess

**Résumé :** Les équations mathématiques des documents techniques peuvent être rédigées et diffusées dans une variété de formats textuels (LaTeX, MathML, ASCIIMath) et visuels (PNG, SVG). Certains formats peuvent limiter l'accessibilité de ces documents aux personnes malvoyantes et compliquer leur référencement dans les moteurs de recherche. La communauté scientifique pourrait bénéficier d'une solution robuste et ouverte pour la conversion d'images de formules imprimées vers LaTeX et MathML. Notre état de l'art identifie des solutions open-source peu robustes, car entraînées sur des jeux de données conventionnels (images nettes, unicité de police et de format). Les solutions closed-source présentent des résultats satisfaisants, mais ne divulguent ni leurs architectures ni leurs entraînements, et nécessitent des abonnements payants. L'entreprise partenaire Luminess convertit semi-automatiquement les numérisations des brevets européens en document XML pour permettre leur publication et leur référencement, à un taux de qualité de 99.99%. Les équations de ces documents ont été traitées et sauvegardées dans un jeu de données original de 360 000 images de formules aux formats, polices et qualités variables, associées à leurs saisies en MathML. Ces données permettent l'élaboration d'un nouveau benchmark et la conduite de nouveaux entraînements pour mesurer et améliorer la robustesse des systèmes de reconnaissance actuels. Une métrique multimodale (texte, image, sémantique) est à définir pour assurer la fiabilité des prédictions. Les premiers essais sur les formules des brevets avec les solutions actuelles montrent une augmentation de la distance d'édition normalisée de 35 points (de 0.12 à 0.47) par rapport au jeu de données conventionnel Im2LaTeX. Ces résultats semblent démontrer une inadéquation de ces systèmes aux particularités des données et à la qualité attendue par Luminess.

## Poster 6

**Titre :** Lecture et compréhension de documents numérisés

**Auteur(s) :** Simon Tom<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> LITIS, Université de Rouen

**Résumé :** Les systèmes de reconnaissance d'écriture sont souvent spécialisés, nécessitant un entraînement spécifique pour chaque nouveau jeu de données, tel qu'une nouvelle langue ou une mise en page différente. Pour améliorer leur adaptabilité, une interaction flexible entre l'utilisateur et le réseau de neurones, sans nécessité de reprise d'entraînement, serait idéale. Cela soulève la question de savoir comment faciliter cette interaction pour optimiser l'adaptation du réseau à de nouveaux scénarios. Dans mes travaux, je propose d'explorer l'approche de l'apprentissage en contexte (in-context learning), une méthode introduite en 2022 avec le modèle multimodal Flamingo. L'idée est de fournir des exemples au modèle sur lesquels il peut s'appuyer pour améliorer ses prédictions. Mes premières recherches se concentreront donc sur le développement d'un réseau de neurones permettant l'apprentissage en contexte sur des images de documents. Cette approche est prometteuse car elle pourrait permettre au modèle de s'adapter à de nouvelles données sans nécessiter un nouvel entraînement complet du réseau.

## Poster 7

**Titre :** Amélioration de la détection et de la reconnaissance des structures de tableaux dans les documents textuels grâce à une approche basée sur DETR et TATR

**Auteur(s) :** Eliott Thomas<sup>1,2</sup>, Aurélie Joseph<sup>2</sup>, Mickaël Coustaty<sup>1</sup>, Vincent Poulain D'andecy<sup>2</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

<sup>2</sup> Yooz

**Résumé :** Les tableaux représentent des structures omniprésentes dans les documents textuels, jouant un rôle crucial dans de nombreuses applications, notamment dans le contexte industriel pour la gestion des factures. La détection précise des lignes de produits au sein de ces tableaux est essentielle pour une analyse approfondie. Cependant, cette tâche est confrontée à plusieurs défis, tels que la présence de tableaux multiples, l'absence de structure tabulaire claire et la variation des formats et styles. Des approches récentes basées sur DETR (détection d'objets), notamment sa variante TATR fine-tunée sur PubTable-1M, ont démontré des résultats prometteurs pour la détection de tableaux. Pour évaluer la performance de ces méthodes, nous utilisons des métriques telles que la précision, le rappel, la pureté et la complétude pour la détection des tableaux, ainsi que l'intersection sur union (IOU) centrée sur les cellules et la liste 2D des relations entre les cellules pour la reconnaissance de la structure. Nous comparerons nos résultats sur des données publiques telles que le jeu de données ICDAR 2013 et un jeu de données de test privé. Pour évaluer l'efficacité de notre approche, nous avons comparé nos résultats préliminaires avec ceux d'ABBY, un système bien établi dans le domaine de la détection de tableaux. Sur notre jeu de données de test privé, notre méthode de détection de tableaux a identifié 185 documents contenant des tableaux, avec un taux d'erreur de seulement 19%. En contraste, ABBY n'a détecté que 165 documents, avec un taux d'erreur plus élevé de 39%. Ces premiers résultats indiquent la promesse de notre approche basée sur DETR et TATR pour améliorer la détection de tableaux. Notre approche pour améliorer l'extraction et la reconnaissance de la structure des tableaux se divise en deux étapes distinctes : la détection des tableaux et la reconnaissance de la structure. Dans la première étape, nous nous concentrons sur l'identification des tableaux dans le document. Pour ce faire, nous avons adopté une approche basée sur le modèle DETR. Nous avons développé deux modules complémentaires : Bbox chooser: Ce module sélectionne les meilleures boîtes englobantes prédites par DETR en utilisant différentes stratégies telles que le score de confiance du modèle, la taille des boîtes et la présence de mots clés pertinents. Bbox pruner: Ce module affine la zone prédite par DETR et sélectionnée par le chooser en éliminant les parties superflues ou incorrectes. La deuxième étape se concentre sur l'affinement de la structure des tableaux reconnue par TATR. Nous utilisons deux modules complémentaires pour cela : Cleaner: Ce module nettoie les prédictions de TATR en s'adaptant à tout modèle de détection d'objets. Il utilise des techniques d'analyse des scores de confiance, des règles

heuristiques basées sur la structure du tableau et des techniques d'apprentissage automatique pour identifier et corriger les erreurs de segmentation. Post-processing: Ce module effectue un dernier traitement sur le tableau final prédit. Il vérifie la cohérence des données, corrige les erreurs de formatage telles que l'alignement des cellules et la suppression des espaces vides, et infère des informations manquantes à partir du contexte.

## Poster 8

**Titre :** LIAS: Layout Information-based Article Separation in Historical Newspapers

**Auteur(s) :** Sun Wenjun<sup>1</sup>, Thi Hong Hanh Tran<sup>1,2</sup>, Carlos-Emiliano González-Gallardo<sup>1</sup>, Mickaël Coustaty<sup>1</sup>, Antoine Doucet<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

<sup>2</sup> Jozef Stefan Institute, Ljubljana, Slovenia

**Résumé :** Nowadays, the increasing digitization of historical document resources not only provides researchers with abundant materials but also introduces a new challenge: how to analyze their contents swiftly and accurately? Article separation emerges as a solution to this issue, overcoming the disorder and inefficiency associated with querying the original document material directly. However, there is limited in-depth research on this topic currently. Given that historical documents contain information in multiple modalities such as text and images, the reasonable integration and analysis of multimodal information pose a further challenge. In response to these challenges, we propose LIAS, a method based on layout information, and conduct experiments on historical newspapers. In comparison to existing work, LIAS introduces a fresh perspective to research in this area. The method initially identifies the separator lines of the newspaper, analyzes the layout information to reconstruct the information flow of the document, performs segmentation based on the semantic relationship of each text block in the information flow, and ultimately achieves article separation. The experiments encompass diverse historical newspapers in French and Finnish, featuring various layouts. Results demonstrate that LIAS outperforms current schemes in most cases, as evidenced by metrics specifically designed for article separation tasks.

## Poster 9

**Titre :** Extraction d'éléments textuels complexes, Application a la détection et a la reconnaissance des onomatopées dans les bandes dessinées.

**Auteur(s) :** John Benson Louis<sup>1</sup>, Jean Christophe Burie<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

**Résumé :** Les bandes dessinées représentent un médium unique et harmonieusement images et écrits, mélangeant dialogues traditionnellement confinés dans des bulles et textes intégrés de manière plus organique au sein du dessin, comme c'est le cas pour les onomatopées. Si les dialogues enfermés dans ces ballons sont relativement simples a détecter et a extraire en raison de leur uniformité, les onomatopées contrastent fortement. Ces éléments textuels particuliers, dont la taille, la forme et le style varient grandement d'une œuvre à l'autre, constituent un réel défi pour la détection automatique. Les méthodes conventionnelles de détection de texte, bien qu'efficaces pour identifier des contenus textuels simples et structurés, se heurtent à la complexité des onomatopées. Nos premières investigations ont mis à l'épreuve diverses techniques fondées sur l'apprentissage profond, révélant leurs insuffisances face à cette tâche spécifique. Forts des enseignements tirés de ces expérimentations initiales, nous avons proposé une méthode, capable non seulement de détecter mais aussi de reconnaître les onomatopées dans leur diversité et leur subtilité graphique.

## Poster 10

**Titre :** Modèles de fondation intégrés pour les Bibliothèques, Archives et Musées

**Auteur(s) :** Solène Tarride<sup>1</sup>, Christopher Kermorvant<sup>1</sup>, William Mocaër<sup>2</sup>, Thierry Paquet<sup>2</sup>, Pierrick Tranouez<sup>2</sup>, Sébastien Cretin<sup>3</sup>, Jean-Philippe Moreux<sup>3</sup>

**Affiliation(s) :**

<sup>1</sup> TEKLIA

<sup>2</sup> LITIS, Université de Rouen

<sup>3</sup> BnF

**Résumé :** La transformation numérique des bibliothèques, qui repose depuis 20 ans sur la technologie de reconnaissance optique de caractères (OCR), présente certaines limites à la fois en termes de qualité, en raison de la diversité des collections et des limites de la technologie OCR, et en termes de valeur ajoutée, en raison d'un manque de structuration et d'indexation avancée. Dans le cadre du projet ANR TSIA Finlam, nous proposons de développer des modèles de fondation multimodaux (texte + image) pour extraire des informations à partir de collections de documents historiques conservés dans des Bibliothèques, Archives et Musées. Notre premier cas d'usage concerne l'extraction d'informations dans des images de presse ancienne. La littérature montre que les recherches dans ce domaine se concentrent principalement sur le traitement de documents commerciaux modernes, tels que les factures ou les tickets de caisse.

# Poster 11

**Titre :** Simple Structural Model for Few-Shot Learning and Multi-Occurrence Field Recognition

**Auteur(s) :** Gaspar Deloin<sup>1</sup>, Elodie Carel<sup>1</sup>, Aurélie Joseph<sup>1</sup>, Vincent Poulain d'Andecy<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> YOOZ

**Résumé :** Ces dernières années, l'utilisation du deep learning pour l'extraction d'informations a été la tendance forte. Toutefois, cela requiert des ressources computationnelles importantes ainsi qu'un volume conséquent de données. De plus, le manque d'explicabilité est un problème dans un contexte industriel. Les travaux présentés ont commencé en utilisant une méthode existante qui crée un modèle spécifique pour chaque fournisseur. Ce modèle représente chaque élément appris sous forme de graphe en étoile en utilisant ses k plus proches mots. Malgré de bons résultats, plusieurs problèmes sont apparus. Le premier, était l'incapacité à rejeter certaines prédictions. Pour y remédier, un système de confiance a été mis en place. De plus, tandis que l'ancienne version augmentait sa taille de façon incontrôlée en ajoutant de nouveaux mots à chaque apprentissage, la nouvelle approche limite un nouvel apprentissage à trois cas distincts. Elle renforce les mots ayant contribué correctement à la décision en cas de prédiction erronée, ajuste la taille de la prédiction lorsque nécessaire, ou crée un nouveau graphe lorsque les deux premiers cas ne s'appliquent pas. Contrairement au système précédent, une mise à jour du modèle n'a lieu que si une prédiction erronée est produite. Cette nouvelle approche a démontré des améliorations significatives en termes de rappel et de précision par rapport à la méthode précédente. Par exemple, après seulement deux apprentissages, les performances sont passées de 96,27% en rappel et 99,38% en précision à 99,37% en rappel et 99,79% en précision sur le dataset privé trié par template. Sur un sous-ensemble du dataset DocILE trié par template, présenté à ICDAR 2023, cela est passé de 94,70% en rappel et 95,55% en précision à 96,98% et 96,97% respectivement. Enfin, une nouvelle fonctionnalité a été ajoutée : la prédiction des champs multi-occurents. Cela se fonde sur le principe que les mots alignés horizontalement dans le graphe peuvent prédire plusieurs champs ayant la même signification dans le document (comme les informations dans des lignes de tableau) . Cela a permis d'extraire parfaitement les listes de bons de livraisons et de bons de commande dans le dataset privé. Avec un nouvel apprentissage, l'extraction des identifiants d'article a vu son rappel passer de 60,10% à 83,94% et sa précision de 66,67% à 73,64%. De plus, une extraction avec un rappel et une précision d'au moins 80% a été obtenue pour presque tous les champs multi-occurents dans le dataset DocILE.

## Poster 12

**Titre :** Évaluation de Large Language Models pour la description de photographies

**Auteur(s) :** Alexis Imbert<sup>1</sup>, Mélodie Boillet<sup>1</sup>, Christopher Kermovant<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> TEKLIA

**Résumé :** Le Musée national des Arts asiatiques - Guimet (MNAAG) et Teklia se sont associés dans le cadre du projet HikarIA pour la mise en valeur de photographies patrimoniales grâce à l'intelligence artificielle et le développement d'outils d'indexation automatisée de photographies anciennes. Au cœur du projet se trouve une collection exceptionnelle de plus de 20 000 photographies du Japon au XIXe et au début du XXe siècle, pour la plupart collectées par le Dr. Joseph Dubois et acquises par le Musée Guimet en 2007. Le but du projet est d'adapter des modèles d'IA à l'analyse de ces photographies et de les mettre en œuvre pour produire des données fiables sur l'iconographie du corpus. Images et métadonnées seront ensuite mises en ligne sur une plateforme d'exploration du corpus librement accessible. Dans cette communication, nous comparons différents modèles de langage multimodaux afin de décrire automatiquement ces photographies anciennes. Ces modèles prennent une image et une instruction (prompt) en entrée, et génèrent une description de l'image en sortie. De nombreux modèles de ce type ont été proposés ces dernières années, tels que LLaVA, BackLLaVA, BLIP, CLIP et MobileCLIP. Dans un premier temps, nous comparons différentes stratégies de prompt pour adapter ces modèles sans ré-entraînement. Puis, nous évaluons l'impact du fine-tuning sur les performances des modèles. Nous discutons également des besoins en ressources de calcul ainsi que de la quantité de données nécessaire. Enfin, nous présentons quelques bibliothèques d'évaluation automatisées de LLMs, telles que promptfoo et benchprompt, ainsi que les métriques d'évaluation pertinentes pour évaluer la sortie d'une combinaison modèle/prompt.

## Poster 13

**Titre :** Apprentissage faiblement supervisé pour la vérification des hologrammes dans les documents d'identité

**Auteur(s) :** Glen Pouliquen<sup>1,2</sup>, Guillaume Chiron<sup>1</sup>, Joseph Chazalon<sup>2</sup>, Ahmad Montaser Awal<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> IDnow

<sup>2</sup> LRE

**Résumé :** Nous proposons une méthode pour vérifier à distance l'authenticité des dispositifs optiquement variables (OVD), souvent appelés "hologrammes", dans des documents d'identité. Notre méthode traite des vidéos capturées avec des smartphones dans des conditions d'éclairage habituelles. Nous l'avons évaluée sur deux ensembles de données publics : MIDV-HOLO et MIDV-2020. Grâce à un apprentissage faiblement supervisé, nous avons optimisé une chaîne d'extraction de caractéristiques et de décision qui atteint une nouvelle performance état de l'art sur MIDV-HOLO, tout en maintenant un rappel élevé sur les documents du dataset MIDV-2020, qui ont été utilisés comme attaque. Notre méthode est également de la première méthode à s'attaquer efficacement à la tâche d'attaque de remplacement de photo. Elle peut être entraînée sur des vidéos authentiques, des vidéos d'attaques ou les deux pour une meilleure performance. En permettant de vérifier les formes et la dynamique des OVD avec peu de supervision, notre travail ouvre la voie à l'utilisation massive de données non étiquetées pour construire des systèmes robustes de vérification de documents d'identité à partir de captures vidéo.

## Poster 14

**Titre :** Plural, flexible and open evaluation of research specialisation indices

**Auteur(s) :** Dipendra Sharma Kaffle<sup>1</sup>, Antoine Doucet<sup>1</sup>, Esma Talhi<sup>2</sup>, Mickael Coustaty<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> L3i, Université La Rochelle, La Rochelle

<sup>2</sup> EIGSI, Université La Rochelle, La Rochelle

**Résumé :** Higher education has undergone significant development in recent years. The intersection between different fields and the diversification of pathways and specialities have led to the emergence of new institutions, universities, research laboratories, teaching programmes, etc. Faced with this growth, students, researchers and university managers need information on the content of the courses on offer and the research areas of the laboratories with which they will be collaborating. In order to identify the teams and universities working on specific themes, it is necessary to carry out a discursive analysis of all the documents describing these activities, in particular communications but also research articles. The latter make it possible to identify a university's research areas by analysing the content and the authors' affiliations. This analysis enhances the methodology by providing it with a knowledge model, organised in the form of taxonomies that can be used to organise and visualise the knowledge extracted: the domains of the universities, their collaboration, research themes, the universities that deal with the same domains, and so on. The goal of this project is to be able to take interdisciplinary activities into account, but also to develop a method that works on the basis of keywords or phrases given as input.

## Poster 15

**Titre :** Graph-based automatic summarization the content of unknown business documents

**Auteur(s) :** Baglan Aitu<sup>1</sup>

**Affiliation(s) :**

<sup>1</sup> Université La Rochelle, La Rochelle

**Résumé :** Document Image Analysis is an essential task in today's digital era, where businesses and organizations rely heavily on processing vast amounts of information contained in various types of documents. However, this task presents numerous challenges due to factors such as poor scan quality, small fonts, inconsistent layouts across different document formats, complex structures, and the presence of domain-specific languages. These difficulties hinder the effective application of Natural Language Processing (NLP) techniques, which are typically designed to handle textual data. While NLP techniques excel at processing unstructured text, they face significant limitations when it comes to documents that incorporate a diverse range of elements beyond plain text. Elements such as tables, figures, graphs, logos, and signatures contribute to the document's structure and semantic meaning. Analyzing and understanding these elements pose a formidable challenge, as they require a more comprehensive approach that goes beyond traditional NLP. To tackle the complexities of document analysis, this work aims to leverage the combined power of computer vision and NLP techniques based on graph neural networks. By fusing visual, textual, and positional information, limitations of purely text-based approaches can be overcome and more robust classification and understanding of documents can be achieved.