

La campagne d'évaluation JOKER à CLEF 2024¹

Analyse Automatique de l'humour

Liana Ermakova¹, Anne-Gwenn Bosser², Tristan Miller^{3,4}, Tremaine Thomas¹,
Victor Manuel Palma Preciado^{1,5}, Grigori Sidorov⁵ and Adam Jatowt⁶

¹Université de Bretagne Occidentale, HCTI, France

²École Nationale d'Ingénieurs de Brest, Lab-STICC CNRS UMR 6285, France

³Département d'informatique, Université du Manitoba, Canada

⁴Institut autrichien de recherche en intelligence artificielle (OFAI), Vienne, Autriche

⁵Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), Mexico City, Mexique

⁶Université d'Innsbruck, Autriche

Abstract

La campagne JOKER de la Conférence et des Laboratoires du Forum d'Évaluation (CLEF) vise à encourager la recherche sur le traitement automatisé de l'humour verbal, y compris des tâches telles que la recherche, la classification, l'interprétation, la génération et la traduction. Malgré le succès retentissant des grands modèles linguistiques, le traitement automatique de l'humour et des jeux de mots est loin d'être un problème résolu. JOKER rassemble des experts en sciences sociales et informatiques et les encourage à collaborer sur des tâches communes à l'aide d'ensembles de données annotées dont la qualité est contrôlée. En 2024, nous proposerons des tâches partagées entièrement nouvelles sur la recherche d'informations basée sur l'humour, ainsi que sur l'analyse fine des sentiments et la classification de l'humour pour les agents conversationnels. Comme lors des précédentes éditions, nous mettrons également nos données à disposition pour une tâche non partagée qui sollicite de nouveaux cas d'utilisation. Dans cet article, nous présentons une brève rétrospective des campagnes JOKER, en mettant l'accent sur les résultats et les enseignements tirés de l'itération de l'année dernière, et nous donnons un aperçu des tâches qui seront organisées lors de JOKER 2024.

Keywords


Jeu de mots, Humour, Recherche d'informations, Génération d'humour, Classification de l'humour, Analyse des sentiments,

1. Introduction

L'analyse et le traitement automatisés de l'humour restent un défi pour les systèmes d'intelligence artificielle (IA) et de traitement automatique du langage naturel (TALN), notamment en raison de sa dépendance à l'égard d'indices contextuels nuancés, d'ambiguïtés lexicales et syntaxiques, de références culturelles implicites et de l'absence de modèles ou de règles linguistiques. Pour la troisième année consécutive, la campagne JOKER¹ de la Conférence

¹Cette soumission est le résumé traduit d'un article accepté à la conférence ECIR'24 [1].

CORIA (CONférence en Recherche d'Information et Applications) les 3 et 4 avril 2024 à La Rochelle

 0000-0002-7598-7474 (L. Ermakova); 0000-0002-0442-2660 (A. Bosser); 0000-0002-0749-1100 (T. Miller); 0009-0004-4210-1931 (T. Thomas); 0000-0001-8711-1106 (V.M. Palma Preciado); 0000-0003-3901-3522 (G. Sidorov); 0000-0001-7235-0665 (A. Jatowt)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://www.joker-project.com/>

et des Laboratoires du Forum d'Évaluation (CLEF) vise à encourager les travaux sur ce sujet difficile et aux multiples facettes. Lors de la première édition de JOKER, qui s'est tenue à CLEF 2022, nous avons présenté des tâches partagées pilotes axées sur la catégorisation, l'interprétation et la traduction de jeux de mots, de calembours et de néologismes humoristiques, en anglais et en français. Dans l'édition 2023 (13 équipes et 176 évaluations), nous nous sommes concentrés sur la détection, la localisation et l'interprétation des jeux de mots en anglais, en français et en espagnol [2, 3], ainsi que sur la traduction automatique des jeux de mots de l'anglais vers le français et de l'anglais vers l'espagnol [4, 5, 6]. Pour JOKER 2024, nous conservons la tâche de traduction de jeux de mots et nous élargissons le champ des tâches en abordant l'humour en général plutôt que les jeux de mots en particulier. Nous étendons également le corpus que nous avons précédemment construit pour la détection des jeux de mots [7, 4] afin de permettre une nouvelle tâche sur la recherche d'informations basée sur l'humour. Enfin, nous prévoyons de proposer des tâches entièrement nouvelles sur l'analyse fine du sentiment de l'humour ainsi que sur sa classification automatique, conçues pour faire progresser l'analyse et la génération d'humour dans les agents conversationnels. Malgré les récentes avancées significatives des LLM, le déploiement de la production d'humour généré dans les agents conversationnels reste risqué car ces modèles d'IA peuvent révéler des stéréotypes indésirables ou produire des blagues qui peuvent être perçues comme offensantes ou inappropriées, l'humour étant souvent lié à des tabous sociaux [8]. D'autre part, la classification de l'humour est utile pour fournir une réponse appropriée à une demande ludique dans les systèmes de dialogue [9].

Les quatre tâches communes de JOKER 2024 peuvent être résumées comme suit : (1) Recherche d'informations humoristiques; (2) Analyse fine des sentiments dans les textes humoristiques courts; (3) Classification de l'humour en fonction du genre et de la technique; (4) Traduction de jeux de mots de l'anglais vers le français.

2. Tâches de JOKER 2024

Tâche 1 : Recherche d'informations humoristiques. Afin d'encourager la recherche d'informations sur l'humour, nous introduisons une nouvelle tâche qui vise à extraire de courts textes humoristiques d'une collection de documents. Pour construire les requêtes, nous utiliserons les emplacements des jeux de mots de la tâche 2 [3] de JOKER 2023. La collection de documents sera une extension de la collection utilisée pour les tâches de JOKER 2023 sur la détection des jeux de mots en anglais et en français [2]. Nous augmenterons les données en introduisant des passages de textes extraits de sources humoristiques et non humoristiques, ainsi que des données générées sur des sujets pertinents pour les requêtes. Nous utiliserons des mesures standard de recherche d'information pour évaluer les systèmes des participants pour cette tâche, telles que la MAP, la NDCG, la précision et le rappel.

Tâche 2 : Analyse fine des sentiments dans les textes humoristiques courts. Notre tâche 2 implique un examen nuancé du sentiment de courts textes humoristiques, visant à identifier et à classer diverses nuances émotionnelles, y compris les sentiments positifs, négatifs et neutres. Le défi consiste à aller au-delà de la classification traditionnelle des polarités à deux ou trois voies et à distinguer les variations subtiles du sentiment, ce qui permet de mieux comprendre

les subtilités émotionnelles de l'humour. Nous demanderons également aux systèmes d'étiqueter les blagues qui propagent des généralisations négatives impliquant des constructions sociales. Nous collecterons de courts textes humoristiques, y compris des blagues, des jeux de mots et des répliques provenant d'une gamme variée de corpus humoristiques accessibles au public ainsi que du sous-corpus de détection de jeux de mots JOKER en anglais. L'ensemble de données est conçu pour englober différentes formes d'humour, de styles linguistiques et de contenus thématiques. Les systèmes seront évalués à l'aide des mesures traditionnelles de classification multi-label en RI (précision, rappel, F_1 et exactitude).

Tâche 3 : Classification de l'humour. La classification de l'humour est une tâche importante dans les systèmes de dialogue car elle peut être utilisée pour fournir une réponse appropriée à une demande ludique [9]. Ainsi, la classification de l'humour de la tâche 3 est une classification dont l'objectif est d'identifier dans un texte cible la technique particulière utilisée pour générer de l'humour. Comme pour la tâche 2, les exécutions de cette tâche seront évaluées selon des mesures standard de classification. Les données pour cette tâche seront un mélange de corpus existants sur la détection de l'ironie et du sarcasme [10, 11] et sur COVID-19 l'humour [12], ainsi que des blagues récupérées sur des sites d'humour publics selon les catégories prédéfinies sélectionnées de manière équilibrée.

Tâche 4 : Traduction des jeux de mots. L'objectif de cette tâche est de traduire des blagues anglaises en français. Les traductions doivent s'efforcer de préserver, dans la mesure du possible, la forme et le sens du jeu de mots original [13, 4]. Nous fournirons une collection de traductions anglaises-françaises de blagues sur les jeux de mots mise à jour. Nous poursuivrons la pratique consistant à demander à des experts formés d'évaluer manuellement les traductions du système en fonction de caractéristiques telles que la préservation du champ lexical, la préservation du sens, la préservation de la forme du jeu de mots, le changement de style, le changement d'humour, etc. et la présence d'erreurs dans la syntaxe, le choix des mots, etc. Nous expérimenterons également d'autres mesures semi-automatiques.

3. Conclusion

Nous avons décrit la configuration prospective de la campagne d'évaluation JOKER de CLEF 2024, qui comprend des tâches partagées sur la recherche d'information basée sur l'humour et deux tâches potentiellement utiles pour la génération d'humour dans les systèmes de dialogue, à savoir l'analyse fine du sentiment de textes humoristiques courts et la classification de l'humour. Les participants sont invités à visiter le site JOKER (<https://joker-project.com>) pour plus de détails sur la campagne. Les inscriptions à l'atelier à CLEF sont ouvertes jusqu'au 22 Avril 2024, les soumissions de résultats sont dues le 6 Mai et les rapports techniques le 31 Mai.

Remerciements

JOKER a reçu une subvention gouvernementale de l'Agence Nationale de la Recherche dans le cadre du programme "Investissements d'avenir" intégré à France 2030, de référence ANR-19-GURE-0001. JOKER est soutenu par la MSHB. Ce projet n'aurait pas été possible sans le soutien de nombreuses personnes ; nous tenons à remercier en particulier les étudiants de l'Université de Bretagne Occidentale pour leur contribution à la construction des données.

References

- [1] L. Ermakova, T. Miller, A.-G. Bosser, V. M. Palma Preciado, T. Thomas, G. Sidorov, A. Jatowt, Clef 2024 joker lab: Automatic humour analysis, in: N. Goharian, N. Tonello, Y. He, A. Lipani, G. McDonald, C. Macdonald, I. Ounis (Eds.), ECIR'24: Advances in Information Retrieval, 46th European Conference on Information Retrieval, Lecture Notes in Computer Science, Springer, 2024.
- [2] L. Ermakova, T. Miller, A.-G. Bosser, V. M. Palma Preciado, G. Sidorov, A. Jatowt, Overview of JOKER 2023 Automatic Wordplay Analysis Task 1 – pun detection, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 – Conference and Labs of the Evaluation Forum, volume 3497 of *CEUR Workshop Proceedings*, 2023, pp. 1785–1803.
- [3] L. Ermakova, T. Miller, A.-G. Bosser, V. M. Palma Preciado, G. Sidorov, A. Jatowt, Overview of JOKER 2023 Automatic Wordplay Analysis Task 2 – pun location and interpretation, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 – Conference and Labs of the Evaluation Forum, volume 3497 of *CEUR Workshop Proceedings*, 2023, pp. 1804–1817.
- [4] L. Ermakova, T. Miller, A.-G. Bosser, V. M. Palma Preciado, G. Sidorov, A. Jatowt, Overview of JOKER – CLEF-2023 track on automatic wordplay analysis, in: A. Arampatzis, E. Kanoulas, T. Tsikrika, S. Vrochidis, A. Giachanou, D. Li, M. Aliannejadi, M. Vlachos, G. Faggioli, N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, volume 14163, Springer Nature Switzerland, Cham, 2023, pp. 397–415. doi:10.1007/978-3-031-42448-9_26.
- [5] L. Ermakova, T. Miller, A.-G. Bosser, V. M. Palma Preciado, G. Sidorov, A. Jatowt, Overview of JOKER 2023 Automatic Wordplay Analysis Task 3 – pun translation, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 – Conference and Labs of the Evaluation Forum, volume 3497 of *CEUR Workshop Proceedings*, 2023, pp. 1818–1827.
- [6] L. Ermakova, T. Miller, A.-G. Bosser, V. M. Palma Preciado, G. Sidorov, A. Jatowt, Science for fun: The CLEF 2023 JOKER track on automatic wordplay analysis, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), *Advances in Information Retrieval: 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2–6, Proceedings, Part III*, volume 13982 of *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2023, pp. 546–556. doi:10.1007/978-3-031-28241-6_63.
- [7] L. Ermakova, A.-G. Bosser, A. Jatowt, T. Miller, The JOKER Corpus: English–French parallel data for multilingual wordplay recognition, in: SIGIR '23: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, New York, NY, 2023, pp. 2796–2806. doi:10.1145/3539618.3591885.
- [8] P. P. Liang, C. Wu, L.-P. Morency, R. Salakhutdinov, Towards understanding and mitigating social biases in language models, in: M. Meila, T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2023, pp. 1011–1020.

Learning Research, PMLR, 2021, pp. 6565–6576. URL: <https://proceedings.mlr.press/v139/liang21a.html>.

- [9] N. Shapira, O. Kalinsky, A. Libov, C. Shani, S. Tolmach, Evaluating humorous response generation to playful shopping requests, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), *Advances in Information Retrieval*, volume 13981, Springer Nature Switzerland, Cham, 2023, pp. 617–626. doi:10.1007/978-3-031-28238-6_53.
- [10] I. Abu Farha, S. V. Oprea, S. Wilson, W. Magdy, SemEval-2022 Task 6: iSarcasmEval, intended sarcasm detection in English and Arabic, in: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, Association for Computational Linguistics, 2022, pp. 802–814. doi:10.18653/v1/2022.semeval-1.111.
- [11] S. Frenda, A. Pedrani, V. Basile, S. M. Lo, A. T. Cignarella, R. Panizzon, C. Marco, B. Scarlina, V. Patti, C. Bosco, D. Bernardi, EPIC: Multi-perspective annotation of a corpus of irony, in: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, volume 1, Association for Computational Linguistics, 2023, pp. 13844–13857. doi:10.18653/v1/2023.acl-long.774.
- [12] N. R. Bogireddy, S. Suresh, S. Rai, I’m out of breath from laughing! I think? A dataset of COVID-19 humor and its toxic variants, in: *Companion Proceedings of the ACM Web Conference 2023*, Association for Computing Machinery, New York, NY, 2023, pp. 1004–1013. doi:10.1145/3543873.3587591.
- [13] L. Ermakova, T. Miller, F. Regattin, A.-G. Bossler, E. Mathurin, G. L. Corre, S. Araújo, J. Boccou, A. Digue, A. Damoy, B. Jeanjean, Overview of JOKER@CLEF 2022: Automatic wordplay and humour translation workshop, in: A. Barrón-Cedeño, G. Da San Martino, M. Degli Esposti, F. Sebastiani, C. Macdonald, G. Pasi, A. Hanbury, M. Potthast, G. Faggioli, N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Thirteenth International Conference of the CLEF Association (CLEF 2022)*, volume 13390 of *Lecture Notes in Computer Science*, 2022, pp. 447–469.