



## Task 7.4

# CLEF: An Evaluation Framework for Multilingual Information Access in Digital Libraries

Carol Peters  
ISTI-CNR

Giorgio Di Nunzio and Nicola Ferro  
University of Padua

# Cross Language Evaluation Forum: Objectives



Promote research and stimulate development of multilingual IR systems for European languages

- Organisation of regular evaluation campaigns for system testing
- Building of an MLIA/CLIR research community
- Construction of reusable test-suites

Conduct research in CLIR evaluation

- design of new evaluation methodologies for CLIR system testing
- development of a scientific infrastructure to support new experimental analyses

Encourage take-up by DL applications

# Cross Language Evaluation Forum: Organization



- Partners:
  - ISTI-CNR (Coordinator)
  - UNIPD
  - SICS
- Mainly dependent on voluntary efforts
- Coordination is distributed for language and for task

## Main Institutions involved in Coordination

- |                            |                             |                        |
|----------------------------|-----------------------------|------------------------|
| ■ Bulgarian Acad. Sciences | ■ Hungarian Acad. Science   | ■ U. Amsterdam, NL     |
| ■ CELCT, Trento, Italy     | ■ IZ- Bonn, Germany         | ■ UC Berkeley, USA     |
| ■ DCU, Ireland             | ■ Linguateca Sintef, Norway | ■ U. Hospitals Geneva  |
| ■ DFKI, Germany            | ■ LSI-UNED, Spain           | ■ U. Limerick, Ireland |
| ■ ELRA/ELDA, France        | ■ Moscow State U. Russia    | ■ U. Padova, Italy     |
| ■ ISTI-CNR, Pisa, Italy    | ■ NIST, USA                 | ■ U. Maryland, USA     |
| ■ ITC-irst, Trento, Italy  | ■ Oregon Health & Sci U.    | ■ U. Sheffield, UK     |

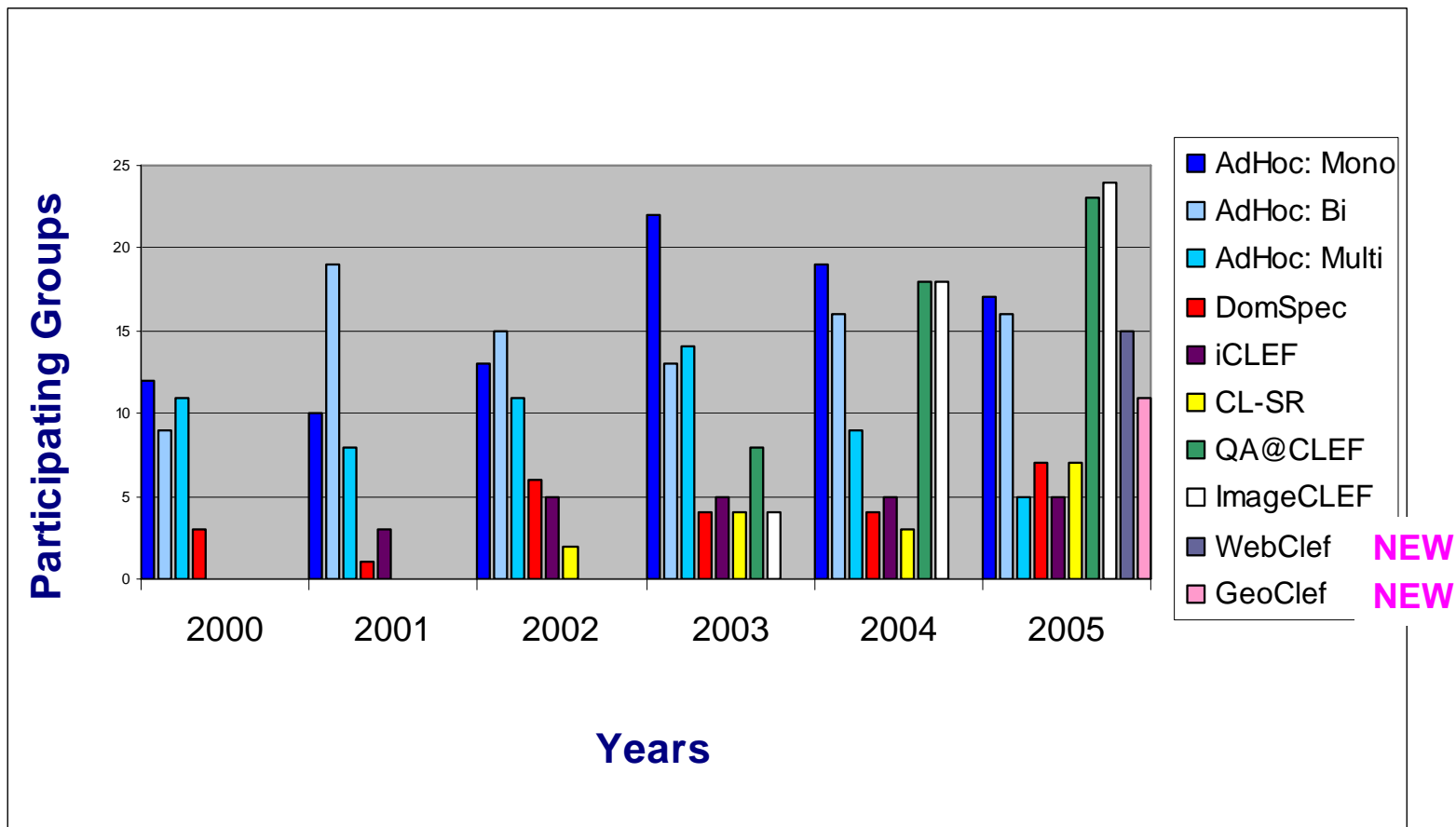
# Organization of Evaluation Campaigns



CLEF 2005 offered eight tracks designed to evaluate the performance of systems for:

- mono-, bi- and multilingual text retrieval (Ad-hoc)
- scientific document retrieval (Domain-specific)
- interactive cross-language retrieval (iCLEF)
- multiple lang. question answering (QA@CLEF)
- cross-lang. retrieval on image collections (ImageCLEF)
- cross-lang. speech retrieval (CL-SDR)
- multilingual web retrieval (WebCLEF) **(New)**
- cross-lang. geographic retrieval (Geo CLEF) **(New)**

# Promotion of Research

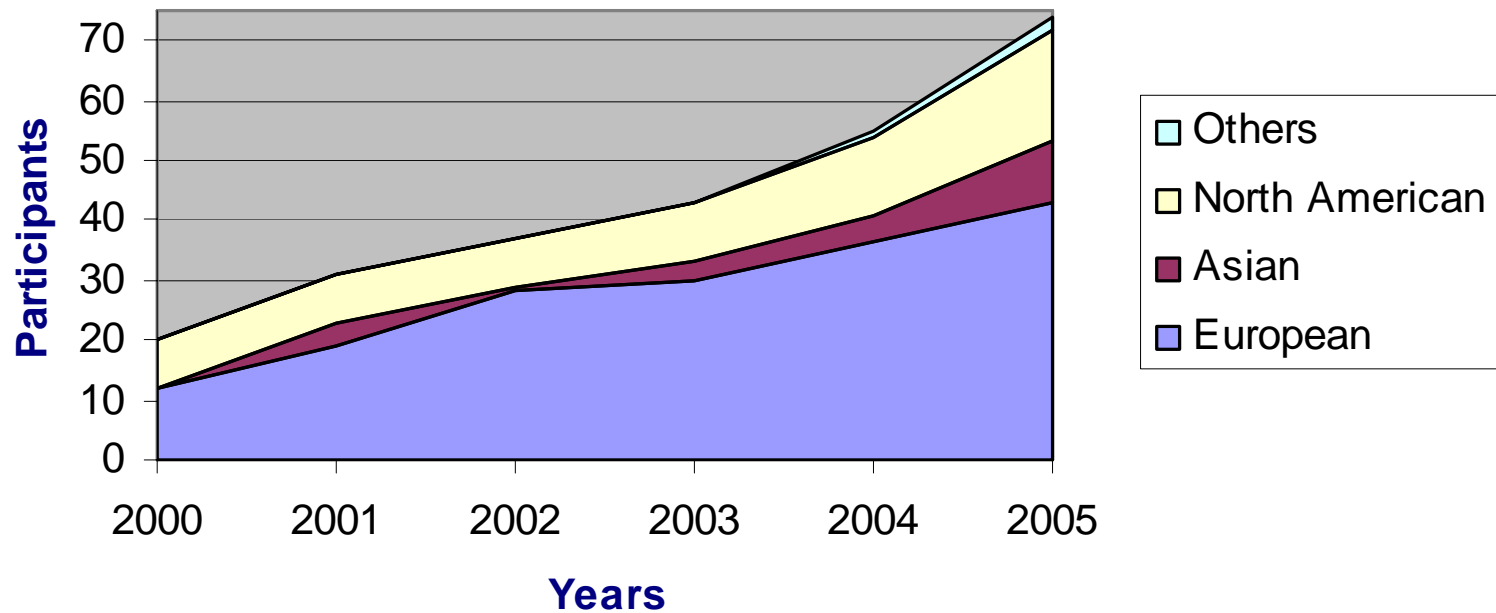


**From cross-language textual document retrieval towards all aspects of multilingual multimedia information access**

# Creation of a MLIA/CLIR Community



**CLEF 2000-2005**



# Creation of Reusable Test Collections



## Text

- CLEF multilingual comparable corpus - more than 2M news docs in 12 languages
- The GIRT-4 social science EN/DE database : more than 300,000 docs
- The Russian Social Science Corpus: almost 100,000 docs

## Image

- St Andrews historical photographic archive: 28,000 images
- CasImage radiological medical database: 9000 images with FR&EN notes
- IRMA collection in EN and DE for automatic medical image annotation: 10,000 images

## Speech

- Malach collection of spontaneous conversational speech derived from the Shoah archives: 589 hours

## Web

- EuroGOV, a multilingual collection of more than 2M webpages crawled from European governmental sites

**Topics for query/question formulation in a large number of languages plus sets of relevance assessments**

# CLEF 2005: Research in New Directions I



- the **ad-hoc track** offered a task aimed at studying in-depth the problem of results merging over collections/over languages and at measuring progress in multilingual IR system development over time;
- the **interactive track** was devoted to the comparative study of user-inclusive cross-language search strategies in two contexts: cross-language question answering and retrieval of annotated images;
- the **question-answering track** focused on building up a common and replicable evaluation framework to test both mono- and cross-language QA systems. New types of natural language questions and new evaluation measures – namely the K1 value and r coefficient – were introduced in order to build more challenging test sets and to explore system self-scoring ability;
- the **image retrieval track** explored the use of both text and content-based retrieval methods for cross-language image retrieval; a major goal was to investigate the effectiveness of combining text and image for retrieval;



# CLEF 2005: Research in New Directions II



- the **speech track** focused on searching spontaneous speech from oral historical interviews (Shoah archives). Aim is to encourage development of technologies facilitating access to spontaneous speech
- the **web track** constructed a multilingual web corpus as an important first step towards a cross-lingual web retrieval test collection. This will serve as an important resource to better understand the challenges of multilingual web retrieval.
- the **geographic information retrieval track** was run as a pilot task with the aim of building an evaluation infrastructure to evaluate the retrieval of multilingual documents with an emphasis on geographic search; this was the first time that GIR systems have been evaluated in a multilingual context. The interest in this initial work, especially from industry, was encouraging.

# Take-up by DL applications



- Need to bridge the gap between research and application communities
- Need to move evaluation context from purely laboratory-based to more user-oriented experiments
- Proposal for a joint action with TEL (The European Library)

# 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
- 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



## Evaluation Campaign CLEF 2006

- Same tracks – new tasks
  - Failure analysis
  - Real-time experiments
  - Work on languages with few resources

## Community Building

- Growth in participation; workshop
- Resource repository
- Joint activities with MUSCLE NoE

## Take-up Activity

- Work with TEL

## Evaluation Research Activity

- Design of infrastructure and Development of new evaluation tools

# Fragmentation within Evaluation Initiatives



- ▶ Organizers set up a new initiative;
- ▶ Each participant acquires the collection and a set of tasks;
- ▶ Participants perform the tasks locally on their own system;
- ▶ Participants return the results to the organizers of the evaluation forum;
- ▶ Organizers make the results of the performances and of the statistical analysis of each participant available;
- ▶ Participants compare the results of their systems with the others.

# Motivations for a Uniform Approach to Evaluation



- ▶ Different moments of an evaluation forum are carried out and completed separately;
- ▶ The tools to analyze and compare results may be different from participant to participant;
- ▶ The format in which results are packed in evaluation forums is useful to exchange and transfer them but not to describe, elaborate, and reference them.

# Objectives

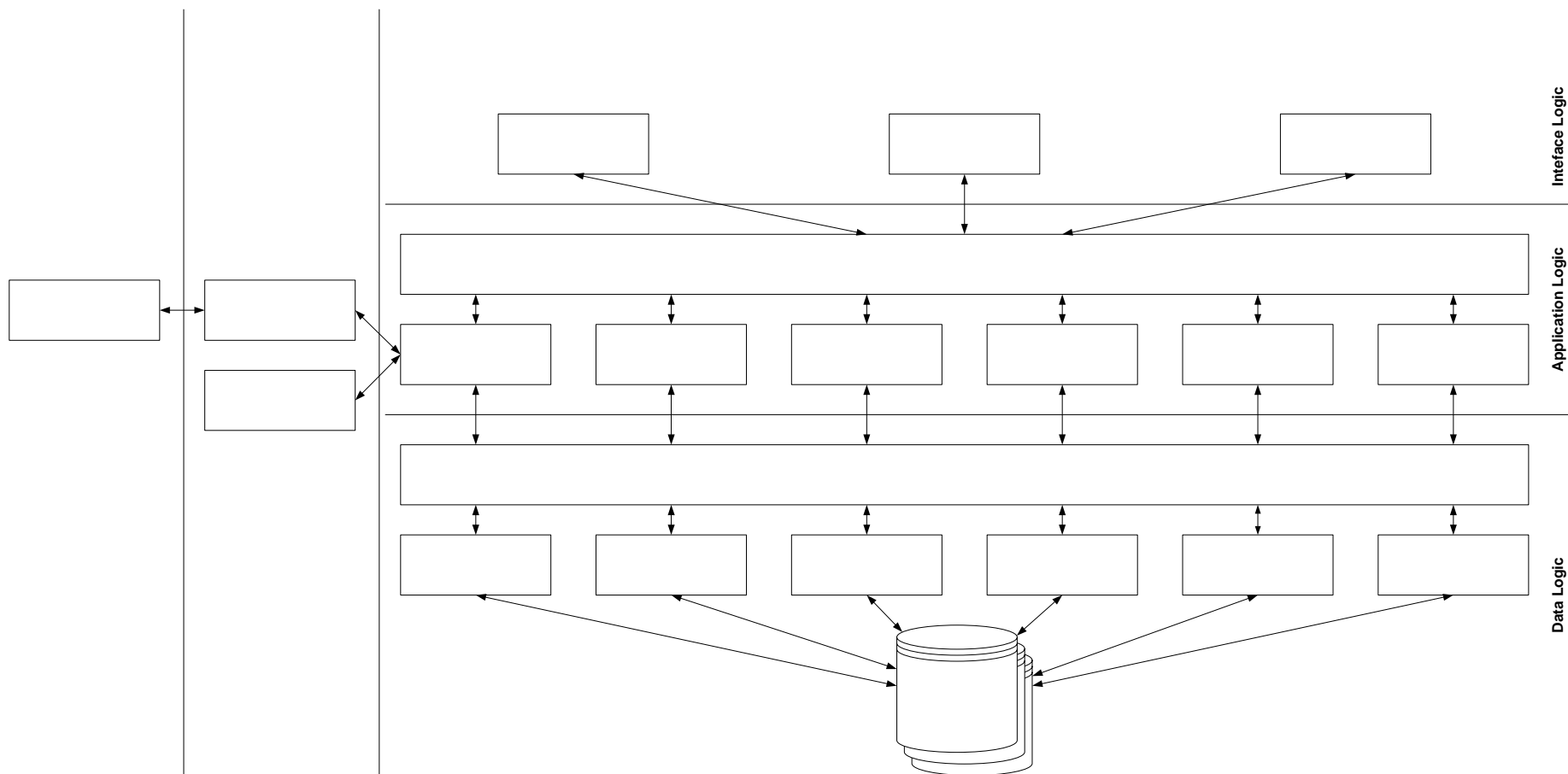


- ▶ a new approach to evaluation initiatives, which is able to address them in a uniform way;
- ▶ to provide participants with a uniform way of performing statistical analysis on their results;
- ▶ to move towards a more aware way of evaluating performances:
  - ▷ to face the lack of an historical/temporal vision of the evaluation.

The word "DIRECT" is rendered in large, bold, black letters with a metallic, embossed texture. Each letter is filled with intricate, repeating patterns that resemble ancient or historical motifs. The letters are arranged in a single horizontal line.

DISTRIBUTED INFORMATION RETRIEVAL EVALUATION CAMPAIGN TOOL

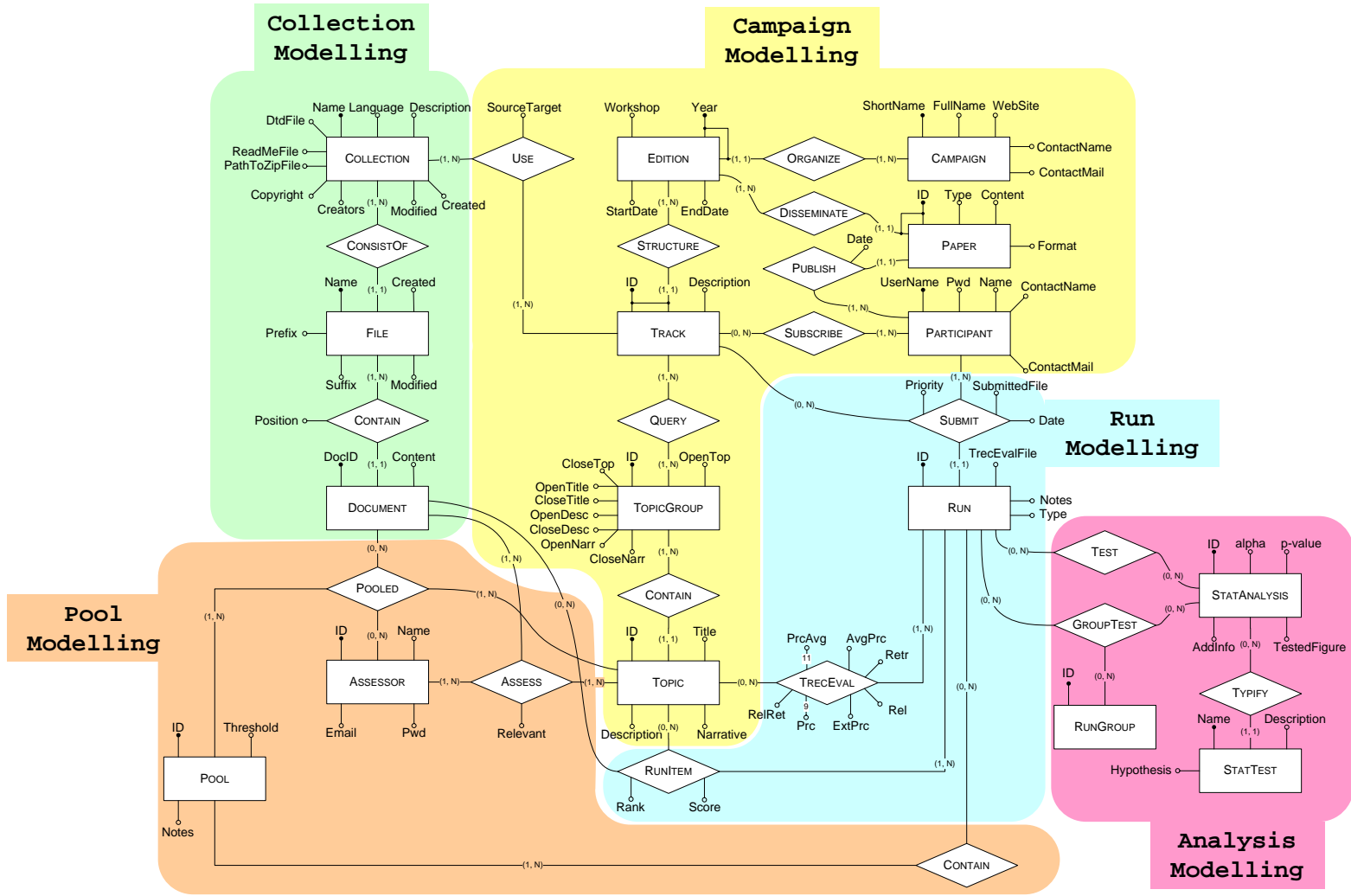
# DIRECT Architecture



**Stand-alone** **Native**



# DIRECT: Entity-Relationship Schema



# DIRECT: Main Page




DIRECT for CLEF 2005 - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indietro Cerca Preferiti

Indirizzo: <http://piscopia:8080/direct/login.jsp> Vai Collegamenti



## Cross Language Evaluation Forum CLEF 2005



powered by Distributed Information Retrieval Evaluation Campaign Tool (DIRECT)  
Wed Sep 14 18:39:58 CEST 2005

---

### Welcome to the DIRECT Login Page for CLEF 2005

Please enter your User Name and Password, and click Enter.

User Name:

Password:

Participant  Assessor  Organizer

---

CLEF is an activity of the DELOS Network of Excellence for Digital Libraries  
under the Sixth Framework Programme of the European Commission



Intranet locale

# DIRECT: Run Submission UI



DIRECT for CLEF 2005 - Submission Page for Ad-Hoc Multilingual Multi-8 2-years-on by cmu - Microsoft Internet Explorer

powered by Distributed Information Retrieval Evaluation Campaign Tool (DIRECT) Wed Sep 14 19:02:21 CEST 2005

User **cmu** logged in Log out

The fields with (\*) are mandatory

**Information about the run**

Run Identifier (*)	Track
<input type="text" value="bkmul8en3"/>	Ad-Hoc Multilingual Multi-8 2-years-on
Run Priority (*)	Run Description
<input type="text" value="1"/>	<input type="text" value="try by Nicola Ferro"/>

**Information about the query**

Query Construction (\*) Topic Fields (\*)

<input checked="" type="radio"/> Automatic	<input checked="" type="checkbox"/> Title
<input type="radio"/> Manual	<input checked="" type="checkbox"/> Description
	<input type="checkbox"/> Narrative

Source language (\*) Topics with less than 1000 retrieved documents (multiple choices by using CTRL + left click)

<input type="text" value="EN"/>	<input type="text" value="161"/> <input type="text" value="162"/> <input type="text" value="163"/> <input type="text" value="164"/> <input type="text" value="165"/> <input type="text" value="166"/>
---------------------------------	---

**Information about the file to upload**

Please upload a zip file or a gzip file (no tar gzip files) (\*)

zip  gzip

CLEF is an activity of the DELOS Network of Excellence for Digital Libraries under the Sixth Framework Programme of the European Commission

DELOS PARTNER OF EXCELLENCE Intranet locale

# DIRECT: Assessors UI



DIRECT for CLEF 2005 - Topic ( 251, Top-bg05) - Assessment by ahbg05-kspo - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti 2

Indietro - Cerca Preferiti

http://piscopia:8080/direct/protected/assessor/assessTopic.jsp

User ahbg05-kspo logged in Log out Back to Main page

**Title**  
Алтернативна медицина

**Description**  
Намерете документи, които дискутират някакъв вид алтернативно или природно медицинско лечение, включително специфичните терапии, като акупунктура, хомеопатия, хиропраксис и др.

**Narrative**  
Подходящите документи ще осигурят обща или специфична информация относно употребата на естествени или алтернативни медицински лечения и практики.

**Documents**

NST2002-01-27-056 Relevant	<a href="#">View</a>
NST2002-01-28-081 Not relevant	<a href="#">View</a>
NST2002-01-28-082 Not relevant	<a href="#">View</a>
NST2002-01-28-087 Relevant	<a href="#">View</a>
NST2002-02-04-023 Not relevant	<a href="#">View</a>
NST2002-02-04-107 Not relevant	<a href="#">View</a>
NST2002-02-08-080 Not relevant	<a href="#">View</a>
NST2002-02-08-107 Not relevant	<a href="#">View</a>
NST2002-02-11-116 Not relevant	<a href="#">View</a>
NST2002-02-11-044 Not relevant	<a href="#">View</a>

**Document ID:** NST2002-01-27-056 Relevant Not Relevant

**Status:** Relevant

Search Clear Prev. doc Next doc

```
<DOCNO> NST2002-01-27-056 </DOCNO>
<TITLE> Лечебният глад влиза като терапевтичен метод в болниците </TITLE>
<AUTHOR> КРУМ БЛАГОВ </AUTHOR>
<DATE> 27 / 01 / 2002 </DATE>
<RUBRIC> faces </RUBRIC>
<LEAD> Две чудеса преживява Лидия Ковачева </LEAD>
<LEAD> Пътят към дълголетието е правилното хранене, твърди природолечителката в новата си книга </LEAD>
<LEAD> " Има едно заболяване и то се лекува по един начин " - проповядва Лидия Ковачева. И штира Хипократ: "
Храната да бъде вашето лекарство и лекарството - вашата храна " . </LEAD>
<PICTURE> <IMGTEXT> <IMGTEXT>
<IMGAUTHOR> фото Стоян Ненов </IMGAUTHOR>
<PICTURE>
<TEXT>
```

На 88 години Лидия Ковачева доживя да се сбъднат две нейни мечти - да довърши третата си книга и да види болница, където лекуват по нейния метод, т. е. без лекарства.

Известната природолечителка е пионер на терапията чрез гладуване в България. Художник по професия, тя намира новото си призвание на 40-годишна възраст. По това време Лидия Ковачева е болна от туберкулоза в последния стадий. Не ѝ дават повече от три месеца живот, когато чува за лечебния глад и метода на руския проф. Юрий Николаев. Рускиня по майка, тя прочита достъпната литература и започва да експериментира, естествено върху себе си. Две години не хапва нищо от животински произход, после включва в диетата си малко мляко, сирене, яйца, но месо не е яла до днес, вече 48 години.

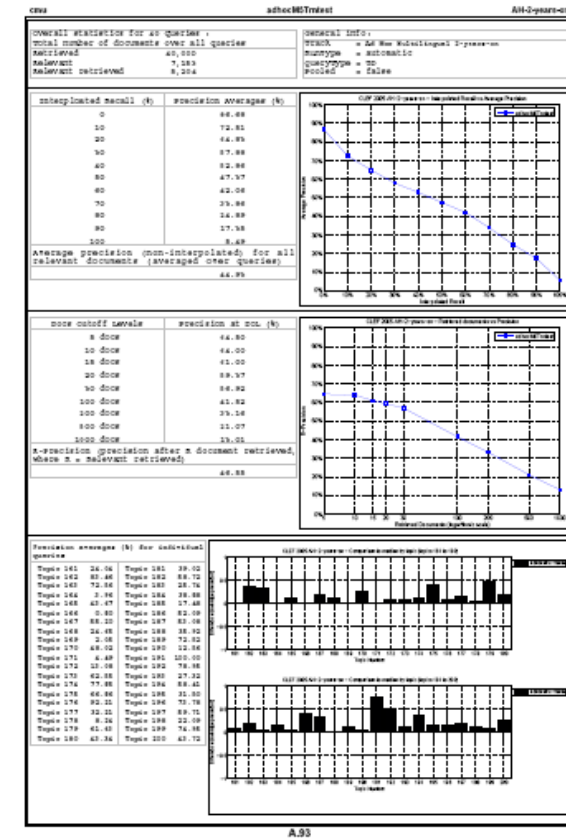
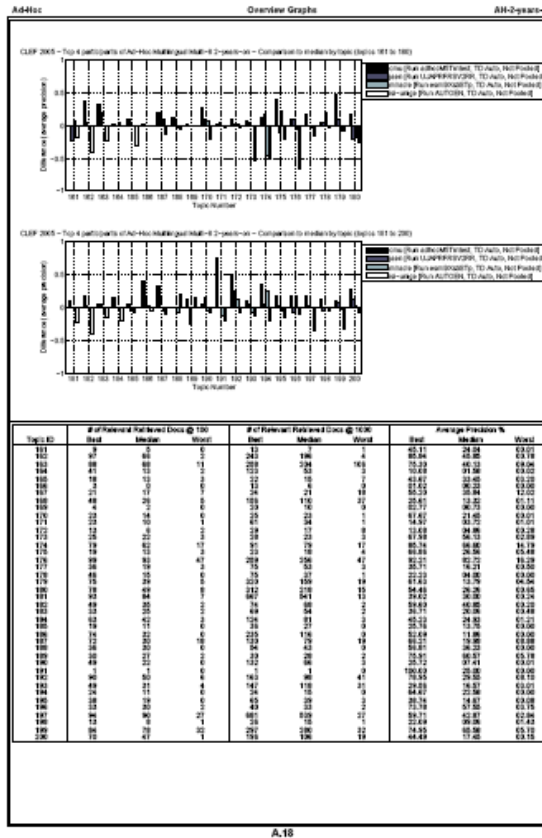
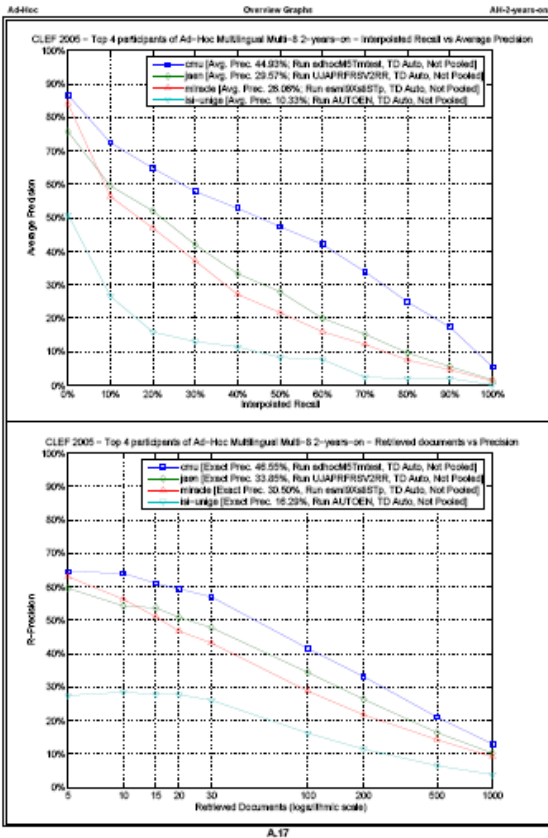
Тя не пропагандира вегетарианството, но смята, че човек по природа не е месояден. С времето създава стройна система за диетично хранене, която е далеч от чуждите образци. Ковачева препоръчва плодове, чайове и бульони.

На 80 години методът ѝ донася второ чудодейно изцеление - чули крак в ябълката, но не остава в болница, а лежи вкъщи на плодова диета. Твърди, че след три седмици костта ѝ зараснала.

Лидия Ковачева започва да описва метода си преди повече от 20 години. Отначало той се разпространява в шиклостилни беседи на Българското есперантско дружество в тираж по 50 екземпляра - толкова разрешава властта. И още тогава започват да печатат нелегално " черни " тиражи. Чак през 1992 г. излиза първата ѝ книга - " Целебният глад ". Втората, " Гладът - приятел и лекарство ", претърпя много издания, издадена е на естонски, украински, а сега се превежда на руски, английски и датски.

Operazione completata Intranet locale

# DIRECT: Performance Evaluation



# DIRECT at CLEF 2005



DIRECT has been successfully adopted during the CLEF 2005 campaign:

- ▶ it has been used by nearly 30 participants spread over 15 different nations;
- ▶ more than 530 experiments have been submitted;
- ▶ more than 160,000 documents in 7 different languages have been assessed by 15 assessors;
- ▶ reports and overview graphs about the submitted experiments have been automatically produced;



# DIRECT: Conclusions and Future Work



- ▶ Issues in evaluating the information access components of a DLMS have been pointed out;
- ▶ A system, called DIRECT, able to address the highlighted issues has been designed and developed in a first prototype;
- ▶ Runs could become first class digital objects and be referenceable by using Digital Object Identifiers (DOIs);
- ▶ Also statistical analyses could become first class digital objects and be referenceable by using DOIs;
- ▶ It is possible to move towards a digital library of scientific data about the evaluation of a DLMS.



Peters, C., (eds.), Working Notes for the CLEF 2005 Workshop, 21-23 September, Vienna, Austria. [http://www.clef-campaign.org/2005/working\\_notes/](http://www.clef-campaign.org/2005/working_notes/)

Peters, C., Clough, P.D., Gonzalo, J, Jones, G., Kluck, M., Magnini, B.(eds.) “Multilingual Information Access for Text, Speech and Images”, CLEF 2004, Bath, UK, 2004. Revised papers”. LNCS 3491, Springer 2005, 845 p.

Peters, C. (2005). “Comparative Evaluation of Cross-language Information Retrieval Systems”. In Matthias Hemmje, Claudia Niederee, Thomas Risse (eds.) From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments, Springer LNCS.

Bacchin, M., Ferro, N. and Melucci, M. “A Probabilistic Model for Stemmer Generation.” Information Processing and Management,” 2005, vol. 41, n. 1, pp. 121–137.

Di Nunzio, G. M. and Ferro, N. (2005). DIRECT: a System for Evaluating Information Access Components of Digital Libraries. In Proc. ECDL 2005, pages 483–484. LNCS 3652, Springer, Heidelberg, Germany.





Di Nunzio, G. M., and Ferro, N. (2005). DIRECT: a Distributed Tool for Information Retrieval Evaluation Campaigns. In Proc. 8th International Workshop of the DELOS Network of Excellence on Digital Libraries on Future Digital Library Management Systems (System Architecture & Information Access), 2005.

Gonzalo, J., Peters, C. (2005). The impact of evaluation on multilingual text retrieval. Proceedings of SIGIR 2005, ACM Press, NY, USA, pp 603-604.

Hansen, P. and Karlgren, J. (2005) Effects of Foreign Language and Task Scenario on Relevance Assessment. Journal of Documentation 61:5.

## **Publications by CLEF participants that cite CLEF results (partial)**

International Journal Articles

2005 = 7, 2004 = 13

International Conferences

2006 = 2 (accepted) 2005 = 27 2004 = 24

Articles or Chapters in Monographs

2005 = 6, 2004 = 1

Theses

2005 = 1, 2004 = 2