

Report on the 19th French Conference on Information Retrieval and Applications (CORIA-2024)

Liana Ermakova

HCTI, Université de Bretagne Occidentale

France

`liana.ermakova@univ-brest.fr`

Philippe Mulhem

Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG

France

Antoine Doucet

La Rochelle Université

France

Carlos-Emiliano Gonzalez-Gallardo

La Rochelle Université

France

Sébastien Fournier

Aix Marseille Université, CNRS, LIS

France

Abstract

We present here a short report on the French Information Retrieval Conference CORIA 2024, held in La Rochelle, France, on the 3rd and 4th of April 2024. We describe the objectives of the CORIA and provide an overview of the event, the submissions, and the keynotes.

Date: 3–4 April 2024.

Website: <http://coria.asso-aria.org/2024/>.

1 Introduction

Founded in 2004 by the French Association for Information Retrieval and Applications ARIA (Association Francophone de Recherche d'Information et Applications), the French Information Retrieval Conference CORIA (Conférence en Recherche d'Information et Applications) is the main francophone conference in Information and Application Research.

The 19th edition was held on 3 and 4 April 2024 alongside the 17th RJCRI meeting (Rencontre des Jeunes Chercheur.euse.s en RI) by the Images et Contenus du Laboratoire Informatique, Image, Interaction (L3i) team from La Rochelle Université. Overall, the organization and program committees of CORIA 2024 were composed of 30 researchers from the French community.

CORIA aims to bring together teams and individuals carrying out scientific work in the field of information search and its applications: information search on the web, social networks or specific

collections, recommendation systems, document, image, audio and video mining, personal assistants and chatbots, etc. Now a daily activity for the general public, information search is essential to many uses of digital technology. The associated scientific and technological activity continues to grow, interacting with other areas of computer science and other disciplines - mathematics, linguistics, cognitive science - as well as with industry and stakeholders in the internet, media, culture, health, and education. Recent models include machine learning, data mining, natural language processing, speech and signal processing, image analysis, and affective computing.

The CORIA conference is open to the entire international scientific community concerned with information search from both a theoretical and applications point of view. CORIA's target audience is academic researchers, including master's and doctoral students, industrialists, and all specialists in the field. All CORIA publications are published in open access on the ARIA website and are indexed by DBLP.

2 Submissions

One role of a national conference is to allow interactions and discussions among the national research community. The outcomes may be collaborations and joint project submissions. Such national conferences must be organized in a way to strengthen the community.

CORIA suffers from several years of difficulty in attracting submissions. The reasons for that are multiple, but the major one is that national conferences are not considered in the recruitment and evaluation process of researchers. Then, it is very hard to convince the French community to spend time submitting its work. After many discussions within the French Association for Information Retrieval and Applications ARIA, in 2024 the decision was made to allow, in addition to traditional short and long papers, *summary abstracts* submissions. Such abstracts (2 pages + references) can be a summarized translation of a previously published paper, a report on negative results, a position paper, or a description of an ongoing research project. In fact, the aim of the conference is to encourage as many meetings as possible between researchers in order to increase collaboration and maximize interactions. Abstracts of already published articles were not anonymized and evaluated by two members of the program committee. The original article must be indicated to avoid any ambiguity. Other submissions were evaluated by at least three reviewers.

With these choices, we were able to get, overall, 31 submissions : 4 long papers (13%), 9 short papers (29%), and 18 abstracts (58%). The authors are mainly from France, but a few of them were from abroad: Canada, Mexico, and Singapore. Regarding the number of submissions and the large proportion of abstracts, the choice of allowing such abstracts clearly boosted the amount of submissions this year. For CORIA 2025, such abstract paper submissions will also be proposed.

The submissions were organized in the following sessions: neural information retrieval, trustworthiness and LLMs in IR, evaluation, information extraction for IR, domain-specific applications, and interactive and multimedia IR. Consistently with what is happening in the main international conferences in IR, the majority of the sessions did contain papers that used deep learning approaches, especially transformer-based ones.

3 Keynotes

Two keynote talks were presented at CORIA 2024:

- “Have a better understanding of language models and the texts they produce”, by Benoît Sagot, from Inria Research Director in Automatic Language Processing and Computational Linguistics.
- “Europeana – challenges, progress and perspectives for a European access service for cultural heritage collections”, by Antoine Isaac, R&D manager at Europeana Foundation, Netherlands.

The talk of Benoît Sagot did outline the need for a deep understanding of deep transformer-based models as they tend to produce vector representations that are not isotropically distributed in space. This anisotropy is related to the way in which these models are learned, which leads to token frequency dominating their representation negatively affecting the training of small models. Avoiding the undesirable effects of this prevalence of frequency information leads to multiple positive consequences, including on the model’s downstream performance. He also presented some results on the detectability of LLM-generated content and showed the importance of a good understanding of how language models work to enable their optimal use, including retrieval-augmented generation and machine translation.

From a very different point of view, the talk of Antoine Isaac showed how real-world applications that use Information Retrieval face problems that the academic community does not always consider. He presented the Europeana search engine which aims to make digitized cultural heritage from thousands of museums, libraries, and archives across Europe accessible to a wide audience.

Both these presentations generated many questions and discussions from the audience. Given their relevance, the audiovisual team of the Information System Department of La Rochelle Université, respecting the right of publicity of the speakers, recorded and published both talks.¹

4 Focus on Young Researchers

The CORIA conference lasted two full days, the 3rd and 4th of April 2024. As we mentioned earlier, the intent of CORIA is to gather together and to make live the community. So, one concern is about building also the community of junior IR French researchers. To achieve that, the young researchers’ party RJRCI was organized the evening before CORIA: the moment was ideal for meeting colleagues in an informal setting and bonding around games and drinks before the official start of the conference.

The CORIA conference also rewards works achieved by young researchers. Two students’ papers were awarded in 2024:

- Soëlie Lerch, Patrice Bellot, Emmanuel Bruno and Elisabeth Muriasco, “EMOLIS App et Dataset pour suggérer des dessins animés proches émotionnellement”.
- Pierre Erbacher, Jian-Yun Nie, Philippe Preux et Laure Soulier, “PAQA: Vers des systèmes ProActifs face aux Questions Ambigües”.

¹<https://videos.univ-lr.fr/sciences-technologies-sante/coria-rjcri-2024/>

5 Conclusion, Open Challenges and Suggestions for Future Editions

CORIA 2024 was a success, 55 participants had the opportunity to present, attend, and exchange with the French Information Retrieval community. In such national conferences, the challenge of receiving submissions is recurrent. As we described above, we were able to get a significant amount of papers submitted, because we allowed abstract submissions. We believe that this may be one solution for the national conferences to survive. Such possibility of abstracts seems then a good idea, however, it is only doable in non-English-language national countries. We also suggest to include more young researchers in the organization of the conferences.

We observed that overall, LLMs have significantly contributed to bringing NLP and IR closer together, resulting in an increasing number of submissions on the intersection of these areas.

Acknowledgements

This edition of CORIA and RJCRI obtained public and private funding in addition to the financial support provided by ARIA. The internal entities providing funding were la Rochelle Université, the L3i laboratory, and the IUT of la Rochelle; whereas, at a regional level, the Charente Maritime Department of France. The private enterprises that supported financially this edition of the conference were ITESOFT, SHIFT, and Yooz. We would like to thank the organizers, volunteers, authors, PC members, and all other people who made the conference a success.