

Cross-Language Evaluation Forum

From CLEF 2003 to CLEF 2004

Carol Peters
Martin Braschler
* Jacques Savoy



Outline



What happened at CLEF 2003

- ◆ Tracks and Tasks
- ◆ Test Collection
- ◆ Participation
- ◆ Results

What is happening in CLEF 2004

NTCIR-4 Workshop

CLEF 2003: Core Tracks



Free-text retrieval on news corpora

- ◆ Multilingual: 2 tasks
 - Small-multilingual: 4 "core" languages (EN,ES,FR,DE)
 - Large-multilingual: 8 languages (+FI,IT,NL,SV)
 - Topics in 12 languages including JP and ZH
- ◆ Bilingual: Aim was comparability
 - IT -> ES FR -> NL
 - DE -> IT FI -> DE
 - x -> RU Newcomers only: x -> EN
- ◆ Monolingual: All languages (except English)

Retrieval on structured, domain-specific data

- ◆ Mono- and CLIR on social science data (DE, EN)

NTCIR-4 Workshop

CLEF 2003: Additional Tracks



- ◆ Interactive Track – iCLEF (coordinated by UNED, UMD)
 - Interactive document selection/query formulation
- ◆ Multilingual QA Track (ITC-irst, UNED, U. Amsterdam, NIST)
 - Monolingual QA for Dutch, Italian and Spanish
 - Cross-language QA to English target collection
- ◆ ImageCLEF (coordinated by U. Sheffield)
 - Cross-language image retrieval using captions
- ◆ Cross-Language Spoken Doc Retrieval (ITC-irst, U. Exeter)
 - Evaluation of CLIR on noisy transcripts of spoken docs
 - Low-cost development of a benchmark

NTCIR-4 Workshop

CLEF 2003 Data Collections



- ◆ Multilingual comparable corpus
 - news docs in 9 languages - DE, EN, ES, FI, FR, IT, NL, RU, SV
 - Common set of 60 topics in 10 languages (+ZH) - core tracks
 - 2 sets of 200 questions for mono- and cross-language QA
- ◆ GIRT4: German and English social science docs
 - plus German/English/Russian thesaurus
 - 25 topics in DE/EN/RU
- ◆ St Andrews University Image Collection
 - historical photo collection with EN captions
 - 50 short topics in DE, ES, FR, IT, NL
- ◆ CL-SDR TREC-8 and TREC-9 SDR collections
 - noisy spoken doc. transcripts in English
 - 100 short topics in DE, ES, FR, IT, NL

NTCIR-4 Workshop

CLEF 2003: Participants

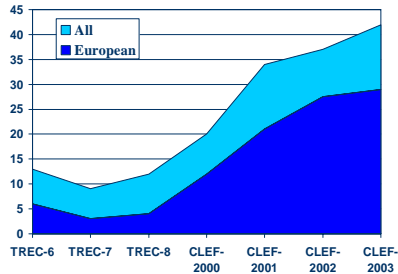


- | | | |
|---------------------------------|-----------------------------|----------------------------|
| • BBN/UMD (US) | • ISI U Southern Cal (US) | • U Amsterdam (NL) ** |
| • CEA/LIC2M (FR) | • ITC-irst (IT) *** | • U Exeter (UK) ** |
| • CLIPS/IMAG (FR) | • JHU-APL (US) *** | • U Oviedo/AIC (ES) |
| • CMU (US) * | • Kermit (FR/UK) | • U Hildesheim (DE) * |
| • Clairvoyance Corp. (US) * | • Medialab (NL) ** | • U Maryland (US) *** |
| • COLE /U La Coruna (ES) * | • NII (JP) | • U Montreal/RALI (CA) *** |
| • Daedalus (ES) | • National Taiwan U (TW) ** | • U Neuchâtel (CH) ** |
| • DFKI (DE) | • OCE Tech. BV (NL) ** | • U Sheffield (UK) *** |
| • DLTG U Limerick (IE) | • Ricoh (JP) | • U Sunderland (UK) *** |
| • ENEA/La Sapienza (IT) | • SICS (SV) ** | • U Surrey (UK) |
| • Fernuni Hagen (DE) | • SINAI/U Jaen (ES) ** | • U Tampere (FI) *** |
| • Fondazione Ugo Bordoni (IT) * | • Tagmatica (FR) * | • U Twente (NL) *** |
| • Hummingbird (CA) ** | • U Alicante (ES) ** | • UC Berkeley (US) *** |
| • IMS U Padova (IT) * | • U Buffalo (US) | • UNED (ES) ** |

42 groups, 14 countries; 29 European, 10 N.American, 3 Asian
32 academia, 10 industry
(*/**/*** = one/two/three previous participations)

NTCIR-4 Workshop

From CLIR-TREC to CLEF Growth in Participation



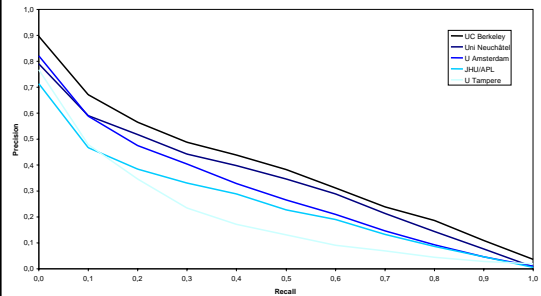
From CLIR-TREC to CLEF Growth in Test Collection (Main Tracks)

	# part.	# lang	# docs.	Size in MB	# assess.	# topics	# ass. per topic
CLEF 2003	33	9	1,611,178	4124	188,475	60 (37)	~3100
CLEF 2002	34	8	1,138,650	3011	140,043	50(30)	~2900
CLEF 2001	31	6	940,487	2522	97,398	50	1948
CLEF 2000	20	4	368,763	1158	43,566	40	1089
TREC8 CLIR	12	4	698,773	1620	23,156	28	827

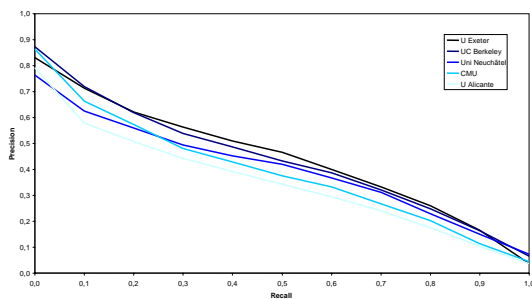
Details of Experiments

Track	# Participants	# Runs/Experiments
Multilingual-8	7	33
Multilingual-4	14	53
Bilingual to FI → DE	2	3
Bilingual to X → EN	3	15
Bilingual to IT → ES	9	25
Bilingual to DE → IT	8	21
Bilingual to FR → NL	3	6
Bilingual to X → RU	2	9
Monolingual DE	13	30
(Monolingual EN)	(5)	11
Monolingual ES	16	38
Monolingual FI	7	13
Monolingual FR	16	36
Monolingual IT	13	27
Monolingual NL	11	32
Monolingual RU	5	23
Monolingual SV	8	18
Domain-specific GRT → DE	4	16
Domain-specific GRT → EN	2	6
Interactive	5	10
Question Answering	8	17
Image Retrieval	4	45
Spoken Document Retrieval	4	29

CLEF 2003 Multilingual-8 Track - TD, Automatic



CLEF 2003 Multilingual-4 Track - TD, Automatic



Trends in CLEF-2003

- ◆ A lot of detailed fine-tuning (per language, per weighting scheme, per translation resource type)
- ◆ People think about ways to “scale” to new languages
- ◆ Merging is still a hot issue; however, no merging approach besides the simple ones has been widely adopted yet
- ◆ A few resources were really popular: Snowball stemmers, UniNE stopwords, some MT systems, “Freelang” dictionaries
- ◆ QT still rules

Trends in CLEF-2003



- ◆ Stemming and decompounding are still actively debated; maybe even more use of linguistics than before?
- ◆ Monolingual tracks were “hotly contested”, some show very similar performance among the top groups
- ◆ Bilingual tracks forced people to think about “inconvenient” language pairs
- ◆ Success of the “additional” tracks

NTCIR-4 Workshop

CLEF-2003 vs. CLEF-2002



- ◆ Many participants were back
- ◆ Many groups tried several tasks
- ◆ People try each other's ideas/methods:
 - collection-size based merging, 2step merging
 - (fast) document translation
 - compound splitting, stemmers
- ◆ Returning participants usually improve performance. (“Advantage for veteran groups”)
- ◆ Scaling up to Multilingual-8 takes its time (?)
- ◆ Strong involvement of new groups in track coordination

NTCIR-4 Workshop

“Effect” of CLEF in 2003



- ◆ Number of Europeans grows more slowly (29)
- ◆ Fine-tuning for individual languages, weighting schemes etc. has become a hot topic
 - are we overtuning to characteristics of the CLEF collection?
- ◆ Some blueprints to “successful CLIR” have now been widely adopted
 - Are we headed towards a monoculture of CLIR systems?
- ◆ Multilingual-8 was dominated by veterans, but Multilingual-4 was very competitive
- ◆ “inconvenient” language pairs for bilingual; stimulated some interesting work
- ◆ Increase of groups with NLP background (effect of QA)

NTCIR-4 Workshop

CLEF 2003 Workshop



- ◆ Results of CLEF 2002 campaign presented at Workshop, 20-21 Aug. 2003, Trondheim
- ◆ 60 researchers and system developers from academia and industry participated
- ◆ Working Notes containing preliminary reports and statistics on CLEF 2003 experiments available on Web site
- ◆ Proceedings to be published by Springer in LNCS series

NTCIR-4 Workshop

CLEF 2004



Reduction of “core” tracks – expansion of “new” tracks

- ◆ Mono-, Bi-, and Multilingual IR on News Collections
 - Just 5 target languages (EN/FI/FR/RU and new language - Portuguese)
- ◆ Mono- and Cross-Language Information Retrieval on Structured Scientific Data
 - GIRT-4 EN and DE social science data

NTCIR-4 Workshop

CLEF 2004



Considerable focus on QA

- ◆ Multilingual Question Answering (QA at CLEF)
 - Mono and Cross-Language QA: target collections for DE/EN/ES/FR/IT/NL/PT
- ◆ Interactive CLIR - iCLEF
 - Cross-Lang. QA from a user-inclusive perspective
 - How can interaction with user help a QA system
 - How should C-L system help users locate answers quickly
 - Coordination with QA track

NTCIR-4 Workshop

CLEF 2004



Importance of non-textual media

- ◆ Cross-Language Image Retrieval (ImageCLEF)
 - Using both text and image matching techniques
 - bilingual ad hoc retrieval task (ES/FR/DE/IT/NL)
 - an interactive search task (tentative)
 - a medical image retrieval task
- ◆ Cross-Lang. Spoken Doc Retrieval (CL-SDR)
 - evaluation of CLIR systems on noisy automatic transcripts of spoken documents
 - CL-SDR from ES/FR/DE/IT/NL
 - retrieval with/without known story boundaries
 - use of multiple automatic transcriptions

NTCIR-4 Workshop

CLEF 2004



- ◆ 60 groups registered
- ◆ Results due end May (dates vary slightly according to the track)
- ◆ QA@CLEF and ImageCLEF particularly popular tasks
- ◆ 16 groups registered for the multilingual task (target document collection in 4 languages: EN, FI, FR, RU)
- ◆ 22 groups registered for QA@CLEF; 19 for ImageCLEF
- ◆ Workshop: 15-17 September, Bath, UK (after European Conference on Digital Libraries)

NTCIR-4 Workshop

Cross-Language Evaluation Forum



For further information see:
<http://www.clef-campaign.org>

or contact:
Carol Peters - ISTI-CNR
E-mail: carol@isti.cnr.it

NTCIR-4 Workshop