Preface

The CLEF 2025 conference is the twenty-sixth edition of the popular CLEF campaign and workshop series that has run since 2000, contributing to the systematic evaluation of multilingual and multimodal information access systems, primarily through experimentation on shared tasks. In 2010, CLEF was launched in a new format, as a conference with research presentations, panels, poster and demo sessions and laboratory evaluation workshops. These are proposed and operated by groups of organizers volunteering their time and effort to define, promote, administer and run an evaluation activity.

CLEF 2025[1] was organized by the Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, from 9 to 12 September 2025. CLEF 2025 was the 16th year of the CLEF Conference and the 26th year of the CLEF initiative as a forum for Information Retrieval (IR) Evaluation. The conference format remained the same as in past years and consisted of keynotes, contributed papers, lab sessions, and poster sessions, including reports from other benchmarking initiatives from around the world. All sessions were organized in person, but also allowed for remote participation for those who were not able to attend physically.

A total of 20 lab proposals were received and evaluated in peer review based on their innovation potential and the quality of the resources created. The 14 selected labs represented scientific challenges based on new datasets and real-world problems in multimodal and multilingual information access. These datasets provide unique opportunities for scientists to explore collections, to develop solutions for these problems, to receive feedback on the performance of their solutions and to discuss the challenges with peers at the workshops. In addition to these workshops, the labs reported results of their year-long activities in overview talks and lab sessions.

We continued the mentorship program to support the preparation of lab proposals for newcomers to CLEF. The CLEF newcomers mentoring program offered help, guidance, and feedback on the writing of draft lab proposals by assigning a mentor to proponents, who helped them in preparing and maturing the lab proposal for submission.

Building on previous experience, the Labs at CLEF 2025 demonstrate the maturity of the CLEF evaluation environment by creating new tasks, new and larger data sets, new ways of evaluation or more languages. Details of the individual Labs are described by the Lab organizers in these proceedings.

The 14 labs running as part of CLEF 2025 comprised mainly labs that continued from previous editions at CLEF (BioASQ, CheckThat!, ELOQUENT, eRisk, EXIST, ImageCLEF, JOKER, LifeCLEF, LongEval, PAN, Quantum-CLEF, SimpleText, and Touché) and a new pilot/workshop activity: Talent-

---

[1] https://clef2025.clef-initiative.eu/

CLEF. In the following, we give a few details for each of the labs organized at CLEF 2025 (presented in alphabetical order):

BioASQ: Large-scale Biomedical Semantic Indexing and Question Answering[2] aims to push the research frontier towards systems that use the diverse and voluminous information available online to respond directly to the information needs of biomedical scientists. This edition of BioASQ offered the following tasks: *Task 1 (13b) – Biomedical Semantic Question Answering:* Benchmark datasets of biomedical questions, in English, along with gold standard (reference) answers constructed by a team of biomedical experts. The participants have to respond with relevant articles, and snippets from designated resources, as well as exact and "ideal" answers. *Task 2 – Synergy: Question Answering for developing problems:* Biomedical experts pose unanswered questions for developing problems, such as COVID-19, receive the responses provided by the participating systems, and provide feedback, together with updated questions in an iterative procedure that aims to facilitate the incremental understanding of developing problems in biomedicine and public health. *Task 3 – MultiClinSum: Multilingual Clinical Summarization:* a shared task on the automatic summarization of lengthy clinical case reports written in different languages. The organizers distribute lengthy clinical case reports written in English, Spanish, French, and Portuguese. The participants generate summaries of the clinical case reports. The evaluation is based on a comparison with manual summaries of the clinical case reports. *Task 4 – BioNNE-L: Nested Named Entity Linking in Russian and English:* A shared task on Natural Language Processing (NLP) challenges in entity linking, also known as medical concept normalization (MCN), for English and Russian languages. The train/dev datasets include annotated mentions of disorders, anatomical structures, and chemicals. The participants normalize the entity mentions to concept names and unique UMLS identifiers. The evaluation is based on a comparison with manual nested named entity linking annotations. *Task 5 – ElCardioCC: Clinical Coding in Cardiology:* The ELCardioCC on the automated clinical coding concerns i) the assignment of cardiology-related ICD-10 codes to discharge letters from Greek hospitals, ii) the extraction of the specific mentions of ICD-10 codes from the discharge letters. The evaluation is based on metrics, such as micro and macro F-measure for Subtask (i) and token F-measure for Subtask (ii). *Task 6 – GutBrainIE: Gut-Brain Interplay Information Extraction:* The GutBrainIE task aims to foster the development of Information Extraction (IE) systems that support experts by automatically extracting and linking knowledge from scientific literature, facilitating the understanding of gut-brain interplay and its role in neurological disease. The task is divided into two subtasks: i) extraction of named entities and linking them to concepts in a reference ontology, and ii) identifying binary relations between entity pairs.

---

[2] `https://www.bioasq.org/workshop2025`

CheckThat! Lab on Checkworthiness, Subjectivity, Persuasion, Roles, Authorities and Adversarial Robustness.[3] The eighth edition of the CheckThat! lab at CLEF presents a diverse set of challenges aimed at advancing technology to support and enhance the journalistic verification process. This edition revisits core tasks in the verification pipeline while also introducing auxiliary tasks such as subjectivity identification, claim normalization, and fact-checking numerical claims, with a particular emphasis on scientific web discourse. These tasks pose complex classification and retrieval problems at both the document level, including in multilingual contexts. The lab was organized into the following tasks: *Task 1 – Subjectivity:* Given a sentence from a news article, determine whether it is subjective or objective. This is a binary classification task and is offered in Arabic, English, Bulgarian, German, and Italian for mono- and multi-lingual settings. Additionally, unseen languages like French and Spanish are considered for zero-shot settings. *Task 2 – Claim Normalization:* Given a noisy, unstructured social media post, the task is to simplify it into a concise form. This is a generation task, offered in 20 languages: English, Arabic, Bengali, Czech, German, Greek, French, Hindi, Korean, Marathi, Indonesian, Dutch, Punjabi, Polish, Portuguese, Romanian, Spanish, Tamil, Telugu, Thai. *Task 3 – Fact-Checking Numerical Claims* This task focuses on verifying claims with numerical quantities and temporal expressions. Numerical claims are defined as those requiring validation of explicit or implicit quantitative or temporal details. Participants must classify each claim as True, False, or Conflicting based on a short list of evidence. *Task 4 – Scientific Web Discourse Processing (SciWeb)* which was further divided into two subtasks. *Subtask 4.1 – SciWeb Discourse Detection:* This task aims at classifying the different forms of science-related online discourse. Namely, given a tweet, this multilabel task aims at detecting if a tweet contains a scientific claim or scientific reference or is referring to science contexts or entities. *Subtask 4.2 – SciWeb Claim-Source Retrieval:* Given a tweet containing a scientific claim and an informal reference to a scientific paper, this task aims at retrieving the scientific paper that serves as the source for the claim from a given pool of candidate scientific papers.

ELOQUENT Lab for Evaluation of Generative Language Model Quality[4] addresses high-level quality criteria through a set of open-ended shared tasks implemented to require minimal human assessment effort. It offered the following tasks: *Task 1 – Voight-Kampff:* Generate text samples for a classifier to distinguish between human-authored and machine-generated text. *Task 2 – Robustness and Consistency:* Explore how much a generative language model's output is affected by stylistic, dialectal, or other non-topical variation in the input. *Task 3 – Preference Score Prediction:* Predict human preferences between sets of LLM-generated responses collected from human assessors, and generate judgments to explain the choice made. *Task*

---

[3] `https://checkthat.gitlab.io/clef2025/`
[4] `https://eloquent-lab.github.io/`

4 – Sensemaking: Given a set of possibly noisy texts, generate questions and answers about the topic.

eRisk: Early Risk Prediction on the Internet[5] explores the evaluation methodology, effectiveness metrics and practical applications (particularly those related to health and safety) of early risk detection on the Internet. This year's edition of eRisk included the following tasks: *Task 1 – Search for Symptoms of Depression:* Rank sentences from users according to their relevance to each of the 21 symptoms of the BDI-II questionnaire. Training data consists of sentence-tagged datasets from 2023 and 2024, with new test data including contextual information (previous and next sentences). *Task 2 – Contextualized Early Detection of Depression:* Participants analyze full conversational interactions to classify users with signs of depression, considering the conversational context beyond isolated user writings. The test phase includes writings with full conversational dynamics, while the training phase uses isolated user submissions. *Pilot Task – Conversational Depression Detection via LLMs:* Participants interact with a persona powered by a large language model (LLM) that is fine-tuned using types of depressive and non depressive users. The objective is to detect signs of depression, with participants limited to a specified number of messages to engage with the LLM.

EXIST: sEXism Identification in Social neTworks[6] aims to capture and categorize sexism, from explicit misogyny to other subtle behaviors, in social networks. In 2024 the EXIST lab included multimedia content in the format of memes, stepping forward research on more robust techniques to identify sexism in social networks. Following this line, in 2025 EXIST introduces TikTok videos in the challenge, thus including in the dataset the three most important sources of sexism spreading: text, images, and videos. Consequently, it is essential to develop automated multimodal tools capable of detecting sexism in text, images, and videos, to raise alarms or automatically remove such content from social network because platforms' algorithms often amplify content that perpetuates gender stereotypes and internalized misogyny. This lab contributes to the creation of applications that identify sexist content in social media across all three formats. EXIST 2025 was divided into three tasks, each split into three subtasks. *Task 1 – Sexism Identification and Characterization in Tweets Subtask 1.1 – Sexism Identification in Tweets:* The first subtask is a binary classification. The systems have to decide whether or not a given tweet contains or describes sexist expressions or behaviors (i.e., it is sexist itself, describes a sexist situation or criticizes a sexist behavior). *Subtask 1.2 – Source Intention in Tweets* This subtask aims to categorize the sexist messages according to the intention of the author in one of the following categories: (i) direct sexist message, (ii) reported sexist message and (iii) judgemental message. *Subtask 1.3 – Sexism Categorization in Tweets* The third subtask is a multiclass task that aims to categorize

---

[5] https://erisk.irlab.org/
[6] https://nlp.uned.es/exist2025/

the sexist messages according to the type or types of sexism they contain (according to the categorization proposed by experts and that takes into account the different facets of women that are undermined): (i) ideological and inequality, (ii) stereotyping and dominance, (iii) objectification, (iv) sexual violence and (v) misogyny and non-sexual violence. *Task 2 – Sexism Identification and Characterization in Memes Subtask 2.1 – Sexism Identification in Memes:* Similar to Subtask 1.1, Subtask 2.1 is a binary classification task where participants must determine when a meme contains or describes sexist expressions or behaviors (i.e., it is sexist itself, describes a sexist situation or criticizes a sexist behavior). *Subtask 2.2 – Source Intention in Tweets:* This subtask aims to categorize the sexist messages according to the intention of the author in one of the following categories: (i) direct sexist message, (ii) judgmental message. *Subtask 2.3 – Sexism Categorization in Memes:* Finally, this subtask addresses the problem of categorizing a sexist meme according to the type of sexism that it encloses: (i) ideological and inequality, (ii) stereotyping and dominance, (iii) objectification, (iv) sexual violence and (v) misogyny and non-sexual violence. *Task 3 - Sexism Identification and Characterization in TikTok Videos Subtask 3.1 – Sexism Identification in Videos:* Similar to Subtasks 1.1 and 2.1, this subtask is a binary classification task where participants must determine when a meme contains or describes sexist expressions or behaviors (i.e., it is sexist itself, describes a sexist situation or criticizes a sexist behavior). *Subtask 3.2 – Source Intention in Videos:* This subtask aims to categorize the sexist messages according to the intention of the author in one of the following categories: (i) direct sexist message, (ii) judgmental message. *Subtask 3.3 – Sexism Categorization in Videos:* Finally, this subtask addresses the problem of categorizing a sexist meme according to the type of sexism that it encloses: (i) ideological and inequality, (ii) stereotyping and dominance, (iii) objectification, (iv) sexual violence and (v) misogyny and non-sexual violence.

ImageCLEF: Multimodal Challenge in CLEF[7] focuses on evaluating technologies for annotating, indexing, classifying, retrieving and generating multimodal data, providing access to large datasets across a variety of scenarios, including medical, social media, and internet-based applications. Building on the success of recent editions, it encourages interdisciplinary methods by engaging participants in diverse domains, providing large amounts of challenging multimodal data and providing am evaluation platform for a large number of use cases. This year's edition of ImageCLEF involved the following tasks: *Task 1 – ImageCLEFmedical:* In its 21st edition, the task will continue all the medical sub-tasks from from the last 2 years, namely: (i) the Caption task with medical concept detection and caption prediction, (ii) the GAN task focused on synthetic medical images, (iii) MEDVQA regarding Visual Question Answering for gastrointestinal data, and (iv) MEDIQA-MAGIC, introducing a new use-case on multimodal dermatology response generation. *Task 2 – Image Retrieval/Generation for Arguments:* As a joint task between

---

[7] https://www.imageclef.org/2025

Touché and ImageCLEF since 2022, the task aims to show the impact of images in arguments, making them more compelling. In this year's task, participants shall find suitable images that convey a given argument. Two submission styles are possible, either as a retrieval task or as prompt generation for an image generator. *Task 3 – ImageCLEFtoPicto:* The aim of this task to convert either speech or text into a meaningful sequence of pictograms, aiding communication for people with language impairments, enhancing user understanding or helping with translation. Therefore, 2 sub-tasks are derived from this: (i) Text-to-Picto, involving generating pictograms starting from a French text and (ii) Speech-to-Picto, which focuses on translating speech to pictograms directly. *Task 4 – MultimodalReason:* is a new task, focusing on Multilingual Visual Question Answering. Participants are given multiple-choice questions and corresponding images and are asked to identify the correct answer, in multiple languages, disciplines and difficulty levels. The task aims to assess the reasoning abilities of modern LLMs across a wide range of real-world situations.

JOKER: Automatic Humour Analysis[8] aims to foster interdisciplinary approaches to the (semi-)automatic analysis and processing of humor and wordplay. *Task 1 – Humor-aware Information Retrieval:* For Task 1, the aim is to retrieve short humorous texts from a document collection based on a given query. The languages are English and Portuguese. *Task 2 – Wordplay Translation:* For Task 2, the goal is to translate English punning jokes into French. *Task 3 – Onomastic Wordplay:* For Task 3, the goal is to classify proper names according to whether they are humorous, and to translate them from English into French. *Task 4 – Controlled Creativity:* For Task 4, the goal is to identify the introduction of distorted or spurious content ("hallucinations") in generated creative texts.

LifeCLEF: Challenges on Species Presence Prediction and Identification, and Individual Animal Identification[9] focuses on advancing Artificial Intelligence (AI)-driven solutions for biodiversity monitoring through challenges on species and individuals recognition and prediction. *Task 1 – AnimalCLEF:* Multi-species individual animal identification. *Task 2 – BirdCLEF:* Bird species identification in soundscape recordings. *Task 3 – FungiCLEF:* Few-shot classification with rare fungi species. *Task 4 – GeoLifeCLEF:* Multi-modal species prediction using remote sensing and large-scale biodiversity data. *Task 5 – PlantCLEF:* Multi-species plant identification in vegetation plot images.

LongEval: Longitudinal Evaluation of Model Performance[10] aims to ignite the development of Information Retrieval systems that can handle temporal data evolution. The retrieval systems evaluated in this task are expected to be persistent in their retrieval efficiency over time, as Web documents and Web queries evolve. To evaluate such features of systems, we

---

[8] http://joker-project.com/
[9] https://www.imageclef.org/LifeCLEF2025
[10] https://clef-longeval.github.io/

rely on collections of documents and queries, corresponding to real data acquired from actual Web search engines. LongEval 2025 included two tasks: *Task 1 – WebRetrieval:* this task uses evolving Web data to evaluate IR system longitudinally, namely, it will assess whether the IR system performance is persistent over time. *Task 2 – SciRetrieval:* Similar to Task 1, this task aims to examine how IR systems' effectiveness changes over time, when the underlying document collection changes, where the documents are scientific publications.

PAN: Lab on Stylometry and Digital Text Forensics[11] aims to advance the state of the art and provide for an objective evaluation on newly developed benchmark datasets in those areas. The tasks proposed by PAN Lab this year included: *Task 1 – Generated Content Analysis:* Given a document, decide if it was written by a human, an AI, or both. *Task 2 – Multilingual Text Detoxification:* Given a toxic piece of text, re-write it in a non-toxic way while saving the main content as much as possible. *Task 3 – Multi-author Writing Style Analysis:* Given a document, determine at which positions the author changes. *Task 4 – Generated Plagiarism Detection:* Given a generated and a human-written source document, identify the passages of reused text between them.

QuantumCLEF: Quantum Computing at CLEF.[12] The second edition of the QuantumCLEF lab is composed of three tasks and aims at: Discovering and evaluating Quantum Annealing approaches compared to their traditional counterpart; Identifying new ways of formulating Information Retrieval and Recommender Systems algorithms and methods, so that they can be solved with Quantum Annealing; Establishing collaborations among researchers from different fields to harness their knowledge and skills to solve the considered challenges and promote the usage of Quantum Annealing. This lab allows participants to use real quantum computers provided by CINECA, one of the most important computing centers worldwide. *Task 1 – Feature Selection:* focuses on formulating the well-known NP-Hard Feature Selection problem and solving it with quantum annealers. Feature Selection is a widespread problem for both Information Retrieval and Recommender systems which requires to identify a subset of the available features (e.g., the most informative, less noisy, etc.) to train a learning model. This problem is very impacting since many of these systems involve the optimization of learning models, and reducing the dimensionality and noise of the input data can improve their performance. *Task 2 – Instance Selection:* focuses on formulating the Instance Selection problem to solve it through Quantum Annealing. Currently, transformer-based architectures, including 1st and 2nd generation transformers (e.g., RoBERTa) as well as current large language models (e.g., Llama3), are used and considered state-of-the-art in several fields. Given the LLMs high-cost application, one of the big challenges is to fine-tune these models efficiently. Instance Selection focuses on selecting

---

[11] http://pan.webis.de/

[12] https://qclef.dei.unipd.it/

a representative subset of instances from a dataset to make the training of these models faster while maintaining a high level of effectiveness of the trained model. *Task 3 – Clustering:* focuses on the formulation of the clustering problem to solve it with a quantum annealer. Clustering is a relevant problem for Information Retrieval and Recommender systems which involves grouping items together according to their characteristics. Clustering can be helpful for organizing large collections, helping users to explore a collection and providing similar results to a query. It can also be used to divide users according to their interests or build user models with the cluster centroids boosting efficiency or effectiveness for users with limited data.

SimpleText: Simplify Scientific Text (and Nothing More)[13] aims at improving accessibility to scientific information for everyone, developing corpora, evaluation measures, and new IR/NLP models able to reduce scientific text complexity with strict faithfulness to the original text. *Task 1 – Text Simplification: simplify scientific text:* aims to simplify scientific text, using aligned biomedical abstracts and lay summaries for sentence-level, paragraph-level, and document-level text simplification. *Task 2 – Controlled Creativity: identify and avoid hallucination:* aims to identify and avoid hallucination, by either post-hoc detection on CLEF submissions with over-generation, or by avoiding creative license of models by design. *Task 3 – SimpleText 2024 Revisited: selected tasks by popular request:* aims to rerun selected tasks by popular request, on scientific passage retrieval and complex terminology detection, and on tracking the state-of-the-art in scholarly papers.

TalentCLEF: Skill and Job Title Intelligence for Human Capital Management[14] aims to drive technological advancement in Human Capital Management by establishing a public benchmark for NLP models that facilitates their application in real-world Human Resources (HR) scenarios, incorporating evaluation criteria including multilingualism, fairness, and cross-industry adaptability. The lab also seeks to build a community for researchers and practitioners to generate, evaluate, and discuss ideas on the use of AI in Human Resources, pushing the state-of-the-art of NLP applications for Human Resources. *Task 1 – Multilingual Job Title Matching:* involves the development of systems that can identify and rank job titles most similar to a given one. For each job title in a provided test set, participants must generate a ranked list of similar job titles from a specified knowledge base. The task includes multilingual and cross-lingual tracks, requiring participants to develop systems adapted to English, Spanish, German, and optionally Chinese. *Task 2 – Job Title-Based Skill Prediction:* involves developing systems capable of retrieving relevant skills associated with a given job title. Participants must train models that can retrieve a list of relevant skills from a provided knowledge base, ranking them according to their relevance to the job title. This task is in English.

---

[13] http://simpletext-project.com/
[14] https://talentclef.github.io/talentclef/

Touché: Argumentation Systems[15] focuses on computational argumentation and causality. Touché 2025 included 4 tasks. *Task 1 – Retrieval-Augmented Debating:* it served to develop generative retrieval systems that argue against their users to support users in forming or confirming opinions or to train their debating skills. *Task 2 – Ideology and Power Identification in Parliamentary Debates:* it concerned with predicting ideology and power in the parliamentary debates on a multi-lingual, multi-country dataset. *Task 3 – Image Retrieval/Generation for Arguments (Joint task with ImageCLEF):* aimed to find images that support a particular point of view. *Task 4 – Advertisement in Retrieval-Augmented Generation:* analyzed possibilities and counter-measures for advertisements in retrieval-augmented search results.

CLEF has always been backed by European projects that complement the incredible amount of volunteering work performed by Lab Organizers and the CLEF community with the resources needed for its necessary central coordination, in a similar manner to the other major international evaluation initiatives such as TREC, NTCIR, FIRE, and MediaEval. Since 2014, the organization of CLEF no longer has direct support from European projects and are working to transform itself into a self-sustainable activity. This is being made possible thanks to the establishment of the CLEF Association,[16] a non-profit legal entity in late 2013, which, through the support of its members, ensures the resources needed to smoothly run and coordinate CLEF.

Acknowledgments

---

[15] https://touche.webis.de/
[16] https://www.clef-initiative.eu/#association

Jaap Kamps, University of Amsterdam, The Netherlands
Evangelos Kanoulas, University of Amsterdam, The Netherlands
Jussi Karlgren, Silo AI, Finland
Johannes Kiesel, Bauhaus-Universität Weimar, Germany
David Losada, University of Santiago de Compostela, Spain
Maria Maistro, University of Copenhagen, Denmark
Alejandro Martín, Universidad Politécnica de Madrid, Spain
María-Teresa Martín-Valdivia, Universidad de Jaén, Spain
Arturo Montejo-Ráez, University of Jaén, Spain
Manuel Montes-Y-Gómez, Instituto Nacional de Astrofísica, Óptica y Electrónica, Mexico
Roser Morante, Universidad Nacional de Educacion a Distancia (UNED), Spain
Josiane Mothe, Université de Toulouse, France
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates
Joakim Nivre, Uppsala University and RISE, Sweden
Javier Parapar, Universidade da Coruña, Spain
Florina Piroi, Technische Universität Wien, Austria
Simone Paolo Ponzetto, University of Mannheim, Germany
Martin Potthast, University of Kassel, hessian.AI, and ScaDS.AI, Germany
Francisco Rangel, Symanto Research, Spain
Eric Sanjuan, Laboratoire Informatique d'Avignon- Université d'Avignon, France
Areg Mikael Sarvazyan, Symanto Research, Spain
Efstathios Stamatatos, University of the Aegean, Greece
Benno Stein, Bauhaus-Universität Weimar, Germany
Sara Tonelli, Fondazione Bruno Kessler, Italy
Theodora Tsikrika, Information Technologies Institute, CERTH, Greece
Rafael Valencia-Garcia, Universidad de Murcia, Spain
David Vilares, Universidade da Coruña, Spain
Esaú Villatoro, Idiap, Switzerland
Matti Wiegmann, Bauhaus-Universität Weimar, Germany
Christa Womser-Hacker, University of Hildesheim, Germany
Eva Zangerle, University of Innsbruck, Austria
Arkaitz Zubiaga, Queen Mary University of London, United Kingdom

Thank you all very much!

July, 2025

Guglielmo Faggioli,
Nicola Ferro,
Paolo Rosso,
Damiano Spina

Organization

CLEF 2025, *Conference and Labs of the Evaluation Forum – Experimental IR meets Multilinguality, Multimodality, and Interaction*, was hosted by the Universidad Nacional de Educación a Distancia (UNED), Spain.

General Chairs

Jorge Carrillo-de-Albornoz, Universidad Nacional de Educación a Distancia (UNED), Spain

Alba García Seco de Herrera, Universidad Nacional de Educación a Distancia (UNED), Spain

Julio Gonzalo, Universidad Nacional de Educación a Distancia (UNED), Spain

Laura Plaza, Universidad Nacional de Educación a Distancia (UNED), Spain

Program Chairs

Josiane Mothe, Université de Toulouse, France

Florina Piroi, Technische Universität Wien, Austria

Lab Chairs

Paolo Rosso, Universitat Politècnica de València, Spain

Damiano Spina, RMIT University, Australia

Proceedings Chairs

Guglielmo Faggioli, University of Padua, Italy

Nicola Ferro, University of Padua, Italy

Local Organization Committee

Víctor Fresno, Universidad Nacional de Educación a Distancia (UNED), Spain

Enrique Amigó, Universidad Nacional de Educación a Distancia (UNED), Spain

CLEF Steering Committee

Steering Committee Chair

Nicola Ferro, University of Padua, Italy

Steering Committee Co-Chairs

Alba García Seco de Herrera, Universidad Nacional de Educación a Distancia (UNED), Spain

Alberto Barrón-Cedeño, University of Bologna, Italy

Deputy Steering Committee Chair for the Conference

Paolo Rosso, Universitat Politècnica de València, Spain

Deputy Steering Committee Chair for the Evaluation Labs

Martin Braschler, Zurich University of Applied Sciences, Switzerland

Members

Avi Arampatzis, Democritus University of Thrace, Greece

Khalid Choukri, Evaluations and Language resources Distribution Agency (ELDA), France

Fabio Crestani, Università della Svizzera italiana, Switzerland

Carsten Eickhoff, University of T ubingen, Germany

Norbert Fuhr, University of Duisburg-Essen, Germany

Petra Galuščáková, University of Stavanger, Norway

Anastasia Giachanou, Utrecht University, The Netherlands

Lorraine Goeuriot, Université Grenoble Alpes, France

Julio Gonzalo, National Distance Education University (UNED), Spain

Donna Harman, National Institute for Standards and Technology (NIST), USA

Bogdan Ionescu, University "Politehnica" of Bucharest, Romania

Evangelos Kanoulas, University of Amsterdam, The Netherlands

Birger Larsen, University of Aalborg, Denmark

Maria Maistro, University of Copenhagen, Denmark

Josiane Mothe, IRIT, Université de Toulouse, France

Henning Müller, University of Applied Sciences Western Switzerland (HES-SO), Switzerland

Jian-Yun Nie, Université de Montréal, Canada

Gabriella Pasi, University of Milano-Bicocca, Italy

Eric SanJuan, University of Avignon, France

Laure Soulier, Pierre and Marie Curie University (Paris 6), France

Theodora Tsikrika, Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Greece

Past Members

Paul Clough, University of Sheffield, United Kingdom

Djoerd Hiemstra, Radboud University, The Netherlands

Jaana Kekäläinen, University of Tampere, Finland

Séamus Lawless, Trinity College Dublin, Ireland

David E. Losada, Universidade de Santiago de Compostela, Spain

Mihai Lupu, Vienna University of Technology, Austria

Carol Peters, ISTI, National Council of Research (CNR), Italy
(Steering Committee Chair 2000–2009)

Emanuele Pianta, Centre for the Evaluation of Language and Communication Technologies (CELCT), Italy

Maarten de Rijke, University of Amsterdam UvA, The Netherlands

Giuseppe Santucci, Sapienza University of Rome, Italy

Jacques Savoy, University of Neuchêtel, Switzerland

Alan Smeaton, Dublin City University, Ireland

Christa Womser-Hacker, University of Hildesheim, Germany