



CLEF: Ongoing Activities and Plans for the Future

Maristella Agosti Giorgio Di Nunzio Nicola Ferro Carol Peters

University of Padua
Italy

{agosti, dinunzio, ferro}@dei.unipd.it

ISTI-CNR, Area di Ricerca
Italy

carol.peters@isti.cnr.it



CLEF Objectives

- ❑ Promote research and stimulate development of multilingual IR systems for European (and non-European) languages, through
 - Creation of evaluation infrastructure and organisation of regular evaluation campaigns for system testing
 - Building of an MLIA/CLIR research community
 - Construction of publicly available test-suites

- ❑ **Major Goal**
 - To anticipate the emerging needs of the R&D community, and
 - Encourage development of truly multilingual, multimodal systems



CLEF History

- ❑ 1997 – First CLIR system evaluation campaigns in US and Japan: Text REtrieval Conference (TREC) and NTCIR
 - CLEF actually began life in 1997 as a track for Cross Language Information Retrieval (CLIR) within TREC. Mainly, English centered tasks (EN -> X, X -> EN).

- ❑ 2000 – CLIR evaluation in Europe: CLEF (extension of CLIR track at TREC)
 - Fully multilingual, multimodal information retrieval systems
 - capable of processing a query in any medium and any language
 - finding relevant information from a multilingual multimedia collection containing documents in any language and form,
 - and presenting it in the style most likely to be useful to the user

- ❑ The European Network of Excellence on Digital Libraries (DELOS) has supported CLEF since 2000 (however, this support will end this year 2007)
 - DELOS under FP5 2000 – 2003 (<http://delos-noe.isti.cnr.it/>)
 - DELOS under FP6 2004 – 2007 (<http://www.delos.info/>)



CLEF Technical Infrastructure History

- ❑ 1997 – 1999: CLEF as CLIR Track at TREC
 - same infrastructure developed at NIST

- ❑ 2000 – 2004: CLEF as CLIR stand-alone European initiative
 - Martin Braschler developed and managed the CLEF infrastructure, starting from the TREC one

- ❑ 2005 – 2007: CLEF as CLIR stand-alone European initiative
 - University of Padua designed and developed a new infrastructure, applying concepts and approaches of the digital library field



CLEF Organisation

- ❑ CLEF is mainly a voluntary activity: the following Institutions 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 Institute for Advanced Computer Studies, U. Maryland, USA
- ❑ Dept. of Computer Science, U. Indonesia
- ❑ Depts. of Computer Science & Medical Informatics, RWTH Aachen U., Germany
- ❑ Dept. of Computer Science and Information Systems, U. Limerick, Ireland
- ❑ Dept. of Computer Science and Information Engineering, National U. Taiwan
- ❑ Dept. of Information Engineering, U. Padua, Italy
- ❑ Dept. of Information Sci, U. Hildesheim, Germany
- ❑ Dept. of Information Studies, U. Sheffield, UK
- ❑ Evaluations and Language Resources Distribution Agency Sarl, Paris, France
- ❑ German Research Centre for Artificial Intelligence, DFKI, Saarbrücken, Germany
- ❑ Information and Language Processing Systems, U. Amsterdam, Netherlands
- ❑ IZ Bonn, Germany
- ❑ Inst. Information Technology, Hyderabad, India
- ❑ LSI-UNED, Madrid, Spain
- ❑ Linguateca, Sintef, Oslo, Norway
- ❑ Linguistic Modelling Lab., Bulgarian Acad Sci
- ❑ NIST, USA
- ❑ Biomedical Informatics, Oregon Health and Science University, USA
- ❑ Research Computing Cente, Moscow State U.
- ❑ Research Institute for Linguistics, Hungarian Academy of Sciences
- ❑ School of Computer Science and Mathematics, Victoria U., Australia
- ❑ School of Computing, DCU, Ireland
- ❑ UC Data Archive and School of Information Management and Systems, UC Berkeley, USA
- ❑ University "Alexandru Ioan Cuza", IASI, Romania
- ❑ U. Hospitals and U.of Geneva, Switzerland

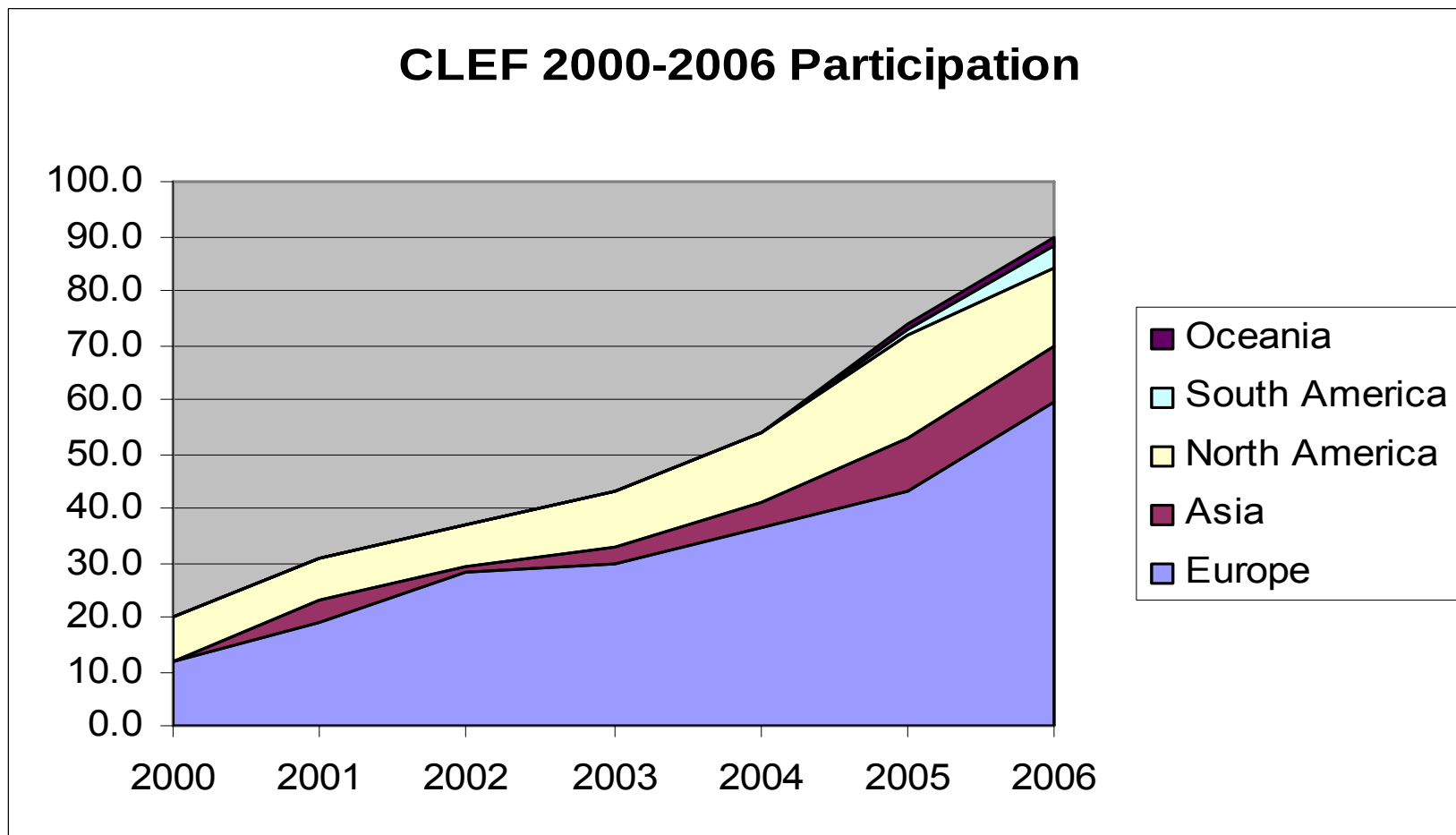


CLEF Track Evolution

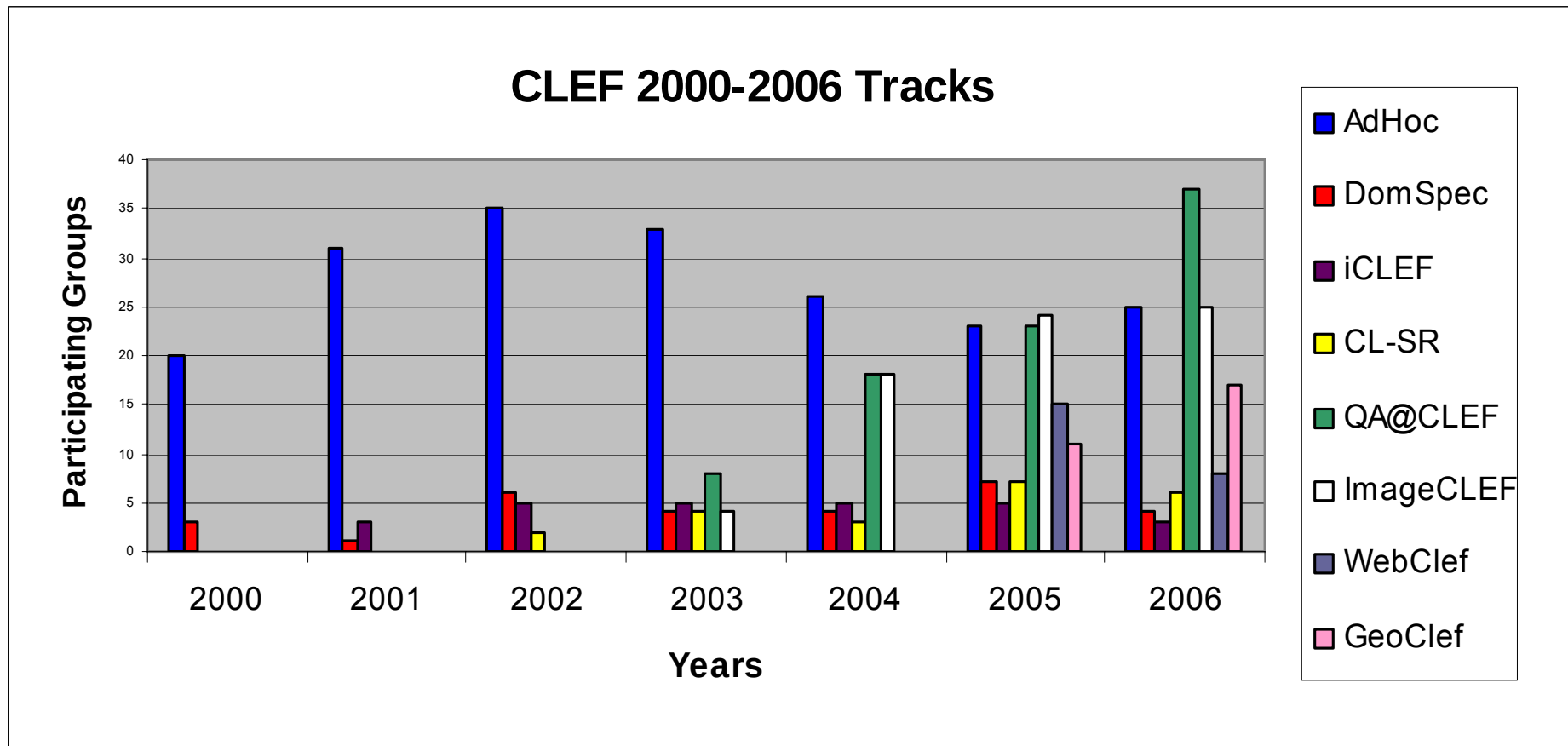
CLEF 2000	❑ mono-, bi- and multilingual text document retrieval (Ad-Hoc) ❑ mono- and cross-language tasks on structured scientific data (Domain-Specific)
CLEF 2001 Added	❑ interactive cross-language retrieval (iCLEF)
CLEF 2002 Added	❑ cross-language spoken document retrieval (CL-SR)
CLEF 2003 Added	❑ multiple language question answering (QA@CLEF) ❑ cross-language retrieval in image collections (ImageCLEF)
CLEF 2005 Added	❑ multilingual retrieval of Web documents (WebCLEF) ❑ cross-language geographical retrieval (GeoCLEF)



CLEF 2000-2006 Participation



CLEF 2006: over 100 groups registered; 90 actually submitted runs for 1 or more tasks.
Most groups from academia; just 9 research groups from industry.



Breakdown of groups per track in 2006: Ad Hoc 25; Domain-Specific 4; iCLEF 3; QAatCLEF 37; ImageCLEF 25; CL-SR 6; WebCLEF 8; GeoCLEF 17



CLEF Tracks: CLIR Text Retrieval

❑ Ad-hoc track (from 2000)

- mono-, bi- and multilingual system performance tested
- bilingual task testing non-European languages (Amharic, Chinese, Hindi, Indonesian, Oromo, Telugu) against English target collection
- “multi-8 two-years-on / merging” checked improvement in system performance over time (CLEF 2003 → 2005) and focused on problem of merging results from different collections/languages
- “robust” task emphasized importance of stable performance over languages instead of high average performance

❑ Domain-specific track (from 2000)

- cross-language retrieval on German and Russian databases of social sciences



CLEF Tracks: Ad-Hoc Tasks

	Monolingual	Bilingual	Multilingual
CLEF 2000	DE;FR;IT	X→EN	X→DE;EN;FR;IT
CLEF 2001	DE; ES; FR; IT; NL	X→EN, X→NL	X→DE;EN;ES;FR;IT
CLEF 2002	DE; ES; FI; FR IT; NL; SV	X→DE;ES;FI;FR;IT;NL;SV X→EN (newcomer)	X→DE;EN;ES;FR;IT
CLEF 2003	DE; ES; FI; FR IT; NL; RU; SV	IT→ES; DE→IT FR→NL; FI→DE X→RU; X→EN	X→DE;EN;ES;FR X→DE;EN;ES;FI FR;IT;NL;SV
CLEF 2004	FI; FR; RU; PT	ES/FR/IT/RU→FI DE/FI/NL/SV→FR X→RU;X→EN	X→FI;FR;RU;PT
CLEF 2005	BG; FR; HU; PT	X→ BG;FR;HU;PT X →EN	Multi8 2yrson Multi8 merge
CLEF 2006	BG; FR; HU; PT	X→ BG;FR;HU;PT X →EN	Robust
CLEF 2007	BG; CZ; HU	X→ BG;CS;HU X →EN	(None)



CLEF Tracks: Ad-Hoc Results Over the Years

❑ TREC-6, 1997:

- EN→FR: 49% of best monolingual French system
- EN→DE: 64% of best monolingual German system

❑ CLEF 2002:

- EN→FR: 83,4% of best monolingual French system
- EN→DE: 85,6% of best monolingual German system

❑ CLEF 2003 enforced the use of “unusual” language pairs:

- IT→ES: 83% of best monolingual Spanish IR system
- DE→IT: 87% of best monolingual Italian IR system
- FR→NL: 82% of best monolingual Dutch IR system

❑ CLEF 2005

- X→FR: 85% of best monolingual French IR system
- X→PT: 88% of best monolingual Portuguese IR system
- X→BG: 74% of best monolingual Bulgarian IR system
- X→HU: 73% of best monolingual Hungarian IR system

❑ CLEF 2006 results of “established” languages:

- X→FR: 94% of best monolingual French IR system
- X→PT: 91% of best monolingual Portuguese IR system
- X→BG: 52% of best monolingual Bulgarian IR system
- X→HU: 51% of best monolingual Hungarian IR system

only one
participant for CLIR
BG and HU



CLEF Tracks:

Cross-Language Information Extraction (1/2)

❑ Multilingual Question Answering – QA@CLEF (from 2003)

- mono- and cross-language QA systems: target collections for BG/DE/EN/ES/FR/IT/NL/PT
- pilot tasks: QA using Wikipedia; answer validation exercise; “time-constrained” exercise
- cross-language QA is around 60% of monolingual QA

❑ Cross-language Geographic Retrieval - GeoCLEF (from 2005)

- Aim: to evaluate retrieval of multilingual documents with an emphasis on geographic search
- Many documents contains geo-references expressed in multiple languages
- Collections:
 - Document languages: English, German, Spanish, Portuguese news
 - Topics: English, German, Portuguese and Spanish
 - Monolingual and bilingual ad-hoc retrieval tasks



CLEF Tracks:

Cross-Language Information Extraction (2/2)

❑ Web track– WebCLEF (from 2005)

- Aim: evaluate systems that address multilingual information needs on the web
- Three tasks organized: mixed monolingual, multilingual, and bilingual English to Spanish, with homepage and named page finding queries for all tasks



CLEF Tracks:

Cross-Language Multimedia Retrieval (1/2)

❑ Cross-Language Speech Retrieval (CL-SR) (from 2002)

- 2002-2004: evaluation of CLIR systems on noisy automatic transcripts of spoken EN documents
 - Written queries in ES/FR/DE/IT/NL
 - retrieval with/without known story boundaries
 - use of multiple automatic transcriptions
- from 2004: encourage development of technologies to facilitate access to spontaneous speech
 - Mono- and cross-language retrieval systems are assessed on the Malach collection of oral interviews (Shoah archives: English & Czech collections)
 - Topics prepared in 5 languages (CZ, EN, FR, DE, ES)
 - A thesaurus, an in-domain expansion collection, and a word lattice are available



CLEF Tracks:

Cross-Language Multimedia Retrieval (2/2)

❑ ImageCLEF (from 2003)

○ Aim

- To promote the evaluation of image retrieval and unite both image and text retrieval communities
- To investigate how visual and textual features can best be combined for retrieval

- 2 main subtracks: photographic and medical image retrieval (ad hoc and annotation tasks)

❑ Interactive CLIR – iCLEF (from 2001)

- Cross-Language IR from a user-inclusive perspective
- Interactive document selection/query formulation
- How can interaction with user help a QA system
- How can interaction with user help an image retrieval system
 - task based on Flickr database: images with textual comments, captions, and titles in many languages by using the so-called folksonomy



CLEF Test Collections

❑ Ad Hoc, QA@CLEF, iCLEF, GeoCLEF

- **CLEF multilingual comparable corpus** of more than 2M news docs in 13 languages:
 - BG, CZ, DE, EN, ES, FI, FR, HU, IT, NL, RU, SV, PT

❑ Domain-Specific

- GIRT-4 social science database in EN and DE: more than 300,000 docs
- Russian Social Science Corpus & ISSS database: almost 250,000 docs

❑ ImageCLEF

- IAPR TC-12 database: 25,000 photos with captions in EN, DE, ES
- CasImage radiological db with case notes in FR and EN: 9,000 images
- PEIR 33,000 images, MIR 2,000, PathoPic 9,000
- IRMA EN/DE collection for medical image annotation: 10,000 images

❑ Cross-Language Speech Retrieval

- Malach collection of spontaneous conversational speech derived from the Shoah archives: in English (ca 750 hours); Czech (ca 500 hours)

❑ WebCLEF

- EuroGOV, more than 3.35 M webpages crawled from European governmental sites



CLEF Overall Results

- ❑ Stimulation of research activity in new, previously unexplored areas, such as cross-language question answering, image and geographic information retrieval
- ❑ Study and implementation of evaluation methodologies for diverse types of cross-language IR systems
- ❑ Documented improvement in system performance for cross-language text retrieval systems
- ❑ Creation of a large set of empirical data about multilingual information access from the user perspective
- ❑ Quantitative and qualitative evidence with respect to best practice in cross-language system development
- ❑ Creation of important, reusable test collections for system benchmarking
- ❑ Building of a strong, multidisciplinary research community



Plans for the Future

- ❑ Challenge in CLEF 2007 is to stimulate transfer of research results to market place
- ❑ Same tracks as in 2006 but new tasks will focus on user-oriented activities (interactive and real-time exercises) and on multilingual web searching.
- ❑ Tasks will be activated in collaboration with 2 important international NLPevaluation campaigns (SemEval-2007 & MorphChallenge 2007)
- ❑ Demos of operational systems planned for the workshop
- ❑ An investigation into ways and forms to continue CLEF after 2007 is now underway. External support for overall coordination is essential if CLEF is to continue after 2007.



We warmly thank all the participants and organizers of CLEF who, with their effort and co-operation, make CLEF a successful initiative



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

or contact:

Carol Peters - ISTI-CNR

E-mail: carol.peters@isti.cnr.it



DIRECT: How to Manage CLEF Scientific Data



CLEF: Need for a Digital Library for Scientific Data

- ❑ The experimental evaluation is a scientific activity and, as such, we have to realise that its outcomes are very valuable scientific data.

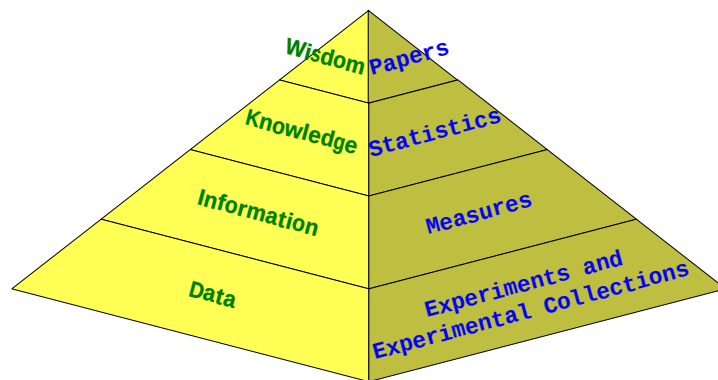
- ❑ Are these scientific data
 - properly modelled,
 - effectively managed and archived,
 - carefully curated and enrichedin the current evaluation methodology?

- ❑ In this context, Digital Library Systems (DLSs) represent the natural choice for managing, making accessible, citing, curating, enriching, and preserving all the information resources produced during an evaluation campaign.



Test Collections and Scientific Data

- ❑ Scientific data, their enrichment and interpretation are essential components of scientific research.
- ❑ The Cranfield methodology traces out how these scientific data have to be produced, while the statistical analysis of experiments provide the means for further elaborating and interpreting the experimental results.
- ❑ Nevertheless, current methodologies do not require any particular coordination or synchronization between the basic scientific data and the analyses on them, which are treated as almost separated items.





Key Points for a New Approach to Design and Develop an Evaluation Infrastructure

❑ Introduce a **conceptual model**

- it makes clear what are the **entities** entailed by the information space of an evaluation campaign, their **features**, and their **relationships**
- **logical models** can be derived from it to **manage** and **preserve** the experimental data
- commonly agreed **data formats** for **exchanging information** can be derived from it

❑ Develop common **metadata formats**

- they provide meaning to the data, and thereby enable their **sharing** and **re-use**
- they allow to keep track of the **lineage** of the managed information

❑ Adopt a **unique identification** and **resolution** mechanism

- it allows for explicit **citation** and **easy access** to the scientific data and it supports the **enrichment** of the scientific data
- **Digital Object Identifiers (DOI)** are a valuable possibility since they have been applied to scientific data and allow metadata to be associated with identified objects

❑ Give to **organizations** responsible for evaluation initiatives an **active role** in developing a comprehensive strategy for long-lived digital data collections

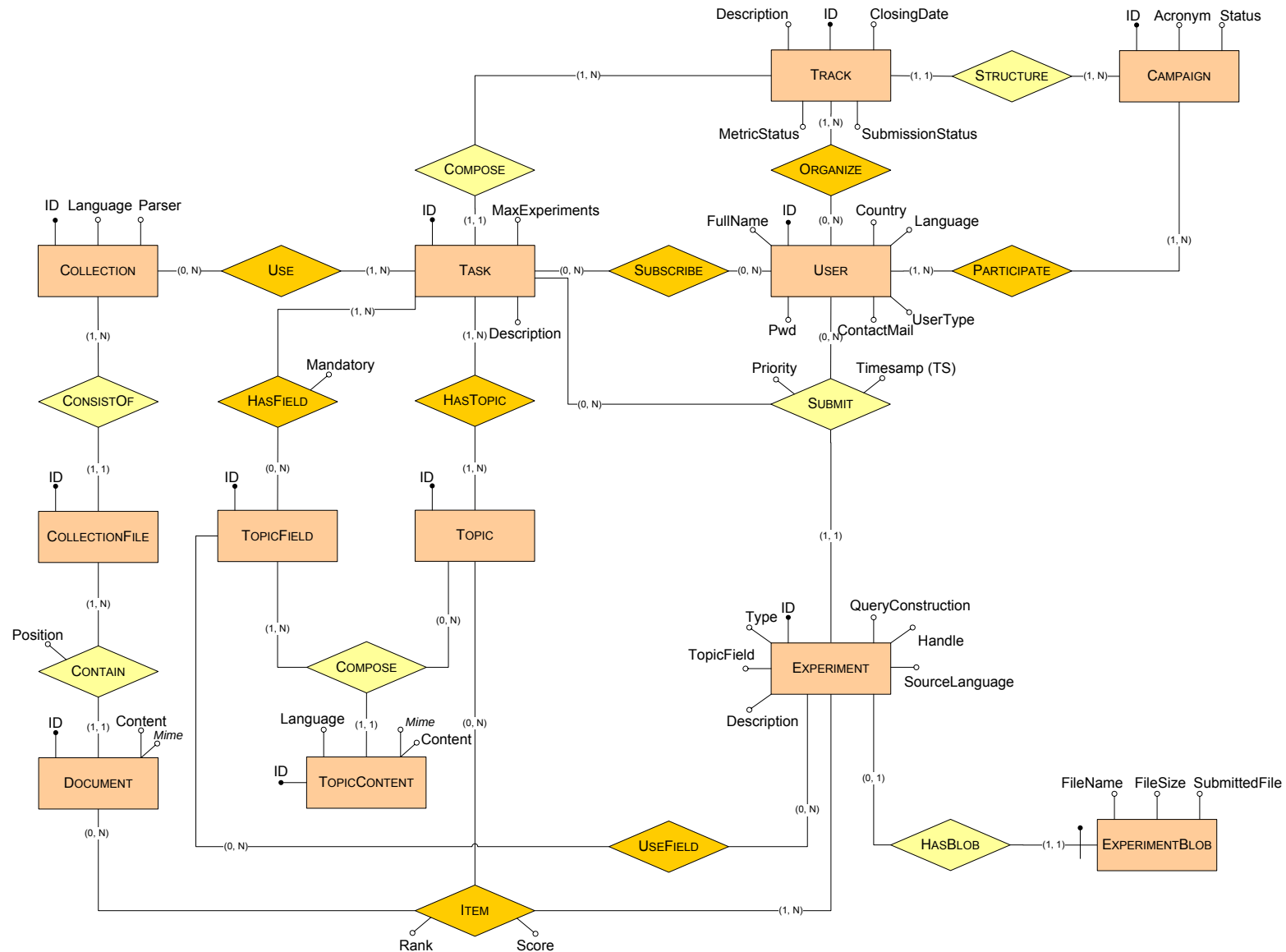


Designing the DLS

- ❑ The design of the scientific DLS is built around five main areas of modelling:
- ❑ **evaluation campaign:**
 - deals with the different aspects of an evaluation forum;
- ❑ **collection:**
 - concerns the different collections made available by an evaluation forum;
- ❑ **experiments:**
 - regards the experiments submitted by the participants and the evaluation metrics computed on these experiments;
- ❑ **pool/relevance assessment:**
 - is about the pooling method where a set of experiments is pooled and the documents retrieved in those experiments are assessed with respect to the topics of the track the experiments belongs to;
- ❑ **statistical analysis:**
 - models the different aspects concerning the statistical analysis of the experimental results.



Conceptual Model





DIRECT: Running Systems Screenshots

❑ Assessor Tasks

- topic creation
- relevance assessments

❑ Participant Tasks

- topics download
- experiment submission
- performance measurements and statistical analyses visualization



DIRECT Topic Creation Interface (1/6)

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) - Selezione dei topic - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://direct.dei.unipd.it/protected/assessor/topicSelection.tiles

Come iniziare Ultime notizie index

Google Vittorio Svegliado padova

Cerca Ortografia Iscriviti Opzioni Vittorio Svegliado padova

Autenticato come dummyuser [Termina sessione]

DIRECT

per il Cross Language Evaluation Forum (CLEF)

DISTRIBUTED INFORMATION RETRIEVAL EVALUATION CAMPAIGN TOOL

Selezione dei topic

Contatti Ultime notizie Pagina principale dell'assessor Aggiorna le informazioni sull'utente Modifica la parola chiave Logout

Stato	Identificativo della track	Descrizione della track
	AH-TOPIC-CREATION-CLEF2007	Ad-Hoc Topic Creation Track
Identificativo del task	Descrizione del task	Numero di topic
AH-EN-TOPIC-CREATION-CLEF2007	Ad-Hoc English Topic Creation Task	50 Scarica i topic
AH-BG-TOPIC-CREATION-CLEF2007	Ad-Hoc Bulgarian Topic Creation Task	50 Scarica i topic
AH-CS-TOPIC-CREATION-CLEF2007	Ad-Hoc Czech Topic Creation Task	50 Scarica i topic
AH-HU-TOPIC-CREATION-CLEF2007	Ad-Hoc Hungarian Topic Creation Task	50 Scarica i topic
AH-FR-TOPIC-CREATION-CLEF2007	Ad-Hoc French Topic Creation Task	50 Scarica i topic
AH-DE-TOPIC-CREATION-CLEF2007	Ad-Hoc German Topic Creation Task	50 Scarica i topic
AH-ES-TOPIC-CREATION-CLEF2007	Ad-Hoc Spanish Topic Creation Task	50 Scarica i topic
AH-IT-TOPIC-CREATION-CLEF2007	Ad-Hoc Italian Topic Creation Task	50 Scarica i topic
AH-AM-TOPIC-CREATION-CLEF2007	Ad-Hoc Amharic Topic Creation Task	50 Scarica i topic
AH-ID-TOPIC-CREATION-CLEF2007	Ad-Hoc Indonesian Topic Creation Task	50 Scarica i topic
AH-OM-TOPIC-CREATION-CLEF2007	Ad-Hoc Oromo Topic Creation Task	50 Scarica i topic
AH-HI-TOPIC-CREATION-CLEF2007	Ad-Hoc Hindi Topic Creation Task	50 Scarica i topic
AH-BN-TOPIC-CREATION-CLEF2007	Ad-Hoc Bengali Topic Creation Task	50 Scarica i topic
AH-TA-TOPIC-CREATION-CLEF2007	Ad-Hoc Tamil Topic Creation Task	50 Scarica i topic
AH-TE-TOPIC-CREATION-CLEF2007	Ad-Hoc Telugu Topic Creation Task	50 Scarica i topic
AH-MR-TOPIC-CREATION-CLEF2007	Ad-Hoc Marathi Topic Creation Task	50 Scarica i topic
AH-ZH-TOPIC-CREATION-CLEF2007	Ad-hoc Chinese Topic Creation Task	50 Scarica i topic
	DS-TOPIC-CREATION-CLEF2007	Domain-Specific Topic Creation Track
Identificativo del task	Descrizione del task	Numero di topic
DS-EN-TOPIC-CREATION-CLEF2007	Domain-Specific English Topic Creation Task	25 Scarica i topic
DS-DE-TOPIC-CREATION-CLEF2007	Domain-Specific German Topic Creation Task	25 Scarica i topic
DS-RU-TOPIC-CREATION-CLEF2007	Domain-Specific Russian Topic Creation Task	25 Scarica i topic
	GC-TOPIC-CREATION-CLEF2007	GeoCLEF Topic Creation Track
Identificativo del task	Descrizione del task	Numero di topic

Trova: Successivo Precedente Evidenzia

Completato



DIRECT Topic Creation Interface (2/6)

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) - Selezione dei topic - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://direct.dei.unipd.it/protected/assessor/topicSelection.tiles

Come iniziare Ultime notizie Index

Google Vittorio Svegliato padova

Cerca Ortografia Iscriviti Opzioni Vittorio Svegliato padova

Autenticato come dummyuser [Termina sessione]

DIRECT per il Cross Language Evaluation Forum (CLEF)

DISTRIBUTED INFORMATION RETRIEVAL EVALUATION CAMPAIGN TOOL

Selezione dei topic

Contatti Ultime notizie Pagina principale dell'assessor Aggiorna le informazioni sull'utente Modifica la parola chiave Logout

Stato Identificativo della track Descrizione della track

Identificativo del task AH-EN-TOPIC-CREATION-CLEF2007 Descrizione del task Ad-Hoc English Topic Creation Task Numero di topic 50 Scarica i topic

Identificativo del topic	Riassunto del topic	Rilevanti	Non rilevanti	
401	Find documents about rises in prices after the introduction of the Euro.	0	0	Modifica il topic
402	Find documents reporting on the exploitation of renewable resources for power/energy production.	0	0	Modifica il topic
403	Find documents that give information about actors playing police roles in films or television.	0	0	Modifica il topic
404	Find reports that describe the on-site security measures at any NATO summit.	0	0	Modifica il topic
405	What conditions can trigger asthma in children?	0	0	Modifica il topic
406	Find documents about animated cartoons that have been considered for a prize or award.	0	0	Modifica il topic
407	Who was the Australian Prime Minister in 2002?	0	0	Modifica il topic
408	We seek any information on human cloning including claims of the production of the first human clone.	0	0	Modifica il topic
409	Retrieve documents reporting the nationality of the victims of a terrorist car bomb in Bali, Indonesia, in 2002.	0	0	Modifica il topic
410	Find documents providing evidence that North Korea is violating the 1994 nuclear agreement to freeze its nuclear weapons program.	0	0	Modifica il topic
411	What movie was awarded the Oscar for best film in March 2002?	0	0	Modifica il topic
412	Find documents that discuss books on politicians.	0	0	Modifica il topic
413	Find documents mentioning factors that reduce the risk of the onset of diabetes.	0	0	Modifica il topic
414	Find documents describing beer festivals such as the Oktoberfest or similar.	0	0	Modifica il topic
415	Find documents that discuss any efforts to solve or control the phenomenon of drug abuse and addiction.	0	0	Modifica il topic
416	How did the Russian authorities end the Moscow theater crisis in which Chechen Rebels held about 700 hostages?	0	0	Modifica il topic
417	We seek documents on both successful and unsuccessful hijacking of planes and information on the airlines or countries involved.	0	0	Modifica il topic
418	Find documents that report the Turkish Prime Minister Bülent Ecevit's public statements.	0	0	Modifica il topic
419	Retrieve documents that discuss some advantages or disadvantages of using designated controlled and centralised sites for the storage of nuclear waste.	0	0	Modifica il topic
420	What health problems are possibly caused at least partly by overweight or obesity?	0	0	Modifica il topic

Trova: Successivo Precedente Evidenzia

Completato



**The NTCIR Workshop 6 Meeting
Tokyo, May 15-18, 2007**



DIRECT Topic Creation Interface (4/6)

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) - Creazione dei topic - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://direct.dei.unipd.it/protected/assessor/topicCreation.tiles

Come iniziare Ultime notizie index

Google Vittorio Svegliado padova Cerca Ortografia Iscriviti Opzioni Vittorio Svegliado padova

DIRECT Indietro

Topic << Topic successivo

GeoCLEF English Topic Creation Task
Informazioni sul topic 101:

Titolo:
Oil and gas extraction found between the UK and the Continent

Description:
To be relevant documents describing oil or gas production between the UK and the European continent will be relevant

Narrative:
Oil and gas fields in the North Sea will be relevant.

Aggiorna topic
Svuota topic

Query Cerca Salva l'interrogazione come una nota [Sintassi dell'interrogazione](#)

Note

Titolo	Body	Utente	Data e ora	Rispondi

Trova: Successivo Precedente Evidenzia

Completato



DIRECT Topic Creation Interface (5/6)

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) - Creazione dei topic - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://direct.dei.unipd.it/protected/assessor/topicCreation.tiles

Come iniziare Ultime notizie Index

Google Vittorio Svegliado padova

Cerca Ortografia Iscriviti Opzioni Vittorio Svegliado padova

DIRECT Indietro

Topic << Topic successivo

GeoCLEF English Topic Creation Task
Informazioni sul topic 101:

Titolo:
Oil and gas extraction found between the UK and the Continent

Description:
To be relevant documents describing oil or gas production between the UK and the European continent will be relevant

Narrative:
Oil and gas fields in the North Sea will be relevant.

Aggiorna topic
Svuota topic

Query oil gas extraction [Cerca] Salva l'interrogazione come una nota Sintassi dell'interrogazione

Rango	Identificativo del documento	Rilevanza	Titolo del documento	Visualizza
0	LA080194-0090	Non ancora valutato	<P> PACIFIC WATCH: VIETNAM </P>	View Document
1	GH950330-000124	Non ancora valutato	North Sea record	View Document
2	GH950104-000112	Non ancora valutato	Oil drilling at castle awaits council stamp	View Document
3	LA090694-0185	Non ancora valutato	<P> PROPERTY VALUES: A WEEKLY COLUMN BY RON GALPERIN: OIL ROYALTIES COULD PROVE MIXED BLESSING; RESOURCES: OFFICIALS COULD SEEK TO MAKE PROPERTY OWNERS WHO HAVE RECEIVED THE FUNDS HELP PAY FOR CLEANUP. </P>	View Document
4	GH950909-000120	Non ancora valutato	Offshore oil and gas firms told to embrace joint development	View Document
5	LA012694-0237	Non ancora valutato	<P> BRIEFING BOOK /CHARLES P. WALLACE: INTERNATIONAL BUSINESS; VIETNAMESE OFFICIALS EAGER FOR U.S. FIRMS TO JOIN SEARCH FOR OIL </P>	View Document
6	GH950809-000099	Non ancora valutato	Manufacturing picks up as North Sea output falls	View Document
7	GH950227-000128	Non ancora valutato	Oil field project unshaken by Kobe earthquake	View Document
8	GH950501-000106	Non ancora valutato	Scottish Enterprise targets Caspian oil	View Document
9	GH950930-000140	Non ancora valutato	Fears for Solway wildlife as oil hunt begins	View Document
10	LA120594-0168	Non ancora valutato	<P> DRILLING ADVANCE REKINDLES SANTA BARBARA OIL WARS; ENERGY: MOBIL WANTS TO REMOVE OFFSHORE RIG, USE NEW METHOD TO BORE BENEATH OCEAN FROM COAST. FOES CRY FOUL. </P>	View Document
11	GH951201-000114	Non ancora valutato	Lecturer queries oil firms' safety figures	View Document
12	LA061194-0072	Non ancora valutato	<P> PAYING A ROYALTY FOR MINING </P>	View Document
13	GH950821-000005	Non ancora valutato	Light week for fresh statistics	View Document
14	LA091994-0161	Non ancora valutato	<P> VIETNAM HOPES TO REFINES ITS OIL INDUSTRY; ENERGY: PETROLEUM IS THE COUNTRY' S BIGGEST EXPORT, BUT FOR NOW IT MUST DEPEND ON SOVIET-AGE EQUIPMENT TO DO THE JOB. </P>	View Document

Note

Aggiungi Nota

Titolo	Body	Utente	Data e ora	Rispondi

Aggiungi Nota

Trova: Successivo Precedente Evidenzia

Completato



DIRECT Topic Creation Interface (6/6)

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) - Creazione dei topic - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://direct.dei.unipd.it/protected/assessor/topicCreation.tiles

Come iniziare Ultime notizie index

Google Vittorio Svegliato padova

Cerca Ortografia Iscriviti Opzioni Vittorio Svegliato padova

DIRECT Indietro

Topic << >>

Documento precedente **Rilevante** Ritorna alla ricerca Non rilevante Documento Successivo

GeoCLEF English Topic Creation Task
Informazioni sul topic 101:

Titolo:
Oil and gas extraction found between the UK and the Continent

Description:
To be relevant documents describing oil or gas production between the UK and the European continent will be relevant

Narrative:
Oil and gas fields in the North Sea will be relevant.

Aggiorna topic
Svuota topic

Identificativo del documento
LA080194-0090

Rilevanza
Non ancora valutato

Titolo del documento
<P> PACIFIC WATCH: VIETNAM </P>

Corpo del documento
<DOCNO>LA080194-0090</DOCNO>
<DOCID> 066024 </DOCID>
<SOURCE>
Los Angeles Times
</SOURCE>
<DATE>
August 1, 1994, Monday, Home Edition
</DATE>
<SECTION>
Business; Part D; Page 5; Column 5; Financial Desk
</SECTION>
<LENGTH>
108 words
</LENGTH>
<HEADLINE>
PACIFIC WATCH: VIETNAM
</HEADLINE> <BYLINE>
From Times Staff and Wire Reports

Note

Titolo	Body	Utente	Data e ora	Rispondi

Aggiungi Nota

Aggiungi Nota

Trova: Successivo Precedente Evidenzia

Completato



DIRECT Assessor Interface (1/2)

DIRECT
DISTRIBUTED INFORMATION RETRIEVAL EVALUATION CAMPAIGN TOOL

for Cross Language Evaluation Forum (CLEF)

assessor

TORINO P.N. LIVORNO-PISA GENOVA 9:46 21 TORINO P.N.
MINTUR. SCAUR REG 9:49 11 MINTUR. SCAUR
FRASCATI REG 9:52 18 FRASCATI
FIUMICINO A. AEROPORTO NON STOP DIR 9:52 29 FIUMICINO

Logged as petya [Logout]

Assessments Messages

Reload Pools

Assessments Status	Pool Name	Pool Description	Assessed Documents / Total Documents	
	AH-BULGARIAN-CLEF2006	Bulgarian Pool for AH-MONO-BG, and AH-BILI-X2BG	17 308 / 17 308	
	Topic ID	Topic Summary	Total Relevant Not Relevant Assessed	Assess Topic
	301	Какви продукти на Нестле се продават по света?	487 2 485 0	
	302	Намерете документи, които описват или дискутират въздействието на бойкоти от страна на потребителите.	413 6 407 0	
	303	Намерете информация за места, където са изложени италиански картини.	367 9 358 0	
	304	Дайте имената и/или нахождението на местата, които са били посочени от ЮНЕСКО за места на световното наследство с изключителна красота или важност.	402 14 388 0	
	305	Коя е настоящата цена на нефта?	436 16 420 0	
	306	Намерете документи, описващи дейността на баската терористична група ETA във Франция.	330 3 327 0	
	307	Намерете каквато и да е информация за филми, чието действие се развива в Шотландия.	274 3 271 0	
	308	Намерете документи, които споменават за пълно или частично слънчево затъмнение.	483 11 472 0	
	309	Какви са опасностите за здравето от нелегалните "твърди" наркотици, като например хероин и кокаин, в сравнение с "леките" наркотици?	250 25 225 0	
	310	Намерете документи, които описват методи, използвани за обработката или съхранението на индустриалните отпадъци.	321 8 313 0	
	311	Намерете информация и цифри за нивото на безработица в европейските страни.	259 5 254 0	
	312	Какви са последствията от нападенията на кучета над хора?	299 26 273 0	
	313	Намерете документи, описващи чествания по случай 100-годишнината на някое събитие.	318 42 276 0	
	314	Какви мерки се вземат в Европа, за да се защитят животните или растенията, които са изложени на риска да изчезнат?	490 30 460 0	
	315	В кои спортове се използват непозволені лекарствени средства?	536 66 470 0	



DIRECT Assessor Interface (2/2)

[Back to Topic Page](#)[Find](#)[Clear Search](#)

Relevant

<<

Document

>>

<<

Relevant Document

>>

[Log Out](#)☐ case-sensitive

Not Relevant

<<

Not yet assessed

>>

<<

Not Relevant Document

>>

Title:
<BG-title> Цените на нефта </BG-title>
Description:
<BG-desc> Коя е настоящата цена на нефта? </BG-desc>
Narrative:
<BG-narr> Само документи, които дават настоящата цена на нефта, са подходящи, т.е. валидна е цената при писането на документа. Препратки към минали цени на нефта или към прогнози за бъдещи промени в цените не са подходящи. </BG-narr>

<DOCID>SEGA20021228-040</DOCID><DOCNO>SEGA20021228-040</DOCNO>
<TITLE> </TITLE>
<AUTHOR> </AUTHOR>
<DATE> 28 /12 /2002 </DATE>
<RUBRIC> Икономика </RUBRIC>
<LEAD> Венецуела изстреля цената на нефта до рекорд от две години </LEAD>
<LEAD> 32, 6 долара искат за барел с доставка през февруари </LEAD>
<TEXT>
Кризата във Венецуела изстреля през последните дни цената на петрола до 32, 55 долара за барел (стандартен варел от 159 литра),което е поскъпване с повече от половин долар в сравнение с цената преди Коледа от 31, 97 долара, предаде Ройтерс. Сравнението бе направено вчера на борсата в Ню Йорк за сделки, сключвани за доставка през февруари. Това е най-високата цена, достигана след януари 2001 г. В Лондон бе отбелязано увеличение също с половин долар и цената стигна 30, 15 долара за барел.

Петролната треска се основава на несигурността, породена от наближаващата война срещу Ирак, но за катализатор ѝ служи кризата във Венецуела, която заедно със Саудитска Арабия и Мексико е най-големият доставчик на петрол за Съединените щати. На нея се падат 13 %от вноса на суров нефт в САЩ.

Ръководството на най-голямата венецуелска държавна петролна компания - " Петролеос де Венесуела " ,гласува в четвъртък да продължи общата стачка, започнала на 2 декември, и да блокира доставките на петрол до оставка на президента Уго Чавес. Той бе избран през 1998 г. и оцеля след опит за преврат през април т. г. Бившият парашутист, който преди 10 години оглави неуспешен преврат, заяви, че няма да свика предсрочни избори, и обвини своите противници, че се стремят към незаконно завземане на властта. Опозицията твърди, че неговата " революция " в полза на бедните и популистката му реторика подкопават икономиката и разпалват класова омраза.

Венецуела е на седмо място в света по запаси от петрол, изчислявани на 48 милиарда барела - над два пъти повече от природните запаси на Съединените щати, които се нареждат на 12-о място с 22 милиарда барела.

РЕЗЕРВ

Съединените щати, които преди повече от година започнаха да попълват стратегическите си запаси от петрол заради намерението си да нападнат Ирак, вече са складирали 287 милиона барела суров нефт, съобщи лондонският в. " Гардиан " . Те използват за целта гигантски подземни резервоари, образували се след прекратяване на добива в стари мини. През последните дни са били закупени още 2, 7 милиона барела. Ако войната срещу Ирак предизвика

SEGA20021228-040
RELEVANT

Document 54 / 436

SEGA20021228-041
NOT_RELEVANT

Document 55 / 436

SEGA20021228-043
NOT_RELEVANT

Document 56 / 436

View

View

View

View

View

DELUG Network of Excellence for Digital Libraries under the Sixth Framework Programme of the European Commission
DIRECT - Copyright © 2005-2009 Information Management Systems (IMS) Research Group, Department of Information Engineering (DIE), University of Padua

The NTCIR Workshop 6 Meeting
Tokyo, May 15-18, 2007

M. Agosti, G.M. Di Nunzio, N. Ferro, C. Peters

35



DIRECT Participant Interface (1/4)

Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) - Main - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti ?

http://direct.dei.unipd.it/login.do

Come iniziare Ultime notizie index

Google Vittorio Svegliado padova

Cerca Ortografia Iscriviti Opzioni Vittorio Svegliado padova

DIRECT for the Cross Language Evaluation Forum (CLEF)
DISTRIBUTED INFORMATION RETRIEVAL EVALUATION CAMPAIGN TOOL

Logged as dummypart [Logout]

Main

Contacts News Participant Main Page Update User Information Change Password Logout

✓ User dummypart successfully logged in.

★ Track Identifier	Track Description	Topic Status	Submission Status	Metrics Status
AH-BILI-CLEF2007	Ad-Hoc Bilingual Track	Available	Not started yet	Not started yet
Task Identifier	Task Description			
AH-BILI-X2BG-CLEF2007	Ad-Hoc Bilingual Bulgarian Task		Download Topics	
AH-BILI-X2CS-CLEF2007	Ad-Hoc Bilingual Czech Task		Download Topics	
AH-BILI-X2EN-CLEF2007	Ad-Hoc Bilingual English Task		Download Topics	
AH-BILI-X2HU-CLEF2007	Ad-Hoc Bilingual Hungarian Task		Download Topics	
AH-MONO-CLEF2007	Ad-Hoc Monolingual Track	Available	Not started yet	Not started yet
Task Identifier	Task Description			
AH-MONO-BG-CLEF2007	Ad-Hoc Monolingual Bulgarian Task		Download Topics	
AH-MONO-CS-CLEF2007	Ad-Hoc Monolingual Czech Task		Download Topics	
AH-MONO-HU-CLEF2007	Ad-Hoc Monolingual Hungarian Task		Download Topics	
AH-MONO-EN-CLEF2007	Ad-Hoc Monolingual English Task (ONLY for English Pool creation and AH-BILI-X2EN participants)		Download Topics	
AH-ROBUST-BILI-CLEF2007	Ad-Hoc Robust Bilingual Track	Available	Not started yet	Not started yet
Task Identifier	Task Description			
AH-ROBUST-BILI-X2FR-TEST-CLEF2007	Ad-Hoc Robust Bilingual Test Task, French target collection(s)		Download Topics	
AH-ROBUST-BILI-X2FR-TRAINING-CLEF2007	Ad-Hoc Robust Bilingual Training Task, French target collection(s)		Download Topics	
AH-ROBUST-MONO-CLEF2007	Ad-Hoc Robust Monolingual Track	Available	Not started yet	Not started yet
Task Identifier	Task Description			
AH-ROBUST-MONO-FR-TEST-CLEF2007	Ad-Hoc Robust Monolingual French Test Task		Download Topics	
AH-ROBUST-MONO-EN-TEST-CLEF2007	Ad-Hoc Robust Monolingual English Test Task		Download Topics	
AH-ROBUST-MONO-PT-TEST-CLEF2007	Ad-Hoc Robust Monolingual Portuguese Test Task		Download Topics	

Trova: Successivo Precedente Evidenzia

Completato



DIRECT Participant Interface (2/4)

C:\Documents and Settings\Administrator\Desktop\Bulgarian_topics_AH-MONO-BG-CLEF2007.xml - Windows Internet Explorer

C:\Documents and Settings\Administrator\Desktop\Bulgarian_topics_AH-MONO-BG-CLEF2007.xml

Google C Cerca RS Segnalibri 0 bloccati Controlla Invia Impostazioni

Ricerca Web Tv Radio DeeJay America We7, canzoni Video Rss [492 new] Arcade Chat News [5 new] E-mail Notifier 15°C

C:\Documents and Settings\Administrator\D...

Topic Identifiers are DOI

```
<?xml version="1.0" encoding="UTF-8" ?>
- <topics>
- <top lang="bg">
  <num>10.2452/401-AH</num>
  <title>Инфлацията на еврото</title>
  <desc>Намерете документи за повишаването на цените след въвеждане на еврото.</desc>
  <narr>Подходящ е всеки документ, който дава информация за повишаването на цените в която и да е държава, въвела единната европейска валута.</narr>
</top>
- <top lang="bg">
  <num>10.2452/402-AH</num>
  <title>Възобновяеми източници на енергия</title>
  <desc>Намерете документи, които дават информация за възобновяеми източници, като например водна, слънчева, вятърна, геотермална енергия и енергия, получена от растителна и животинска биомаса. Информацията не е подходяща.</narr>
</top>
- <top lang="bg">
  <num>10.2452/403-AH</num>
  <title>Филмова роля на ченге</title>
  <desc>Намерете документи, даващи информация за актьори, които играят полицаи във кинофилми или в телевизията.</desc>
  <narr>Подходящите документи ще споменават заглавието на филма, пиесата или сериала, както и името на актьора или актрисата, които са в ролята на полицаи/полицайка.</narr>
</top>
- <top lang="bg">
  <num>10.2452/404-AH</num>
  <title>Сигурността на срещите на най-високо равнище на НАТО</title>
  <desc>Намерете информация, която описва локалните мерки за сигурност на която и да е среща на най-високо равнище на НАТО.</desc>
  <narr>Търси се информация, която се отнася до мерките за сигурност, взети в страните, където се провеждат срещите на най-високо равнище на НАТО по света. Информация, която се отнася до инциденти, демонстрации и протестни акции срещу срещите на най-високо равнище, не е подходяща.</narr>
</top>
- <top lang="bg">
  <num>10.2452/405-AH</num>
  <title>Астма в детска възраст</title>
  <desc>Кои предпоставки могат да предизвикат астма у децата?</desc>
  <narr>Подходящите документи ще цитират някои от основните фактори, за които се смята, че са отговорни за пристъпите на астма у децата.</narr>
</top>
- <top lang="bg">
  <num>10.2452/406-AH</num>
  <title>Анимационни филми</title>
  <desc>Намерете документи за анимационни филми, които са били предлагани за награда.</desc>
  <narr>Подходящите документи споменават поне един анимационен филм, който или е бил представян на филмов фестивал с надеждата да спечели награда, или е бил предлаган за награда без връзка с конкретен фестивал.</narr>
</top>
- <top lang="bg">
  <num>10.2452/407-AH</num>
  <title>Австралийският министър-председател</title>
  <desc>Кой беше австралийският министър-председател през 2002 година?</desc>
  <narr>Подходящите документи трябва да споменават името на действащия през 2002 г. министър-председател на Австралия.</narr>
</top>
</topics>
</xml>
```

Fine

Risorse del computer 100%



DIRECT Participant Interface (3/4)

DIRECT
DISTRIBUTED INFORMATION EVALUATION CAMPAIGN TOOL

for Cross Language Evaluation Forum (CLEF)

participant

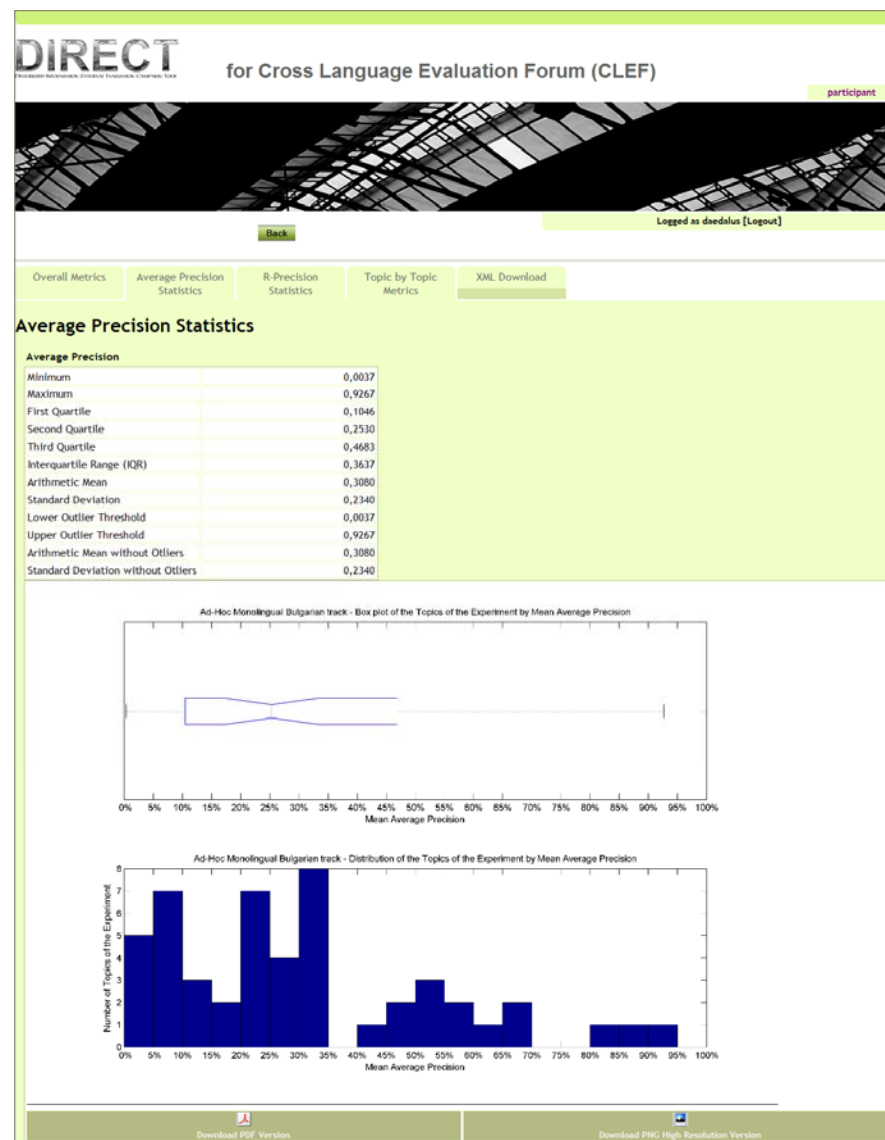
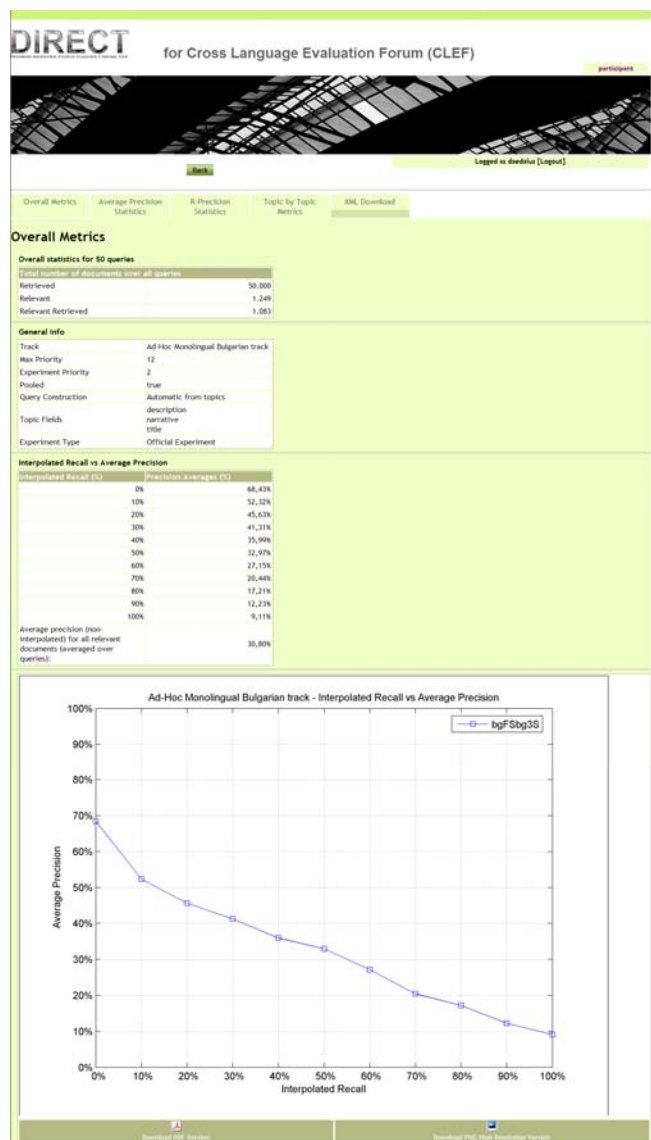
Logged as daedalus [Logout]

Main

Messages

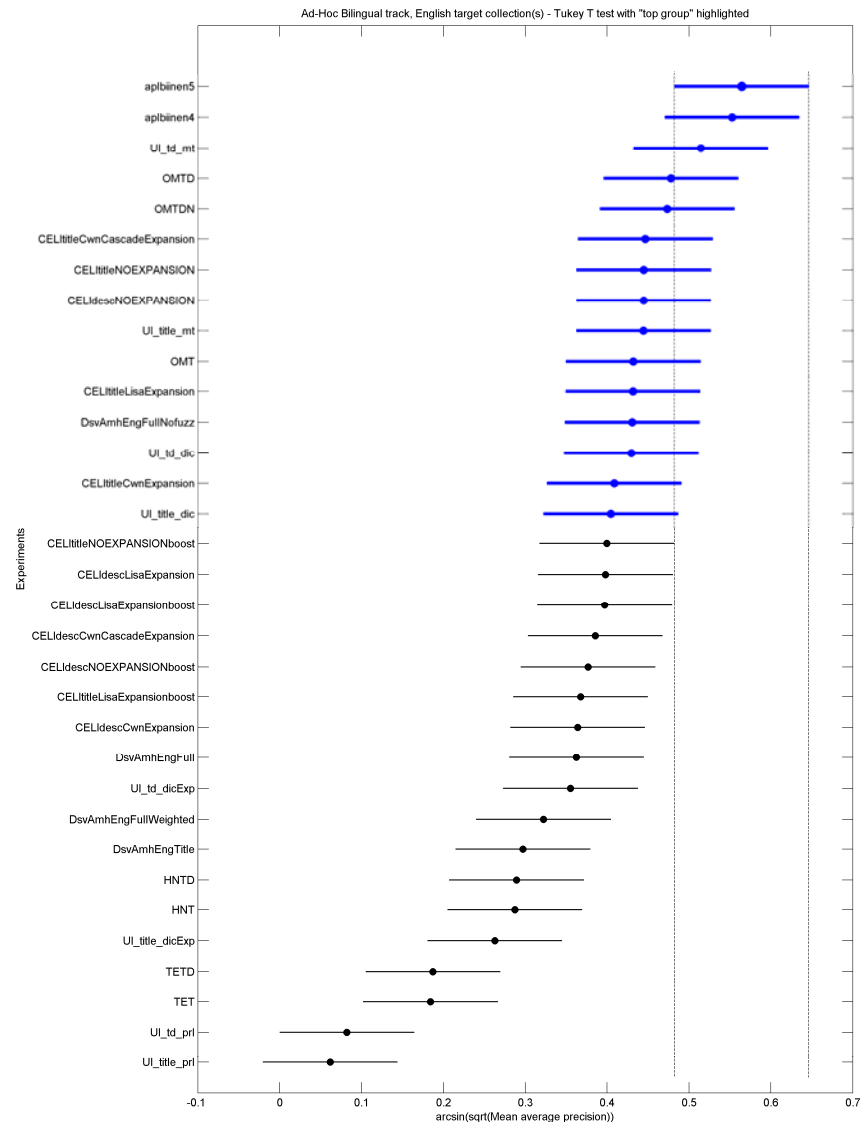
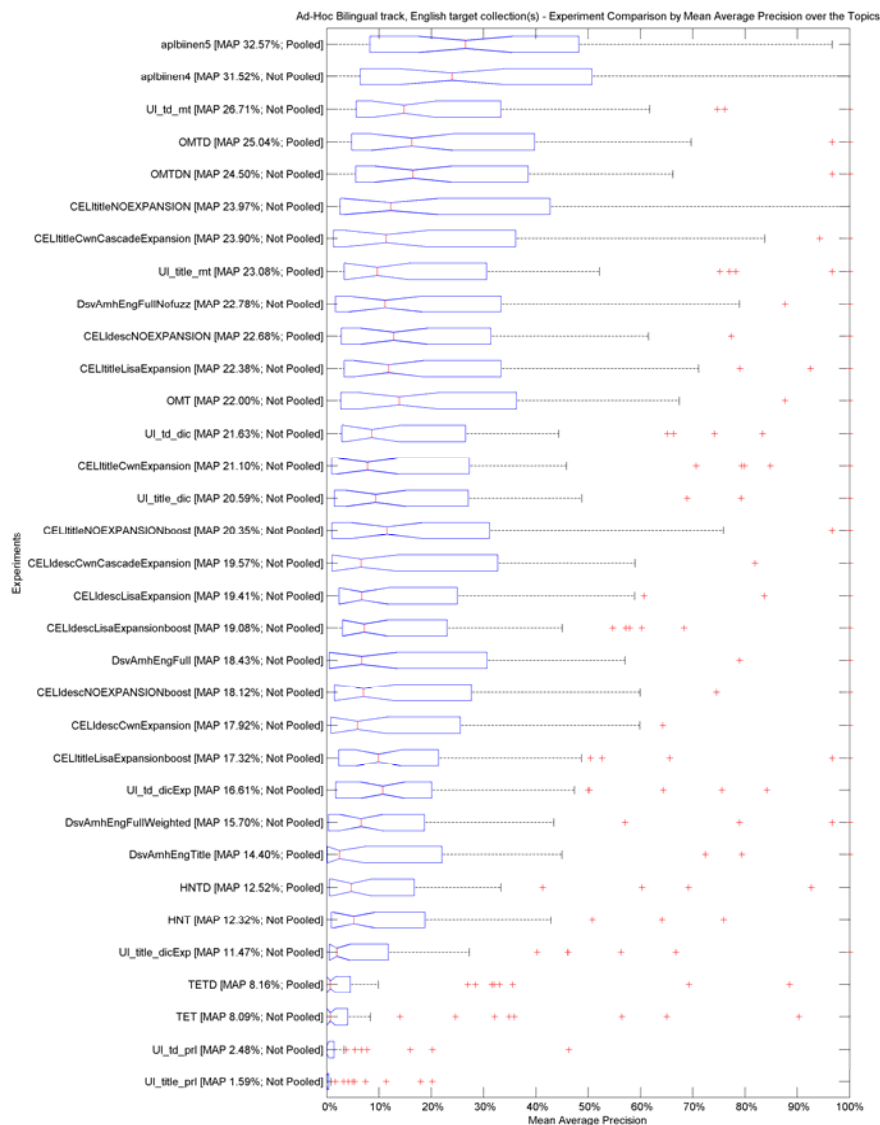
Submission Status	Track group	Submission Closing Date	Assessments and Metrics
-------------------	-------------	-------------------------	-------------------------

	AH-BILI-CLEF2006 - Ad-Hoc Bilingual task	04-jun-2006	Available	↓
	AH-MONO-CLEF2006 - Ad-Hoc Monolingual task	04-jun-2006	Available	↑
AH-MONO-BG-CLEF2006		Ad-Hoc Monolingual Bulgarian track	Download Assessments	↑
			Download Track Info as XML	
Run ID	Timestamp	Priority	Topic Fields	Description
bgFSbg3S View Run Info View Experiment Metrics	viernes 2 de junio de 2006 at 00H22' CEST	2	description narrative title	TDN run
bgFSbg2S View Run Info View Experiment Metrics	viernes 2 de junio de 2006 at 00H20' CEST	3	description title	Baseline mandatory TD run
bgx011bgFW4FS4 View Run Info View Experiment Metrics	viernes 2 de junio de 2006 at 00H24' CEST	1	description narrative title	Asymmetric combination of TDN runs
AH-MONO-EN-CLEF2006		Ad-Hoc Monolingual English track	Download Assessments	↓
			Download Track Info as XML	





Statistical Analyses



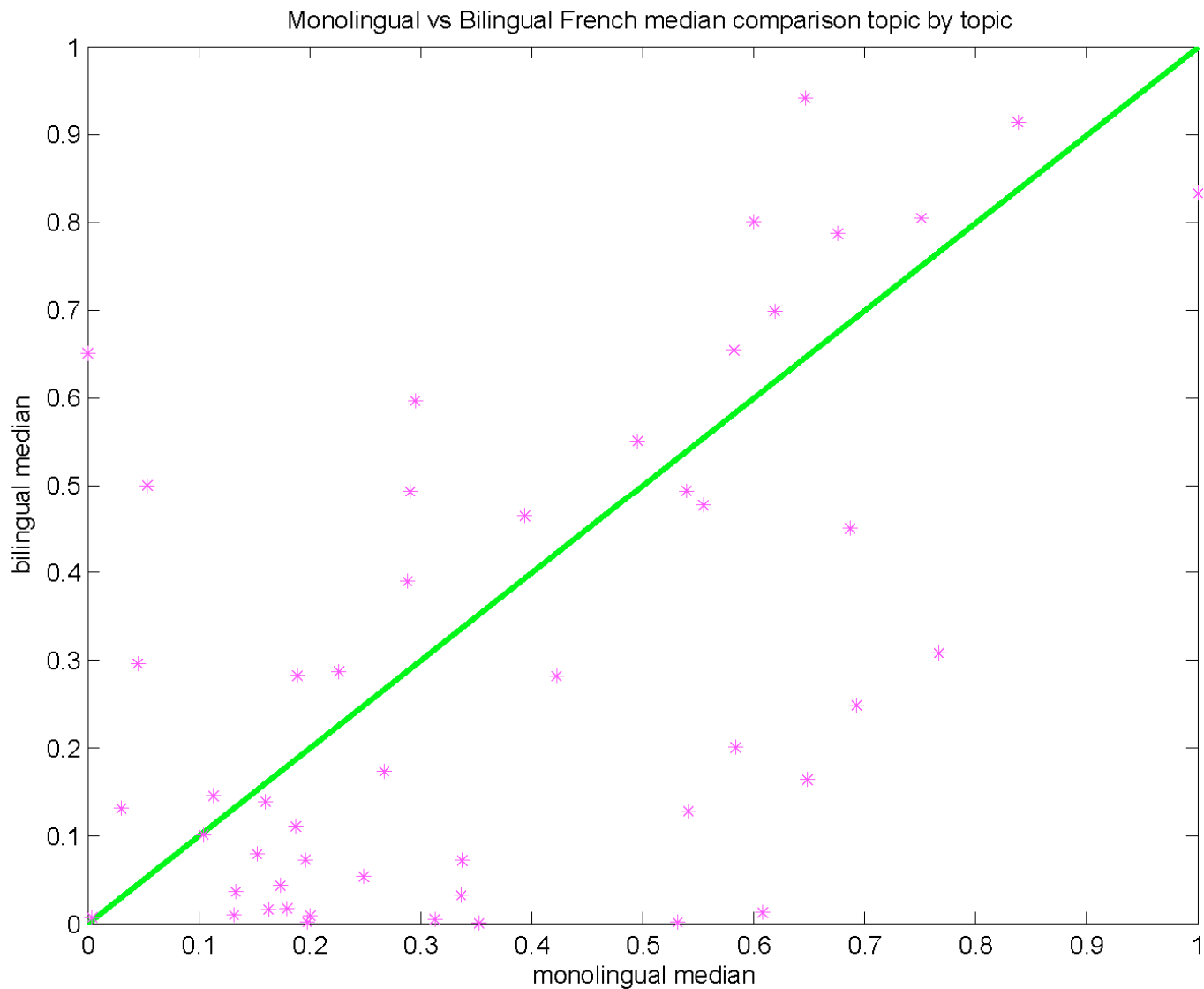


...and More Statistical Analyses

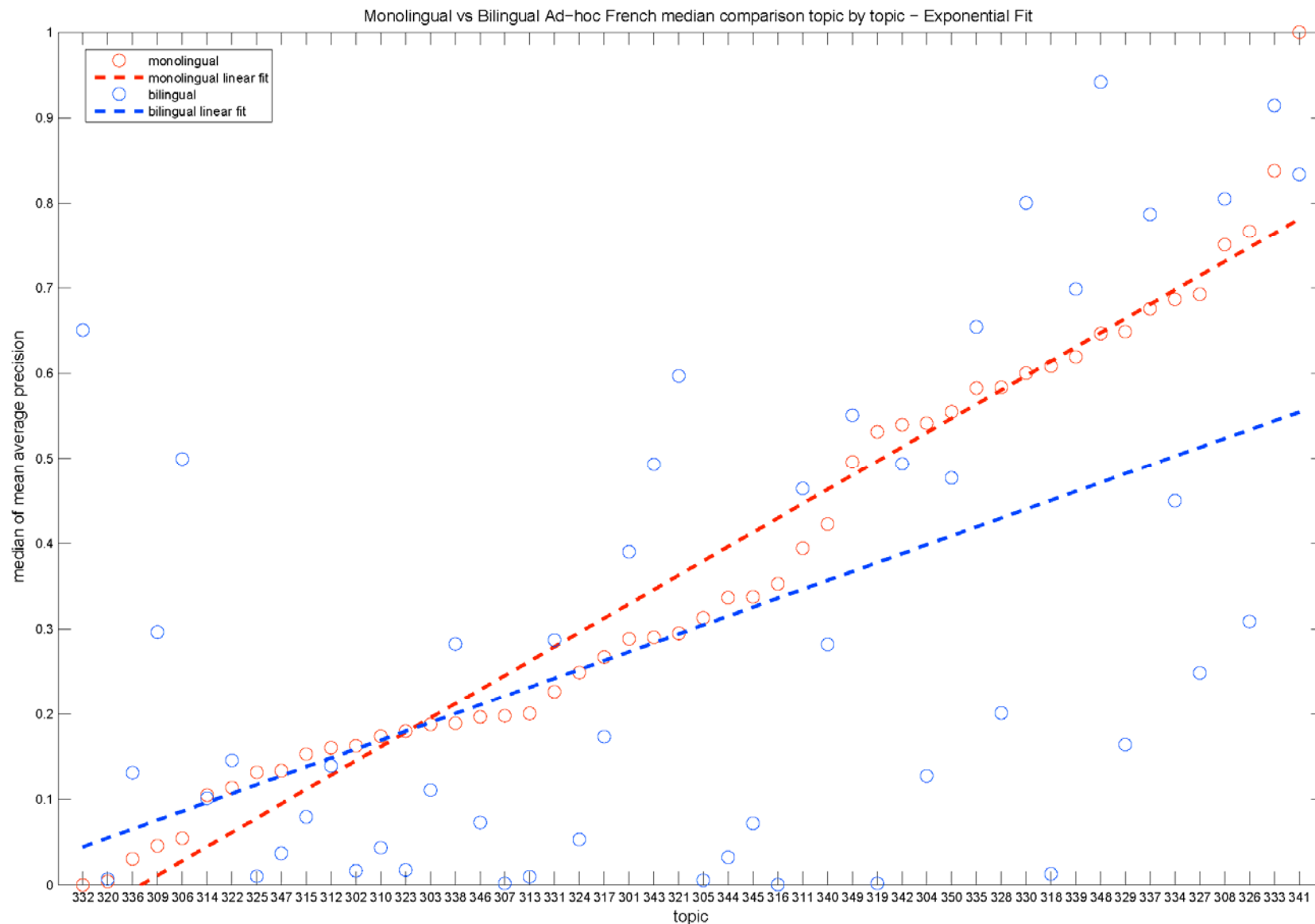
- ❑ In order to improve the current way of comparing bilingual and monolingual retrieval and provide better methods and tools for assessing the performances we need:
 - more sophisticated analysis techniques to assess performances;
 - effective methods and tools for communicating the results.

- ❑ Given the average performance for each single topic of the monolingual and bilingual task, we want to study the distribution of these performances and employ different statistical tests to verify the following conditions:
 - the distributions of the performances are similar.
 - the variances of the two distribution are similar.
 - the mean of the two distributions are different.

Mono- vs Bili Comparison (1/2)



Mono- vs Bili Comparison (2/2)





For further information see:
<http://direct.dei.unipd.it>

or contact:
E-mail: direct@dei.unipd.it