

What happened in CLEF 2006

Introduction to the Working Notes

Carol Peters

Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), Pisa, Italy
carol.peters@isti.cnr.it

The objective of CLEF is to promote research in the field of multilingual system development. This is done through the organisation of annual evaluation campaigns in which a series of tracks designed to test different aspects of mono- and cross-language information retrieval (IR) are offered. The intention is to encourage experimentation with all kinds of multilingual information access – from the development of systems for monolingual retrieval operating on many languages to the implementation of complete multilingual multimedia search services. This has been achieved by offering an increasingly complex and varied set of evaluation tasks over the years. The aim is not only to meet but also to anticipate the emerging needs of the R&D community and to encourage the development of next generation multilingual IR systems.

These Working Notes contain descriptions of the experiments conducted within CLEF 2006 – the sixth in a series of annual system evaluation campaigns¹. The results of the experiments will be presented and discussed in the CLEF 2006 Workshop, 20-22 September, Alicante, Spain. The final papers - revised and extended as a result of the discussions at the Workshop - together with a comparative analysis of the results will appear in the CLEF 2006 Proceedings, to be published by Springer in their Lecture Notes for Computer Science series.

As from CLEF 2005, the Working Notes are published in electronic format only and are distributed to participants at the Workshop on CD-ROM together with the Book of Abstracts in printed form. All reports included in the Working Notes will also be inserted in the DELOS Digital Library, accessible at <http://delos-dl.isti.cnr.it>.

Both Working Notes and Book of Abstracts are divided into eight sections, corresponding to the CLEF 2006 evaluation tracks. In addition appendices are included containing run statistics for the Ad Hoc, Domain-Specific, GeoCLEF and QA tracks, plus a list of all participating groups showing in which track they took part.

The main features of the 2006 campaign are briefly outlined here below in order to provide the necessary background to the experiments reported in the rest of the Working Notes.

1. Tracks and Tasks in CLEF 2006

CLEF 2006 offered eight tracks designed to evaluate the performance of systems for:

- mono-, bi- and multilingual textual document retrieval on news collections (Ad Hoc)
- mono- and cross-language information on structured scientific data (Domain-Specific)
- interactive cross-language retrieval (iCLEF)
- multiple language question answering (QA@CLEF)
- cross-language retrieval in image collections (ImageCLEF)
- cross-language spoken document retrieval (CL-SR)
- multilingual retrieval of Web documents (WebCLEF)
- cross-language geographical retrieval (GeoCLEF)

Although these tracks are the same as those offered in CLEF 2005, many of the tasks offered are new.

Multilingual Text Retrieval (Ad Hoc): Similarly to last year, the 2006 track offered mono- and bilingual tasks on target collections in French, Portuguese, Bulgarian and Hungarian. The topics (i.e. statements of information needs from which queries are derived) were prepared in a wide range of European languages (Bulgarian, English, French, German, Hungarian, Italian, Portuguese, Spanish). We also offered a bilingual task aimed at encouraging system testing with non-European languages against an English target collection. Topics were supplied in: Amharic, Chinese, Hindi, Indonesian, Oromo and Telugu. This choice of languages was determined by the demand from participants. In addition, a new “robust” task was offered; this task emphasized the importance of stable performance over languages instead of high average performance in mono-, bilingual and multilingual IR. It made use of test collections previously developed at CLEF. The track is coordinated jointly by ISTI-CNR and U.Padua (Italy) and U.Hildesheim (Germany).

¹ CLEF is included in the activities of the DELOS Network of Excellence on Digital Libraries, funded by the Sixth Framework Programme of the European Commission. For information on DELOS, see www.delos.info.

Cross-Language Scientific Data Retrieval (Domain-Specific): This track studied retrieval in a domain-specific context using the GIRT-4 German/English social science database and two Russian corpora: Russian Social Science Corpus (RSSC) and the ISSS collection of sociology and economics documents. Multilingual controlled vocabularies (German-English, English-German, German-Russian, English-Russian) were available. Monolingual and cross-language tasks were offered. Topics were prepared in English, German and Russian. Participants could make use of the indexing terms inside the documents and/or the Social Science Thesaurus provided, not only as translation means, but also for tuning relevance decisions of their system. The track is coordinated by IZ Bonn (Germany).

Interactive CLIR (iCLEF): For CLEF 2006, the interactive track joined forces with the image track to work on a new type of interactive image retrieval task to better capture the interplay between image and the multilingual reality of the internet for the public at large. The task was based on the popular image perusal community Flickr (www.flickr.com), a dynamic and rapidly changing database of images with textual comments, captions, and titles in many languages and annotated by image creators and viewers cooperatively in a self-organizing ontology of tags (a so-called “folksonomy”). The track is coordinated by UNED (Spain), U. Sheffield (UK) and SICS (Sweden).

Multilingual Question Answering (QA@CLEF): This track, which has received increasing interest at CLEF since 2003, evaluated both monolingual (non-English) and cross-language QA systems. The main task evaluated open domain QA systems. Target collections were offered in Bulgarian, Dutch, English (bilingual only), French, German, Italian, Portuguese or Spanish. In addition, three pilot tasks were organized: a task that assessed question answering using Wikipedia, the online encyclopaedia; an Answer Validation exercise; and a “Time-constrained” exercise to be conducted during the workshop. A number of institutions (one for each language) collaborated in the organization of the main task; the Wikipedia activity was coordinated by U. Amsterdam (The Netherlands), the Answer Validation exercise by UNED (Spain) and the Time-constrained exercise by U. Alicante (Spain). The overall coordination of this track is by ITC-irst and CELCT, Trento (Italy).

Cross-Language Retrieval in Image Collections (ImageCLEF):

This track evaluated retrieval of images described by text captions based on queries in a different language; both text and image matching techniques were potentially exploitable. Two main sub-tracks were organised for photographic and medical image retrieval. Each track offered two tasks: bilingual ad hoc retrieval (collection in English, queries in a range of languages) and an annotation task in the first case; medical image retrieval (collection with casenotes in English, French and German, queries derived from short text plus image - visual, mixed and semantic queries) and automatic annotation for medical images (fully categorized collection, categories available in English and German) in the second. The tasks offered different and challenging retrieval problems for cross-language image retrieval. Image analysis was not required for all tasks and a default visual image retrieval system was made available for participants as well as results from a basic text retrieval system. The track coordinators are University of Sheffield (UK) and the University and Hospitals of Geneva (Switzerland). Oregon Health and Science University (USA), Victoria University, Melbourne (Australia), RWTH Aachen University (Germany), and Vienna University of Technology (Austria) collaborate in the task organization.

Cross-Language Speech Retrieval (CL-SR): In 2005, the CL-SR track built a reusable test collection for searching spontaneous conversational English speech using queries in five languages (Czech, English, French, German and Spanish), speech recognition for spoken words, manually and automatically assigned controlled vocabulary descriptors for concepts, dates and locations, manually assigned person names, and hand-written segment summaries. The 2006 CL-SR track included a second test collection containing about 500 hours of Czech speech. Multilingual topic sets were again created for five languages. The track was coordinated by the University of Maryland (USA) and Dublin City University (Ireland).

Multilingual Web Retrieval (WebCLEF): WebCLEF 2006 used the EuroGOV collection, with web pages crawled from European governmental sites for over 20 languages/countries. It was decided to focus this year on the mixed-monolingual known-item topics. The topics were a mixture of old topics and new topics. The old topics were a subset of last year's topics; the new topics were provided by the organizers, using a new method for generating known-item test beds and some human generated new topics. The experiments explored two complementary dimensions: old vs new topics; topics generated by participants vs automatically topics generated by the organizers

Cross-Language Geographical Retrieval (GeoCLEF): The track provided a framework in which to evaluate GIR systems for search tasks involving both spatial and multilingual aspects. Participants were offered a TREC style ad hoc retrieval task based on existing CLEF collections. The aim was to compare methods of query translation, query expansion, translation of geographical references, use of text and spatial retrieval methods separately or combined, retrieval models and indexing methods. Given a multilingual statement describing a spatial user need (topic), the challenge was to find relevant documents from target collections in English and

Portuguese German and/or Spanish news documents. Monolingual and cross-language tasks were activated. 25 topics were prepared in the target languages and in Japanese. Spatial analysis was not required to participate in this task but could be used to augment text-retrieval methods. A number of groups collaborated in the organization of the track; the overall coordination was by UC Berkley (USA) and U. Sheffield (UK).

2. Test Collections

The CLEF test collections, created as a result of the evaluation campaigns, consist of topics or queries, documents, and relevance assessments. Each track was responsible for preparing its own topic/query statements and for performing relevance assessments of the results submitted by participating groups. A number of different document collections were used in CLEF 2006 to build the test collections:

- CLEF multilingual comparable corpus of more than 2 million news docs in 12 languages (see Table 1) ; this corpus was unchanged from 2005. Parts of this collection were used in three tracks: Ad-Hoc (all languages except Finnish, Swedish and Russian), Question Answering (all languages except Finnish, Hungarian, Swedish and Russian) and GeoCLEF (English, German, Portuguese and Spanish).
- The CLEF domain-specific collection consisting of the GIRT-4 social science database in English and German (over 300,000 documents) and two Russian databases: the Russian Social Science Corpus (approx. 95,000 documents) and the Russian ISSS collection for sociology and economics (approx. 150,000 docs). The ISSS corpus was new this year. Controlled vocabularies in German-English and German-Russian were also made available to the participants in this track. This collection was used in the domain-specific track.
- The ImageCLEF track used four collections:
 - the ImageCLEFmed radiological medical database based on a dataset containing images from the Casimage, MIR, PEIR, and PathoPIC datasets (about 50,000 images) with case notes in English (majority) but also German and French.
 - the IRMA collection in English and German of 10,000 images for automatic medical image annotation
 - the IAPR TC-12 database of 25,000 photographs with captions in English, German and Spanish
 - a general photographic collection for image annotation provide by LookThatUp (LTUtech) database
- The Speech retrieval track used the Malach collection of spontaneous conversational speech derived from the Shoah archives in English (more than 750 hours) and Czech (approx 500 hours)
- The WebCLEF track used a collection crawled from European governmental sites, called EuroGOV. This collection consists of more than 3.35 million pages from 27 primary domains. The most frequent languages are Finnish (20%), German (18%), Hungarian (13%), English (10%), and Latvian (9%).

3. Technical Infrastructure

The CLEF technical infrastructure is managed by the DIRECT system. DIRECT manages the test data plus results submission and analyses for the ad hoc, question answering and geographic IR tracks. It has been designed to facilitate data management tasks but also to support the production, maintenance, enrichment and interpretation of the scientific data for subsequent in-depth evaluation studies.

The technical infrastructure is thus responsible for:

- the track set-up, harvesting of documents, management of the registration of participants to tracks;
- the submission of experiments, collection of metadata about experiments, and their validation;
- the creation of document pools and the management of relevance assessment;
- the provision of common statistical analysis tools for both organizers and participants in order to allow the comparison of the experiments;
- the provision of common tools for summarizing, producing reports and graphs on the measured performances and conducted analyses.

DIRECT was designed and implemented by Giorgio Di Nunzio and Nicola Ferro and is described in more detail in a paper in these Working Notes.

Table 1: Sources and dimensions of the CLEF 2006 multilingual comparable corpus

Collection	Added in	Size (MB)	No. of Docs	Median Size of Docs. (Bytes)	Median Size of Docs. (Tokens) ²	Median Size of Docs (Features)
Bulgarian: Sega 2002	2005	120	33,356	NA	NA	NA
Bulgarian: Standart 2002	2005	93	35,839	NA	NA	NA
Dutch: Algemeen Dagblad 94/95	2001	241	106483	1282	166	112
Dutch: NRC Handelsblad 94/95	2001	299	84121	2153	354	203
English: LA Times 94	2000	425	113005	2204	421	246
English: Glasgow Herald 95	2003	154	56472	2219	343	202
Finnish: Aamulehti late 94/95	2002	137	55344	1712	217	150
French: Le Monde 94	2000	158	44013	1994	361	213
French: ATS 94	2001	86	43178	1683	227	137
French: ATS 95	2003	88	42615	1715	234	140
German: Frankfurter Rundschau94	2000	320	139715	1598	225	161
German: Der Spiegel 94/95	2000	63	13979	1324	213	160
German: SDA 94	2001	144	71677	1672	186	131
German: SDA 95	2003	144	69438	1693	188	132
Hungarian: Magyar Hirlap 2002	2005	105	49,530	NA	NA	NA
Italian: La Stampa 94	2000	193	58051	1915	435	268
Italian: AGZ 94	2001	86	50527	1454	187	129
Italian: AGZ 95	2003	85	48980	1474	192	132
Portuguese: Público 1994	2004	164	51751	NA	NA	NA
Portuguese: Público 1995	2004	176	55070	NA	NA	NA
Portuguese: Folha 94	2005	108	51,875	NA	NA	NA
Portuguese: Folha 95	2005	116	52,038	NA	NA	NA
Russian: Izvestia 95	2003	68	16761	NA	NA	NA
Spanish: EFE 94	2001	511	215738	2172	290	171
Spanish: EFE 95	2003	577	238307	2221	299	175
Swedish: TT 94/95	2002	352	142819	2171	183	121

SDA/ATS/AGZ = Schweizerische Depeschagentur (Swiss News Agency)

EFE = Agencia EFE S.A (Spanish News Agency)

TT = Tidningarnas Telegrambyrå (Swedish newspaper)

² The number of tokens extracted from each document can vary slightly across systems, depending on the respective definition of what constitutes a token. Consequently, the number of tokens and features given in this table are approximations and may differ from actual implemented systems.

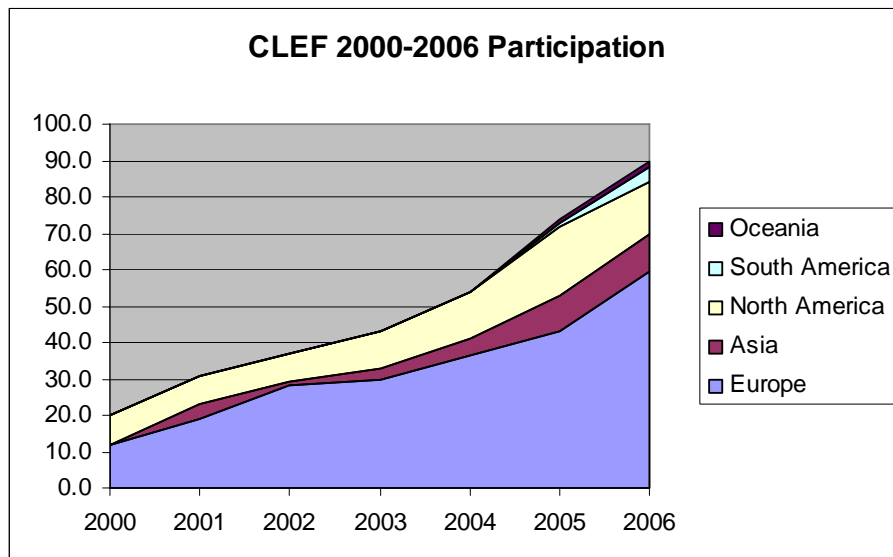


Figure 1. CLEF 2000 – 2006: Increase in Participation

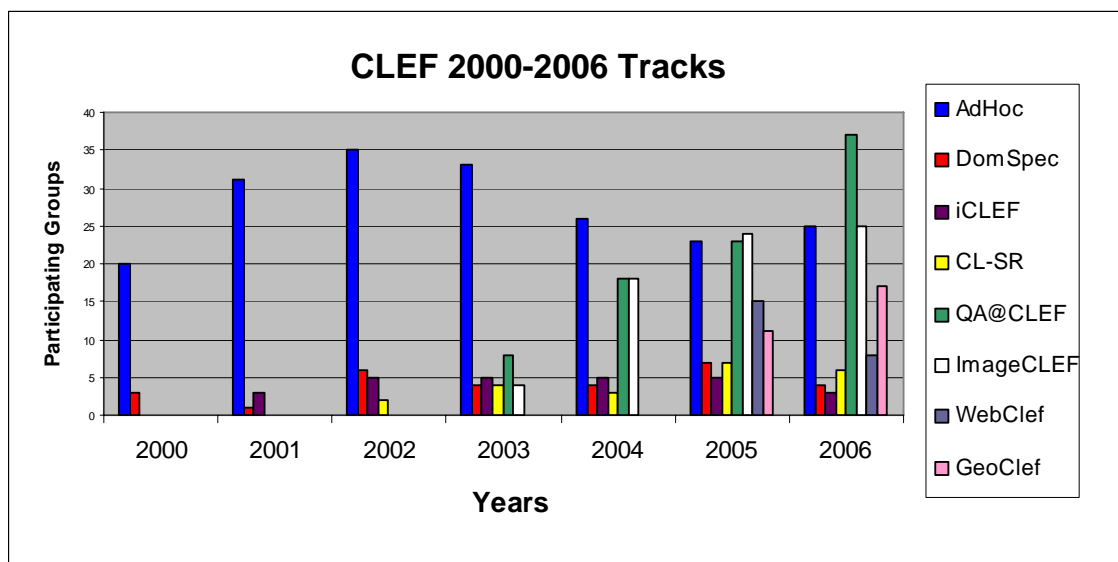


Figure 2. CLEF 2000 – 2006: Shift in Participation

4. Participation

A total of 90 groups submitted runs in CLEF 2006, as opposed to the 74 groups of CLEF 2005: 59.5(43) from Europe, 14.5(19) from N.America; 10(10) from Asia, 4(1) from S.America and 2(1) from Australia³. Last years' figures are given between brackets. The breakdown of participation of groups per track is as follows: Ad Hoc 25; Domain-Specific 4; iCLEF 3; QAatCLEF 37; ImageCLEF 25; CL-SR 6; WebCLEF 8; GeoCLEF 17. As in previous years, participating groups consisted of a nice mix of new-comers (34) and groups that had participated in one or more previous editions (56). Most of the groups came from academia; there were just 9 research groups from industry. A list of groups and indications of the tracks in which they participated is given in Appendix to these Working Notes.

Figure 1 shows the growth in participation this year and Figure 2 shows the shift in focus over the years as new tracks have been added.

³ The 0.5 figures result from a Mexican/Spanish collaboration.

5. Workshop

CLEF aims at creating a strong CLIR/MLIR research and development community. The Workshop plays an important role by providing the opportunity for all the groups that have participated in the evaluation campaign to get together comparing approaches and exchanging ideas. The work of the groups participating in this year's campaign will be presented in plenary and parallel paper sessions and an afternoon poster session. There will also be break-out sessions for more in-depth discussion of the results of individual tracks and intentions for the future. The final sessions will include discussions on ideas for new tracks in future campaigns. Overall, the Workshop should provide an ample panorama of the current state-of-the-art and the latest research directions in the multilingual information retrieval area. I very much hope that it will prove an interesting, worthwhile and enjoyable experience to all those who participate.

The final programme and the presentations at the Workshop will be posted on the CLEF website at <http://www.clef-campaign.org>.

Acknowledgements

I could not run the CLEF evaluation initiative and organize the annual workshops without considerable assistance from many people. CLEF is organized on a distributed basis, with different research groups being responsible for the running of the various tracks. My gratitude goes to all those who have been involved in the coordination of the 2006 campaign. A list of the principal institutions involved is given on the following page. However, it is really impossible for me to list here the names of all the people involved in the organization of the different tracks. Here below, let me just mention those responsible for the overall coordination:

- Giorgio Di Nunzio, Nicola Ferro and Thomas Mandl for the Ad Hoc Track
- Maximilian Stempfhuber, Stefan Baerisch and Natalia Loukachevitch for the Domain-Specific track
- Julio Gonzalo, Paul Clough and Jussi Karlgren for iCLEF
- Bernardo Magnini, Danilo Giampiccolo, Fernando Llopis, Elisa Noguera, Anselmo Peñas and Maarten de Rijke for QA@CLEF
- Paul Clough, Thomas Deselaers, Michael Grubinger, Allan Hanbury, William Hersh, Thomas Lehmann and Henning Müller for ImageCLEF
- Douglas W. Oard and Gareth J. F. Jones for CL-SR
- Krisztian Balog, Leif Azzopardi, Jaap Kamps and Maarten de Rijke for Web-CLEF
- Fredric Gey, Ray Larson and Mark Sanderson as the main coordinators of GeoCLEF

I apologise for those I have not mentioned here. However, I really must express my appreciation to Diana Santos and her colleagues at Linguateca in Norway and Portugal, for all their efforts aimed at supporting the inclusion of Portuguese in CLEF activities. I also thank all those colleagues who have helped us by preparing topic sets in different languages and in particular the NLP Lab. Dept. of Computer Science and Information Engineering of the National Taiwan University for their work on Chinese.

I should also like to thank the members of the CLEF Steering Committee who have assisted me with their advice and suggestions throughout this campaign.

Furthermore, I gratefully acknowledge the support of all the data providers and copyright holders, and in particular:

- The Los Angeles Times, for the American-English data collection
- SMG Newspapers (The Herald) for the British-English data collection
- Le Monde S.A. and ELDA: Evaluations and Language resources Distribution Agency, for the French data
- Frankfurter Rundschau, Druck und Verlagshaus Frankfurt am Main; Der Spiegel, Spiegel Verlag, Hamburg, for the German newspaper collections
- InformationsZentrum Sozialwissenschaften, Bonn, for the GIRT database
- SocioNet system for the Russian Social Science Corpora
- Hypersystems Srl, Torino and La Stampa, for the Italian data
- Agencia EFE S.A. for the Spanish data
- NRC Handelsblad, Algemeen Dagblad and PCM Landelijke dagbladen/Het Parool for the Dutch newspaper data
- Aamulehti Oyj and Sanoma Osakeyhtiö for the Finnish newspaper data
- Russika-Izvestia for the Russian newspaper data
- Público, Portugal, and Linguateca for the Portuguese (PT) newspaper collection
- Folha, Brazil, and Linguateca for the Portuguese (BR) newspaper collection
- Tidningarnas Telegrambyrå (TT) SE-105 12 Stockholm, Sweden for the Swedish newspaper data

- Schweizerische Depeschagentur, Switzerland, for the French, German and Italian Swiss news agency data
- Ringier Kiadoi Rt. [Ringier Publishing Inc.] and the Research Institute for Linguistics, Hungarian Acad. Sci. for the Hungarian newspaper documents
- Sega AD, Sofia; Standart Nyuz AD, Sofia, and the BulTreeBank Project, Linguistic Modelling Laboratory, IPP, Bulgarian Acad. Sci, for the Bulgarian newspaper documents
- St Andrews University Library for the historic photographic archive
- University and University Hospitals, Geneva, Switzerland and Oregon Health and Science University for the ImageCLEFmed Radiological Medical Database
- Aachen University of Technology (RWTH), Germany for the IRMA database of annotated medical images
- The Survivors of the Shoah Visual History Foundation, and IBM for the Malach spoken document collection

Without their contribution, this evaluation activity would be impossible.

Last and not least, I should like to express my gratitude to Alessandro Nardi in Pisa and José Luis Vicedo, Patricio Martínez Barco and Maximiliano Saiz Noeda, U. Alicante, for their assistance in the organisation of the CLEF 2006 Workshop.

Coordination

CLEF is coordinated by the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa. The following institutions have contributed to the organisation of the different tracks of the CLEF 2006 campaign:

- Centre for the Evaluation of Human Language and Multimodal Communication Technologies (CELCT), Trento, Italy
- Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Trento, Italy
- College of Information Studies and Inst. for Advanced Computer Studies, University of Maryland, USA
- Department of Computer Science, University of Helsinki
- Department of Computer Science, University of Indonesia
- Department of Computer Science Department, RWTH Aachen University, Germany
- Department of Computer Science and Information Systems, University of Limerick, Ireland
- Department of Computer Science and Information Engineering, National University of Taiwan
- Department of Information Engineering, University of Padua, Italy
- Department of Information Science, University of Hildesheim, Germany
- Department of Information Studies, University of Sheffield, UK
- Department of Medical Informatics, Aachen University of Technology (RWTH), Germany
- Evaluations and Language Resources Distribution Agency Sarl, Paris, France
- German Research Centre for Artificial Intelligence, DFKI, Saarbrücken,
- Information and Language Processing Systems, University of Amsterdam, Netherlands
- InformationsZentrum Sozialwissenschaften, Bonn, Germany
- Institute for Information technology, Hyderabad, India
- Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, Madrid, Spain
- Linguateca, Sintef, Oslo, Norway;
- Linguistic Modelling Laboratory, Bulgarian Academy of Sciences
- National Institute of Standards and Technology, Gaithersburg MD, USA
- Biomedical Informatics, Oregon Health and Science University, USA
- Research Computing Center of Moscow State University
- Research Institute for Linguistics, Hungarian Academy of Sciences
- School of Computer Science and Mathematics, Victoria University, Australia
- School of Computing, Dublin City University, Ireland
- UC Data Archive and School of Information Management and Systems, UC Berkeley, USA
- University "Alexandru Ioan Cuza", IASI, Romania
- University Hospitals and University of Geneva, Switzerland

CLEF Steering Committee

Maristella Agosti, University of Padova, Italy
Martin Braschler, Zurich University of Applied Sciences Winterthur, Switzerland
Amedeo Cappelli, ISTI-CNR & CELCT, Italy
Hsin-Hsi Chen, National Taiwan University, Taipei, Taiwan
Khalid Choukri, Evaluations and Language resources Distribution Agency, Paris, France
Paul Clough, University of Sheffield, UK
Thomas Deselaers, RWTH Aachen University, Germany
David A. Evans, Clairvoyance Corporation, USA
Marcello Federico, ITC-irst, Trento, Italy
Christian Fluhr, CEA-LIST, Fontenay-aux-Roses, France
Norbert Fuhr, University of Duisburg, Germany
Frederic C. Gey, U.C. Berkeley, USA
Julio Gonzalo, LSI-UNED, Madrid, Spain
Donna Harman, National Institute of Standards and Technology, USA
Gareth Jones, Dublin City University, Ireland
Franciska de Jong, University of Twente, Netherlands
Noriko Kando, National Institute of Informatics, Tokyo, Japan
Jussi Karlgren, Swedish Institute of Computer Science, Sweden
Michael Kluck, German Institute for International and Security Affairs, Berlin, Germany
Natalia Loukachevitch, Moscow State University, Russia
Bernardo Magnini, ITC-irst, Trento, Italy
Paul McNamee, Johns Hopkins University, USA
Henning Müller, University & University Hospitals of Geneva, Switzerland
Douglas W. Oard, University of Maryland, USA
Maarten de Rijke, University of Amsterdam, Netherlands
Jacques Savoy, University of Neuchâtel, Switzerland
Diana Santos, Linguatca, Sintef, Oslo, Norway
Peter Schäubel, Eurospider Information Technologies, Switzerland
Richard Sutcliffe, University of Limerick, Ireland
Max Stempfhuber, Informationszentrum Sozialwissenschaften Bonn, Germany
Hans Uszkoreit, German Research Center for Artificial Intelligence (DFKI), Germany
Felisa Verdejo, LSI-UNED, Madrid, Spain
José Luis Vicedo, University of Alicante, Spain
Ellen Voorhees, National Institute of Standards and Technology, USA
Christa Womser-Hacker, University of Hildesheim, Germany