

PROPOSITION DE SUJET DE THÈSE

CONTRATS DOCTORAUX

2021–2024

Appel ciblé : Contrat doctoral ministériel ED 536

- Directeur de thèse :** Richard Dufour email : richard.dufour@univ-avignon.fr
Laboratoire Informatique d'Avignon
- Co-encadrant :** Vincent Labatut email : vincent.labatut@univ-avignon.fr
Laboratoire Informatique d'Avignon
- Titre en français :** Traitement du langage et modélisation de relations pour la représentation unifiée de documents narratifs
- Titre en anglais :** Natural language processing and relationship modelisation for the unified representation of narrative documents
- Résumé :** Dans le domaine du traitement du langage, la question de la construction de représentations pertinentes de mots ou de phrases est capitale pour de nombreuses applications. Or, il existe peu de travaux sur la représentation de documents ou de corpus à grande échelle. L'objectif de cette thèse est de proposer de nouvelles approches permettant la représentation multimodale de documents narratif (articles de presse, notices biographiques, romans) ou de corpus de tels documents, en s'appuyant notamment sur des graphes d'interactions entre acteurs. Cette représentation multimodale ayant vocation à être utilisée dans la résolution de tâches concrètes, l'évaluation de celle-ci sur plusieurs applications telles que la classification, la recommandation et la segmentation de textes permettra de valider sa pertinence.
- Mots-clés :** Traitement automatique du langage naturel, Apprentissage profond, Analyse de graphes, Documents narratifs

1	Présentation détaillée du sujet	1
1.1	Contexte et enjeux	1
1.2	Objectifs	2
1.3	Méthode	3
1.4	Organisation	5
2	Profil du candidat ou de la candidate	6
3	Opportunités de mobilité	6
4	Références bibliographiques	6

1 Présentation détaillée du sujet

1.1 Contexte et enjeux

Ces dernières années, le domaine du traitement automatique du langage naturel (TALN) est en plein essor grâce à de nombreux progrès liés à l'utilisation d'approches fondées sur l'apprentissage profond. Plusieurs architectures neuronales telles que les réseaux de neurones convolutifs, les réseaux

de neurones récurrents [7], ou encore les transformers [22], ont été utilisées avec succès dans la résolution de différentes tâches de TALN, notamment à l'écrit.

Une grande partie de ces avancées réside dans l'utilisation de ces architectures neuronales pour le développement de nouvelles méthodes permettant la représentation des mots dans des espaces de grandes dimensions sous forme de vecteurs (plongements lexicaux ou *word embeddings*). Ces vecteurs se révèlent extrêmement utiles, puisqu'ils peuvent, par exemple, constituer les données d'entrée d'autres algorithmes d'apprentissage, comme des classificateurs. L'un des points de départ de ces évolutions est la technique Word2Vec [17], une méthode d'apprentissage non-supervisée où la couche cachée unique constitue un plongement lexical. L'un des intérêts de cette approche est l'apprentissage des mots et de leur contexte dans un même espace de représentation, permettant un calcul de similarité entre les mots, et faisant apparaître des rapprochements sémantiques et/ou syntaxiques entre eux. Plus récemment, les modèles profonds pré-entraînés permettant la vectorisation contextuelle des mots tels que BERT [5] ou XLNET [27] ont contribué à faire considérablement progresser l'état de l'art dans de nombreuses tâches de TALN. En tentant de résoudre une tâche très générale sur un très grand nombre de données, comme la prédiction d'un mot masqué dans un texte, ces modèles dérivent des représentations très polyvalentes. Il est ensuite possible d'affiner ces représentations pour résoudre des tâches en aval (classification de documents, reconnaissance d'entités nommées...) : on parle alors de *d'apprentissage par transfert* [19].

Si des techniques permettant la représentation de documents existent également [21, 3], celles-ci se limitent en général à l'échelle de la phrase ou du paragraphe. Comparativement, la question de la représentation de documents à plus grande échelle (articles, livres...), voire d'ensembles de documents, est peu explorée par la littérature. À cette échelle, si l'on considère qu'un corpus de documents *narratifs* est constitué d'un ensemble d'acteurs et de leurs interactions (par exemple, les personnages d'un roman, ou des personnalités publiques citées dans un corpus journalistique), une représentation sous forme de graphe peut se révéler plus adaptée. De tels graphes peuvent être construits sur la base des résultats produits par les outils de TALN mentionnés au paragraphe précédent. Ce type de représentation possède plusieurs avantages : tout d'abord, il permet de rendre compte explicitement des relations entre acteurs, voire de les visualiser. En outre, l'utilisation de graphes dynamiques (*i.e* qui évoluent de façon diachronique) permettent de témoigner de l'aspect temporel du document ou du corpus. Enfin, les avantages d'une représentation vectorielle ne sont pas pour autant perdus : il est toujours possible d'obtenir un vecteur utilisable comme entrée d'autres algorithmes, en utilisant par exemple des techniques de plongements de graphes [2] (*graph embeddings*).

Un autre aspect peu développé dans la représentation de documents est l'intégration de données multimodales ou multisources. Or, les graphes se prêtent bien à leur représentation. La littérature contient ainsi des exemples de méthodes permettant de construire des réseaux de personnages représentant l'intrigue d'un film à partir de ses données vidéo, audio, de son script, et d'autres sources textuelles [12]. Les approches neuronales sont quant à elles appropriées au traitement des données multimodales, et aboutissent généralement à de meilleurs résultats que les approches unimodales [18]. Il existe un grand nombre de pistes à explorer pour élaborer une représentation pertinente de corpus de documents narratifs via des approches multimodales, et pour ce faire les approches neuronales modernes semblent particulièrement prometteuses. Reste la question de l'évaluation de cette qualité de représentation : les performances obtenues sur des tâches de haut niveau linguistique permettraient de déterminer si la représentation proposée est capable de capturer la composante narrative des documents, nécessaire à la réalisation de ces tâches.

1.2 Objectifs

Les objectifs principaux identifiés pour ce sujet de doctorat sont les suivants.

Le **premier objectif** est d'étudier, de mettre en oeuvre et d'adapter des techniques de TALN modernes pour l'identification d'acteurs mentionnés à l'échelle de documents ou d'ensembles de documents, et l'extraction et la caractérisation de leurs relations. Ces relations peuvent être disponibles immédiatement car décrites explicitement dans le texte, ou bien identifiées via le recoupement d'évènements distincts dans lesquels ces acteurs sont impliqués.

Le **deuxième objectif** est la définition de méthodes permettant la construction de graphes dynamiques à partir des techniques de TALN explorées précédemment. Les graphes obtenus pourront notamment être utilisés pour proposer des visualisations, mais seront principalement destinés à jouer

le rôle de représentation narrative du corpus. Une évaluation de la qualité des graphes obtenues sera mise en place, en les comparant à des graphes de références grâce à des métriques de similarité.

Le **troisième objectif** vise à utiliser les graphes obtenus dans le développement d'une représentation de documents ou de corpus de documents pertinente sur des applications spécifiques. On pourra notamment étudier l'influence des métriques sur les graphes précédemment explorées dans le cadre de ces applications. Plusieurs applications sont envisagées, notamment la classification de texte, la segmentation de textes et la recommandation.

1.3 Méthode

Tâches de TALN La première étape du traitement vise à résoudre les trois tâches de TALN que nous avons identifiées comme constituant le pré-traitement nécessaire à l'extraction des réseaux d'interaction. Tout d'abord, il faut identifier les entités d'intérêt apparaissant dans le texte, une tâche appelée *Reconnaissance d'Entités Nommées*. Exprimée simplement, il s'agit de repérer les noms propres correspondant à des personnes, lieux, objets, etc. La deuxième tâche est appelée *Résolution de Co-références* et a pour but d'identifier dans le texte les constructions anaphoriques (notamment les pronoms et syntagmes nominaux) faisant référence aux entités détectées sans les mentionner explicitement. Enfin, la troisième tâche, qui s'appuie sur les deux premières, est l'*Extraction d'Évènements*. Son but est de déterminer *qui a fait quoi, où, quand, et avec qui, sur qui/quoi*.

La résolution de ces trois tâches pourra notamment s'appuyer sur les progrès récents du domaine déjà cités, particulièrement en ce qui concerne les approches neuronales. Il conviendra donc d'abord d'effectuer l'état de l'art des méthodes existantes. Ce travail sera réalisé non seulement d'un point de vue théorique, dans le but de comprendre les principes des méthodes proposées, mais aussi d'un point de vue pratique, en considérant leur niveau de performance.

La détection d'entités nommées est une tâche pour laquelle un travail conséquent a déjà été publié, à la fois via des méthodes classiques [8] et neuronales [25]. Dans une moindre mesure, cela vaut également pour la résolution de co-références [10]. Ces deux tâches sont toutefois loin d'être complètement résolues. Il conviendra donc d'évaluer les performances de l'état de l'art dans ces deux tâches et, si le besoin s'en fait sentir, de proposer des axes d'améliorations.

Pour la tâche d'extraction d'évènements, si des approches classiques existent [9], l'utilisation d'approches neuronales est beaucoup plus récente [23, 26], et moins explorée. De plus, la notion d'évènement peut se définir de nombreuses façons différentes, et la nature de la définition retenue affecte significativement les performances de ces méthodes. Dans le cadre de cette thèse, notre objectif est d'identifier des évènements spatiaux-temporels, permettant non seulement de mettre en relation des entités représentant des acteurs et d'identifier la nature de l'action qui les relie, mais également de contextualiser cette action en identifiant le lieu et la date à laquelle elle se produit, ainsi que, le cas échéant, les objets mis en oeuvre (par exemple deux auteurs écrivant un livre ensemble). L'une des pistes que nous envisageons est de nous appuyer sur les architectures neuronales existantes et d'en proposer des adaptations spécifiques à notre conception de ce qu'est un évènement.

Outre la résolution indépendante des trois tâches de TALN décrites plus haut, une piste de recherche serait le développement de méthodes capables de traiter les trois tâches de TALN simultanément. Ces méthodes pourront être fondées sur des approches neuronales, mais également sur des approches hybrides mêlant apprentissage profond et approches classiques de fusion par exemple.

Extraction des graphes La deuxième étape du traitement est l'extraction de graphes à partir des résultats fournis par les tâches de TALN : entités, mentions, évènements. Techniquement, plusieurs formes de graphes différents peuvent être ciblées, comme en témoigne la littérature existante sur l'extraction de réseaux de personnages fictionnels [12], un problème sur lequel l'équipe d'encadrement travaille déjà.

Une première approche consistera à transposer et appliquer des méthodes déjà développées dans ce contexte. Celles-ci permettent de détecter des relations **explicites** entre les acteurs identifiés dans le texte. Pour être bref, cela peut se faire de façon *approximative* en assimilant une co-occurrence d'acteurs (deux acteurs apparaissent au même endroit dans le texte) à une relation. Dans notre cas, l'information issue de la reconnaissance d'entités nommées et de la résolution de co-références est suffisante pour mettre en oeuvre cette approche. Une approche plus *précise*, mais

plus difficile en termes de TALN, consiste à relever une relation entre des acteurs seulement si elle est exprimée sous la forme d'une action impliquant les acteurs (un verbe a pour sujets ou objets les acteurs en question). Cette méthode peut être mise en oeuvre sur la base des événements identifiés par notre troisième tâche de TALN.

En fonction du type de document étudié, il est possible que nous ayons accès à des informations supplémentaires, notamment du point de vue de la contextualisation spatio-temporelle des événements. En effet, si l'on prend l'exemple d'un article journalistique ou d'une notice biographique, les événements sont souvent datés et/ou localisés. Il est donc possible de détecter des relations **implicites** entre les acteurs, un objectif qui n'a pour ainsi dire pas encore été exploré dans la littérature. Prenons le cas de deux événements possiblement mentionnés dans deux textes différents, et impliquant deux acteurs différents, mais dont le contexte est identique. Notre hypothèse est qu'il est possible, dans une certaine mesure, de supposer que ces deux acteurs entretiennent une relation d'une certaine forme, même si celle-ci n'est mentionnée explicitement à aucun moment dans le texte. Par exemple, deux personnes dont les biographies respectives indiquent qu'elles sont toutes les deux diplômées d'une même école et appartiennent à la même promotion se connaissent très certainement, même si leurs biographies respectives ne l'indiquent pas explicitement.

L'identification de ces relations implicites repose sur la détection d'événements réalisée à l'étape de TALN. Le défi majeur consiste ici à unifier les événements identifiés, c'est à dire à déterminer lesquels sont similaires, et dans quelle mesure, afin de relier les sommets concernés. La première méthode à mettre en oeuvre est descendante, dans le sens où l'on part d'une conception *a priori* de ce que l'on considère comme des événements similaires. Il s'agit de développer manuellement une mesure de similarité entre événements, qui nous permettra d'obtenir une performance de référence sur cette tâche. Il faudra pour cela utiliser la sémantique correspondant aux termes du texte décrivant l'action associée à l'événement, ainsi qu'à son contexte. En effet, cette sémantique nous permettra d'effectuer des inférences, et ainsi de déterminer par exemple que les termes *Bordeaux* et *CUB*¹ ne sont pas exactement identiques, mais de sens suffisamment similaires pour en conclure une proximité géographique. Ceci implique d'utiliser les grandes bases de connaissances disponibles publiquement en ligne, telles que WikiData². La seconde méthode que nous voulons explorer est ascendante, dans le sens où elle se base sur les données pour établir ce que sont deux événements similaires. Elle consiste à utiliser des approches neuronales pour apprendre automatiquement une mesure de similarité entre événements. Cette approche a l'avantage de ne pas reposer sur des pré-supposés pour définir cette similarité, cependant, elle implique l'utilisation d'un corpus annoté approprié. Ce corpus pourra être annoté manuellement, mais on pourra également explorer la possibilité d'une génération automatique d'événements permettant l'obtention d'un corpus synthétique, ou d'une approche hybride utilisant de l'augmentation de données pour augmenter la taille d'un corpus manuellement annoté de petite taille.

Une fois les graphes construits, que ce soit via l'identification de relations explicites ou implicites, il sera nécessaire d'estimer leur pertinence. Ceci implique non seulement de constituer une *vérité terrain* en extrayant manuellement des graphes similaires à titre de référence, mais également d'appliquer une méthode de comparaison de graphes. La littérature est riche sur ce point, et présente un grand nombre de méthodes fondées sur différents principes de théorie des graphes [1]. Nous comptons également explorer les approches neuronales pour traiter ce point [15], qui permettent d'apprendre la mesure de similarité à partir des données, au lieu de la définir *a priori*. Sur cet aspect, le candidat pourrait bénéficier du travail déjà réalisé dans le cadre de la thèse de Noé Cécillon sur les méthodes de plongements de graphes [2], qui se déroule actuellement au LIA sous la direction des mêmes encadrants.

Représentations et Applications En se servant des graphes obtenus précédemment, il sera possible de développer une représentation permettant la résolution de tâches de recherche d'information ou de fouille de textes. Cette représentation pourra être adaptée en fonction de la tâche à résoudre, mais il s'agira de développer un mécanisme de fusion des informations d'entrée suffisamment générique pour être utilisable dans plusieurs contextes. Ces informations d'entrée supplémentaires pouvant être de types divers (texte, vidéo, audio...), il conviendra notamment de se soucier de l'aspect multimodal de la fusion. Pour ce faire, nous envisageons d'explorer essentiellement deux pistes. La première est

1. Communauté urbaine de Bordeaux

2. <https://www.wikidata.org/>

d'exploiter séparément les méthodes de *representation learning* propres à chaque modalité, afin d'en réaliser une fusion tardive et d'aboutir à une représentation unique. La seconde, plus ambitieuse, consiste dans un premier temps à effectuer une intégration de toutes les modalités dans le graphe, en exploitant le concept de *feature-rich network* qui permet d'inclure dans le graphe des attributs décrivant individuellement ses nœuds et ses liens. Puis, dans un second temps, à proposer une méthode de *representation learning* capable de traiter ce graphe augmenté pour obtenir une représentation pertinente pour la tâche considérée. Même si d'autres tâches pourront être rajoutées en fonction de l'évolution de la thèse, nous avons *a priori* identifié trois tâches de haut niveau pour évaluer la qualité de nos représentations : la classification de documents, la segmentation de textes et la recommandation de documents.

La *classification de documents* est une tâche de TALN relativement bien étudiée, et qui a l'avantage de déboucher sur des applications intéressantes dans notre contexte d'étude. L'approche générale est d'exploiter un large corpus de documents, tel que le *Projet Gutenberg*³ pour les romans ou *IMSDb*⁴ pour les scripts de films, pour entraîner un classificateur ou un régresseur capable de prédire un aspect de ces documents à partir de leurs autres caractéristiques. Il existe de nombreuses variantes de cette tâche, comme par exemple prédire le genre ou le courant littéraire d'un roman [24], la note attribuée à un film par des spectateurs [28], ou la popularité d'une série télévisée [6]. Le fait que cette tâche soit déjà explorée sous différentes formes dans la littérature nous fournira à la fois des performances de référence (*baselines*) et des corpus directement exploitables.

La *segmentation de texte* a pour but de diviser un document en plusieurs blocs sémantiques. La nature de cette division dépend grandement du type de document considéré. Ainsi, un article journalistique pourrait être scindé en plusieurs blocs, chacun contenant le traitement d'une thématique particulière. Mais dans le cas de documents purement narratifs, on cherche plus généralement à identifier les différentes intrigues et sous-intrigues s'entremêlant pour constituer le récit. Notre hypothèse est que l'aspect temporel des graphes dynamiques extraits pourra permettre à un algorithme, par observation de l'évolution des relations entre acteurs du graphe, de réaliser une segmentation efficace. Là aussi, des travaux existent déjà pour traiter ce problème, par exemple dans des films [4], des romans [11], et des articles de Wikipedia [20], fournissant ainsi à la fois des données pour évaluer notre approche, ainsi que des performances de référence auxquelles se comparer.

La *recommandation de documents* consiste à proposer à l'utilisateur des documents qui sont susceptibles de l'intéresser. Cet intérêt est généralement déterminé sur la base des retours déjà exprimés par l'utilisateur, et les approches existantes fonctionnent souvent par analogie, en renvoyant des documents similaires à ceux déjà appréciés par l'utilisateur ou par des utilisateurs aux goûts similaires. Pour ce qui concerne les documents narratifs, la littérature est riche, avec des travaux portant sur la recommandation de films [13], livres [16], ou même bandes-dessinées [14]. Comme la classification, il s'agit d'un problème relativement exploré, notamment avec des approches neuronales [29], ce qui nous donnera de nombreux points de comparaison.

1.4 Organisation

Les principales étapes du travail de thèse proposé suivent la décomposition détaillée en Section 1.3 :

1. **Identification automatique des acteurs et de leurs mentions.** Évaluation des méthodes existantes, et développement de méthodes spécialisées.
2. **Extraction automatique d'évènements.** Mise en place et évaluation d'une première version basée sur les approches existantes, puis amélioration progressive basée sur les métriques d'évaluation.
3. **Construction de graphes d'interactions.** Mise en place d'un algorithme de construction se basant sur les informations précédemment extraites. Évaluation des graphes obtenus par comparaison avec des graphes de référence.
4. **Applications.** Utilisation des représentations obtenues dans plusieurs applications. Étude de l'influence de la qualité des graphes utilisés dans la performance sur ces applications.

3. <https://www.gutenberg.org/>

4. <https://imsdb.com/>

2 Profil du candidat ou de la candidate

La candidate ou le candidat devra posséder les compétences suivantes :

- Détenteur d'un diplôme de master en informatique ou d'ingénieur en informatique.
- Maîtrise du français et bon niveau d'anglais, à l'oral et à l'écrit.
- Maîtrise des langages de programmation C++, Java, et/ou Python.
- A suivi un cours de traitement automatique du langage naturel écrit et/ou possède une expérience pratique dans une tâche de TALN écrit : détection d'entités nommées, résolution de co-références, extraction d'évènements, etc.
- A suivi un cours sur l'apprentissage automatique profond, et/ou possède une expérience pratique mettant en oeuvre ce type de méthodes.

3 Opportunités de mobilité

Nous envisageons de mettre en place un stage doctoral sur la base d'un financement Hubert Curien STAR (Science and Technology Amical Relationship). Il s'agit de projets binationaux France-Corée du Sud, impliquant une équipe dans chacun des deux pays, et permettant à certains de leurs membres d'effectuer des missions chez l'autre. En l'occurrence, nos contacts en Corée sont O-Joun Lee (Pohang University of Science and Technology) et Jason J. Jung (Chung-Ang University). Ils travaillent sur l'extraction de réseaux de personnages dans des scripts, et leur caractérisation par des méthodes neuronales afin d'élaborer des plongements permettant de réaliser différentes tâches de recherche d'information sur ces représentation vectorielles. Leur travail et les aspects méthodologiques de cette thèse sont donc très complémentaires et susceptibles de se compléter.

4 Références bibliographiques

- [1] M. Baur et M. Benkert. "Network comparison". In : *Network Analysis : Methodological Foundations*. T. 3418. Lecture Notes in Computer Science. 2005, p. 318–340. doi : [10.1007/978-3-540-31955-9_12](https://doi.org/10.1007/978-3-540-31955-9_12).
- [2] N. Cécillon et al. "Graph embeddings for Abusive Language Detection". In : *Springer Nature Computer Science 2* (2021), p. 37. doi : [10.1007/s42979-020-00413-7](https://doi.org/10.1007/s42979-020-00413-7).
- [3] D. Cer et al. "Universal Sentence Encoder". In : *arXiv cs.CL* (2021), p. 1803.11175. url : <https://arxiv.org/abs/1803.11175>.
- [4] D. Chao et al. "Representing stories as interdependent dynamics of character activities and plots : A two-mode network relational event model". In : *Digital Scholarship in the Humanities* 34.3 (2018), p. 471–481. doi : [10.1093/l1c/fqy062](https://doi.org/10.1093/l1c/fqy062).
- [5] J. Devlin et al. In : *Conference of the North American Chapter of the Association for Computational Linguistics*. 2019, p. 4171–4186. doi : [10.18653/v1/n19-1423](https://doi.org/10.18653/v1/n19-1423).
- [6] A. Fronzetti Collado et M. Naldi. "Predicting the performance of TV series through textual and network analysis : The case of Big Bang Theory". In : *PLoS ONE* 14.11 (2019), e0225306. doi : [10.1371/journal.pone.0225306](https://doi.org/10.1371/journal.pone.0225306).
- [7] I. Goodfellow, Y. Bengio et A. Courville. *Deep learning – Adaptive computation and machine learning*. MIT Press, 2016. url : <https://www.deeplearningbook.org/>.
- [8] A. Goyal, V. Gupta et M. Kumar. "Recent Named Entity Recognition and Classification techniques : A systematic review". In : *Computer Science Review* 29 (2018), p. 21–43. doi : [10.1016/j.cosrev.2018.06.001](https://doi.org/10.1016/j.cosrev.2018.06.001).
- [9] F. Hogenboom et al. "An Overview of Event Extraction from Text". In : *Detection, Representation, and Exploitation of Events in the Semantic Web*. T. 779. CEUR Workshop Proceedings. 2011, p. 48–57. url : <http://ceur-ws.org/Vol-779/>.
- [10] M. Joshi et al. "BERT for Coreference Resolution : Baselines and Analysis". In : *Conference on Empirical Methods in Natural Language Processing / 9th International Joint Conference on Natural Language Processing*. 2019, p. 5803–5808. doi : [10.18653/v1/D19-1588](https://doi.org/10.18653/v1/D19-1588).
- [11] O. Koshorek et al. "Text Segmentation as a Supervised Learning Task". In : *Conference of the North American Chapter of the Association for Computational Linguistics*. 2018, p. 469–473. doi : [10.18653/v1/n18-2075](https://doi.org/10.18653/v1/n18-2075).
- [12] V. Labatut et X. Bost. "Extraction and Analysis of Fictional Character Networks : A Survey". In : *ACM Computing Surveys* 52.5 (2019), p. 89. doi : [10.1145/3344548](https://doi.org/10.1145/3344548).
- [13] O.-J. Lee et J. J. Jung. "Modeling affective character network for story analytics". In : *Future Generation Computer Systems* 92 (2019), p. 458–478. doi : [10.1016/j.future.2018.01.030](https://doi.org/10.1016/j.future.2018.01.030).
- [14] O.-J. Lee et J.-T. Kim. "Learning Multi-modal Representations of Narrative Multimedia : a Case Study of Webtoons". In : *ACM International Conference on Research in Adaptive and Convergent Systems*. 2020, p. 1–6. doi : [10.1145/3400286.3418216](https://doi.org/10.1145/3400286.3418216).
- [15] G. Ma et al. "Deep Graph Similarity Learning : A Survey". In : *arXiv cs.LG* (2019), p. 1912.11615. url : <https://arxiv.org/abs/1912.11615>.
- [16] O. Mathew, B. Kuriakose et V. Hegde. "Book Recommendation System through content based and collaborative filtering method". In : *International Conference on Data Mining and Advanced Computing*. 2016. doi : [10.1109/sapience.2016.7684166](https://doi.org/10.1109/sapience.2016.7684166).
- [17] Tomas Mikolov et al. "Efficient Estimation of Word Representations in Vector Space". In : *ICLR*. 2013.
- [18] J. Ngiam et al. "Multimodal Deep Learning". In : *28th International Conference on International Conference on Machine Learning*. 2011, p. 689–696. url : <https://dl.acm.org/doi/10.5555/3104482.3104569>.

- [19] S. J. Pan et Q. Yang. "A Survey on Transfer Learning". In : *IEEE Transactions on Knowledge and Data Engineering* 22.10 (2010), p. 1345–1359. doi : [10.1109/tkde.2009.191](https://doi.org/10.1109/tkde.2009.191).
- [20] C. Pethe, A. Kim et S. Skiena. "Chapter Captor : Text Segmentation in Novels". In : *Conference on Empirical Methods in Natural Language Processing*. 2020, p. 8373–8383. doi : [10.18653/v1/2020.emnlp-main.672](https://doi.org/10.18653/v1/2020.emnlp-main.672).
- [21] N. Reimers et I. Gurevych. "Sentence-BERT : Sentence Embeddings using Siamese BERT-Networks". In : *Conference on Empirical Methods in Natural Language Processing / 9th International Joint Conference on Natural Language Processing*. 2019, p. 3982–3992. doi : [10.18653/v1/d19-1410](https://doi.org/10.18653/v1/d19-1410).
- [22] A. Vaswani et al. "Attention is All you Need". In : *31st Conference on Neural Information Processing Systems*. 2017. url : <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.
- [23] D. Wadden et al. "Entity, Relation, and Event Extraction with Contextualized Span Representations". In : *Conference on Empirical Methods in Natural Language Processing / 9th International Joint Conference on Natural Language Processing*. 2019, p. 5784–5789. doi : [10.18653/v1/d19-1585](https://doi.org/10.18653/v1/d19-1585).
- [24] J. Worsham et J. Kalita. "Genre Identification and the Compositional Effect of Genre in Literature". In : *27th International Conference on Computational Linguistics*. 2018, p. 1963–1973. url : <https://www.aclweb.org/anthology/C18-1167/>.
- [25] I. Yamada et al. "LUKE : Deep Contextualized Entity Representations with Entity-aware Self-attention". In : *Conference on Empirical Methods in Natural Language Processing*. 2020, p. 6442–6454. doi : [10.18653/v1/2020.emnlp-main.523](https://doi.org/10.18653/v1/2020.emnlp-main.523).
- [26] S. Yang et al. "Exploring Pre-trained Language Models for Event Extraction and Generation". In : *57th Annual Meeting of the Association for Computational Linguistics*. 2019, p. 5284–5294. doi : [10.18653/v1/p19-1522](https://doi.org/10.18653/v1/p19-1522).
- [27] Z. Yang et al. "XLNet : Generalized Autoregressive Pretraining for Language Understanding". In : *33rd Conference on Neural Information Processing Systems*. 2019. url : <https://papers.nips.cc/paper/2019>.
- [28] Y. Zeng. "Prediction of Film Score : Based on Character Relations". In : *International Conference on Humanities, Cultures, Arts and Design*. 2019, p. 26–34. doi : [10.25236/ICHCAD.2019.006](https://doi.org/10.25236/ICHCAD.2019.006).
- [29] S. Zhang et al. "Deep Learning Based Recommender System : A Survey and New Perspectives". In : *ACM Computing Surveys* 52 (2019), p. 5. doi : [10.1145/3285029](https://doi.org/10.1145/3285029).