

First International Workshop on Evaluating Information Access (EVIA 2007)
Tokyo, May 15, 2007

A Proposal to Extend and Enrich the Scientific Data Curation of Evaluation Campaigns

Maristella Agosti, Giorgio Maria Di Nunzio, Nicola Ferro

Information Management Systems (IMS) Research Group
Department of Information Engineering
University of Padua, Italy



DEPARTMENT OF
INFORMATION
ENGINEERING
UNIVERSITY OF PADOVA





Outline

- ❑ Background on experimental evaluation
- ❑ Motivations and objectives
- ❑ Discussion on the current methodology
- ❑ Possible extensions to the current methodology
- ❑ A concrete example: the DIRECT system



Experimental Evaluation Background

- ❑ The Cranfield evaluation methodology is a very well understood paradigm
 - its main focus is on experiment comparability and performance evaluation

- ❑ Several successful evaluation initiatives (TREC, CLEF, NTCIR, ...) have adopted this paradigm
 - they have produced a huge amount of data, promoted the research in the IR field, and favoured the creation of cross-disciplinary communities

- ❑ Steve Robertson, in his keynote at ECIR 2007, pointed out
 - the IR field has a long tradition in evaluation and “the tradition that began with Cranfield is alive and kicking, half a century later”
 - we need to understand “how and when to push its boundaries, how and when to transcend it without throwing it away”
 - such evaluation initiatives are an “extremely valuable infrastructure for the [IR] field”.



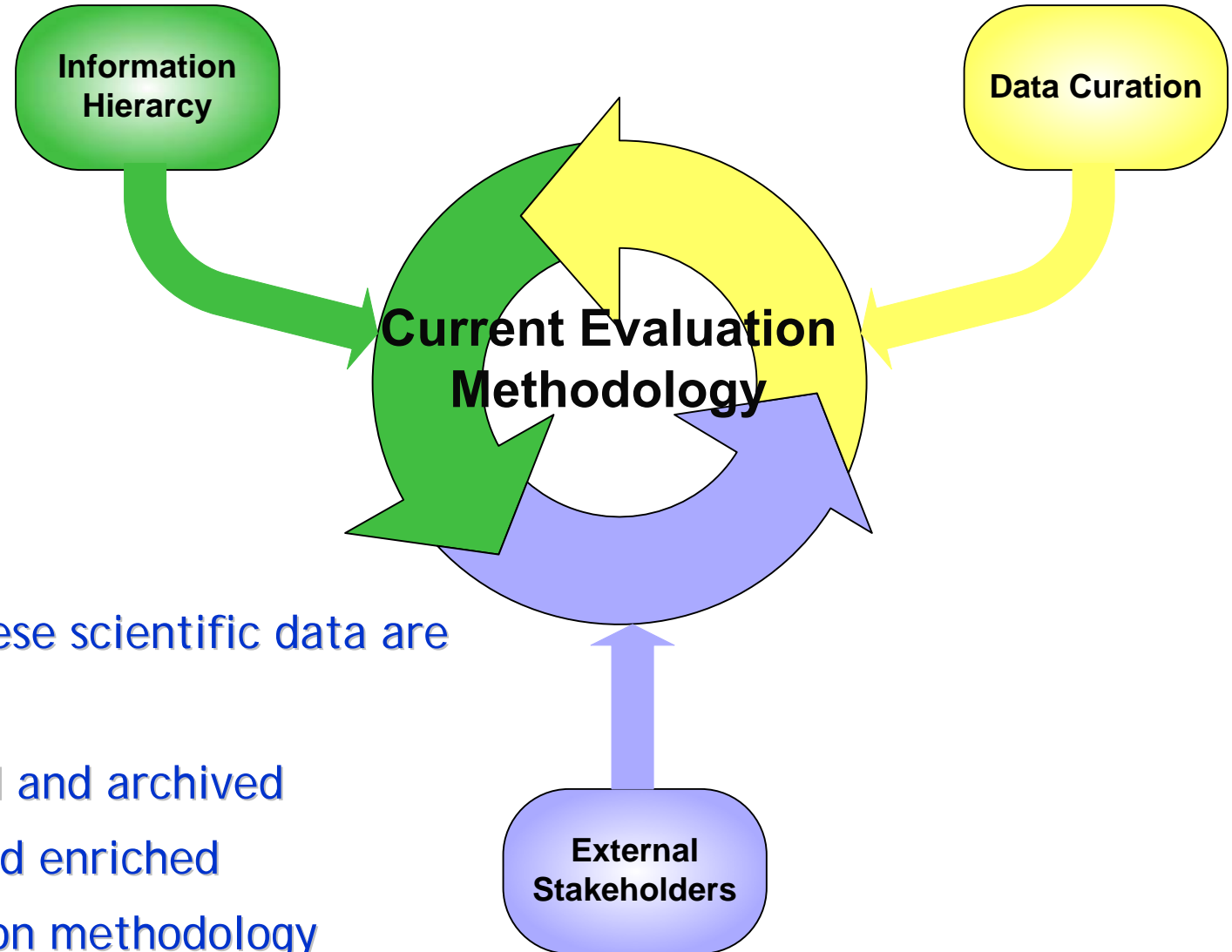
Motivation

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

- ❑ Comparable experiments
- ❑ Performance measurements concerning the experiments
- ❑ Descriptive statistics about a collection of experiment
- ❑ Hypothesis tests for in-depth analysis of the experiments



Objectives



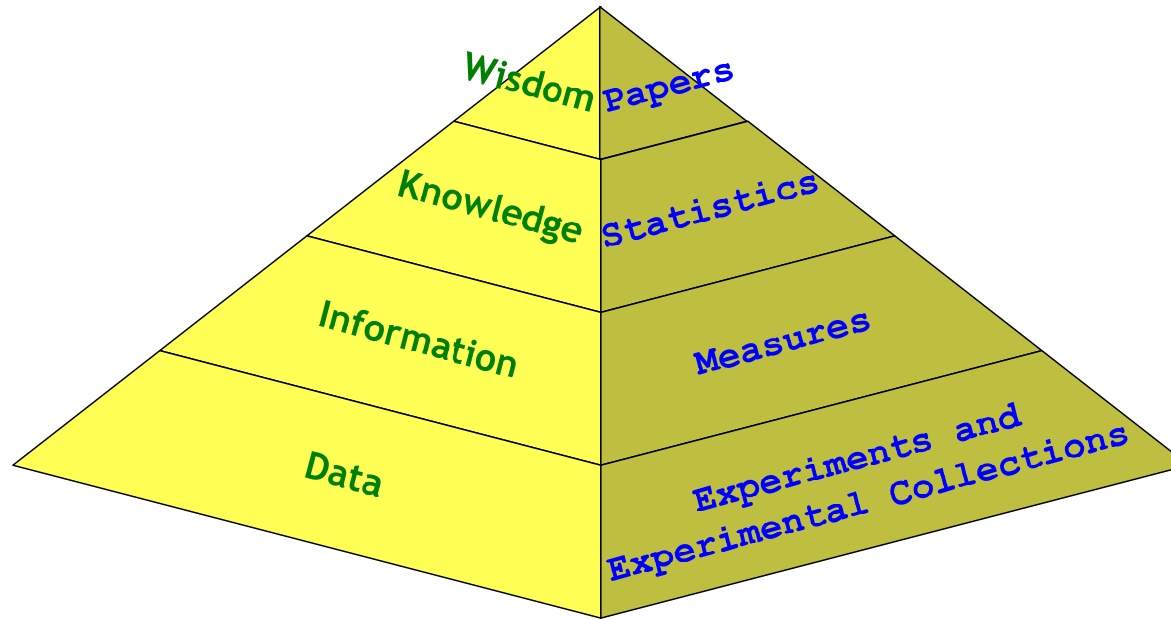
Investigate whether these scientific data are

- properly modelled
- effectively managed and archived
- carefully curated and enriched

in the current evaluation methodology



Information Hierarchy



- ❑ experimental collections and the experiments are data, since they are the raw, basic elements needed for any further investigation
- ❑ performance measurements are information, since they are the result of computations and processing on the data,
- ❑ descriptive statistics and the hypothesis tests are knowledge, since they are a further elaboration of the information carried by the performance measurements
- ❑ theories, models, algorithms, and techniques are wisdom, since they provide interpretation, explanation, and formalization of the content of the previous levels.



Data Curation

- ❑ Scientific data should be archived, preserved, maintained over the time, and made easily accessible to interested users;
 - Re-use of data for new research
 - Retention of expensive or difficult to generate data

- ❑ Their lineage should be tracked since it allows us to judge the quality and applicability of information for a given use;
 - Validation of published research results

- ❑ Scientific data should be enriched progressively adding further analyses and interpretations on them;
 - Enhancement of existing data available for research projects

- ❑ It should be possible to cite scientific data and their further elaboration
 - cross-dissemination of scientific results to research communities and industrial partners



External Stakeholders

- ❑ The EC in the 7FP i2010 Digital Library Initiative states that
 - digital repositories of scientific information are essential elements to build European eInfrastructure for knowledge sharing and transfer, feeding the cycles of scientific research and innovation up-take

- ❑ The US National Scientific Board points out that
 - organizations make choices on behalf of the current and future user community on issues such as collection access; collection structure; technical standards and processes for data curation; ontology development; annotation; and peer review

- ❑ The Australian Working Group on Data for Science suggests to
 - establish a nationally supported long-term strategic framework for scientific data management, including guiding principles, policies, best practices and infrastructure



Extending the Approach to the Evaluation (1/2)

□ 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 mechanism

- it allows for explicit citation and easy access to the scientific data and it supports the **enrichment** of the scientific data



Extending the Approach to the Evaluation (2/2)

- ❑ Provide common tools for statistical analyses
 - they allow for judging whether measured differences between retrieval methods can be considered statistically significant
 - a uniform way of performing statistical analyses on experiments make the analysis and assessment of the experiments comparable too

- ❑ Design and develop a **Digital Library System (DLS)** for IR scientific data
 - it is well suited for managing and making accessible the scientific data and the experiments produced during the course of an evaluation campaign
 - it also provides tools for analyzing, comparing, and citing the scientific data of an evaluation campaign, as well as curating, preserving, annotating, enriching, and promoting the re-use of them

- ❑ Give to **organizations** responsible for evaluation initiatives an **active role** in this process
 - they should take a leadership role in developing a comprehensive strategy for long-lived digital data collections and drive the research community through this process in order to improve the way of doing research
 - they should take care also of defining guiding principles, policies, best practices for making use of the scientific data produced during the evaluation campaign itself



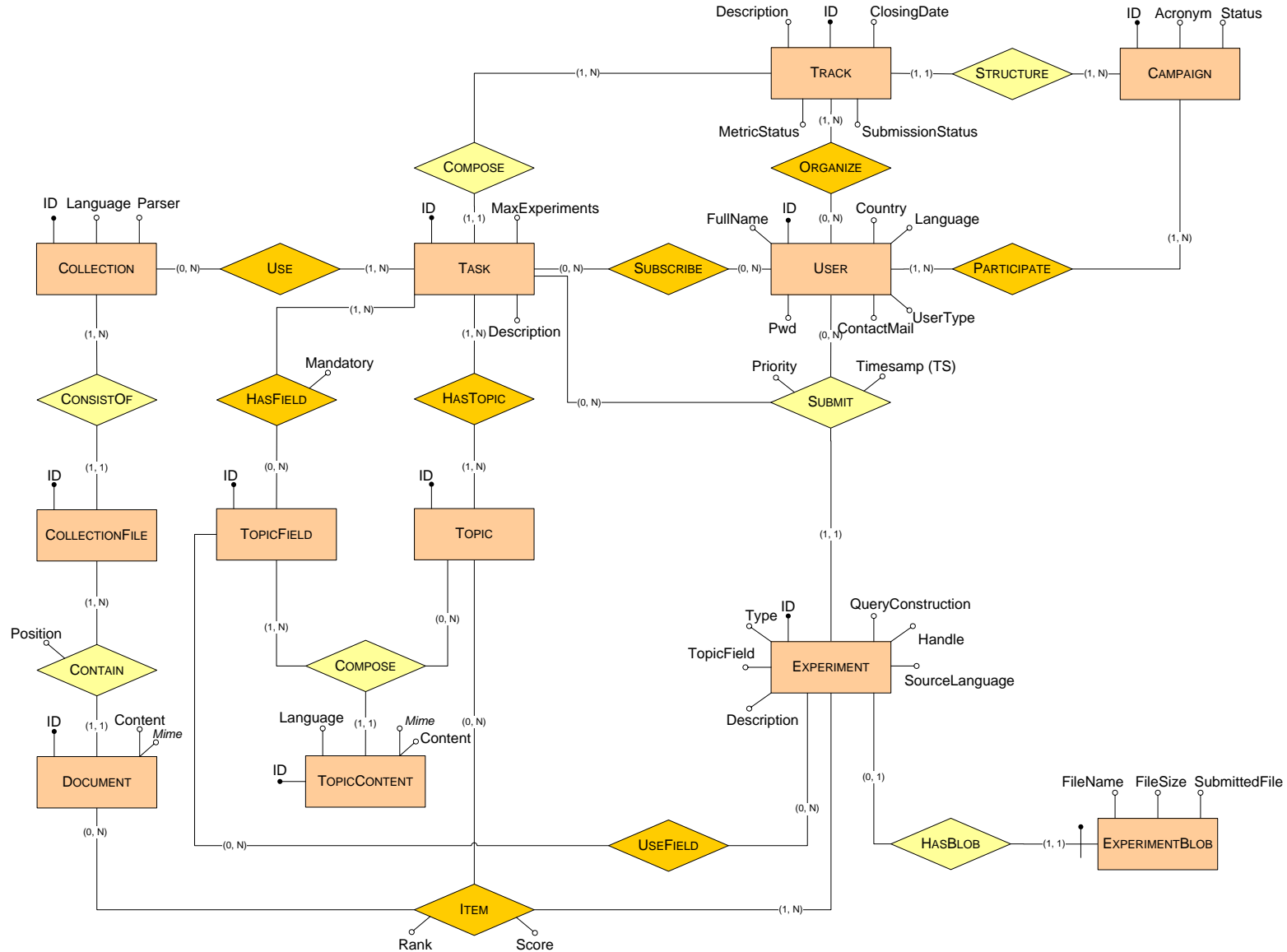
DIRECT: a DLS for IR Scientific Data

- ❑ DIRECT (Distributed Information Retrieval Evaluation Campaign Tool) is a digital library system for managing the scientific data produced during an evaluation campaign

- ❑ DIRECT has been adopted in
 - CLEF 2005: 30 participants spread over 15 different nations submitted more than 530 experiments; 15 assessors assessed more than 160,000 documents in 7 different languages (Latin and Cyrillic alphabets)
 - CLEF 2006: nearly 75 participants spread over 25 different nations submitted around 570 experiments; 40 assessors assessed more than 198,500 documents in 9 languages (Latin and Cyrillic alphabets)
 - CLEF 2007: ongoing



DIRECT Conceptual Model



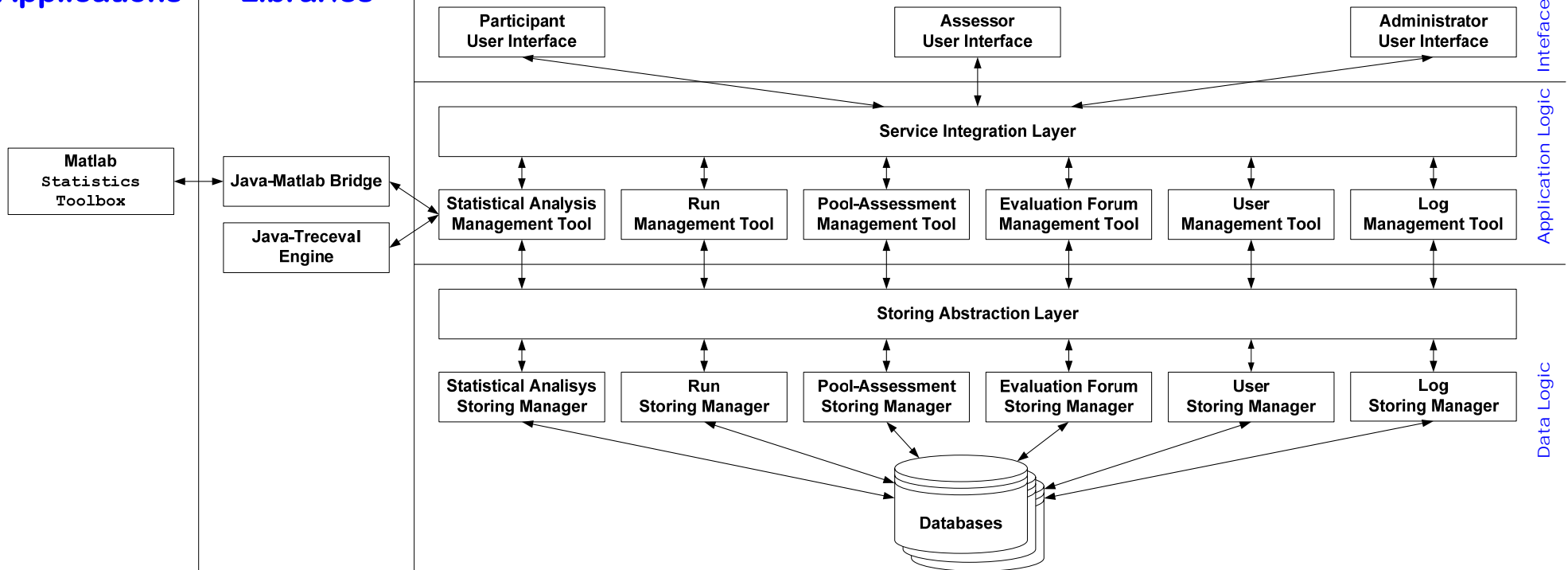


DIRECT Architecture

Stand-alone Applications

Native Libraries

Service for Supporting the Evaluation of Information Access Components



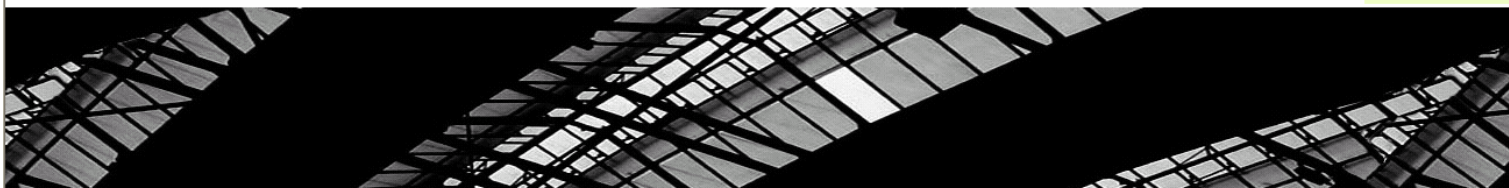


DIRECT Participant Interface (1/2)

DIRECT
DISTRIBUTED INFORMATION EVALUATION CAMPUS 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
View Run Info View Experiment Metrics bgF5bg35	viernes 2 de junio de 2006 at 00H22' CEST	2	description narrative title	TDN run
View Run Info View Experiment Metrics bgF5bg25	viernes 2 de junio de 2006 at 00H20' CEST	3	description title	Baseline mandatory TD run
View Run Info View Experiment Metrics bgx011bgFW4FS4	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	



DIRECT Participant Interface (2/2)

DIRECT for Cross Language Evaluation Forum (CLEF) participant

Logged as daedalus [Logout]

[Back](#)

Overall Metrics | **Average Precision Statistics** | R-Precision Statistics | Topic by Topic Metrics | XML Download

Overall Metrics

Overall statistics for 50 queries

Total number of documents over all queries	
Retrieved	50,000
Relevant	1,249
Relevant Retrieved	1,083

General info

Track	Ad Hoc Monolingual Bulgarian track
Max Priority	12
Experiment Priority	2
Pooled	true
Query Construction	Automatic from topics
Topic Fields	description narrative title
Experiment Type	Official Experiment

Interpolated Recall vs Average Precision

Interpolated Recall (%)	Precision Averages (%)
0%	68,478
10%	52,328
20%	45,678
30%	41,318
40%	35,998
50%	32,978
60%	27,158
70%	25,448
80%	17,218
90%	12,138
100%	9,118

Average precision (non-interpolated) for all relevant documents (averaged over queries): 30,80%

Download PDF Version | Download PNG High Resolution Version

DIRECT for Cross Language Evaluation Forum (CLEF) participant

Logged as daedalus [Logout]

[Back](#)

Overall Metrics | Average Precision Statistics | **R-Precision Statistics** | Topic by Topic Metrics | XML Download

Average Precision Statistics

Average Precision

Minimum	0,0037
Maximum	0,9267
First Quartile	0,1046
Second Quartile	0,2530
Third Quartile	0,4683
Interquartile Range (IQR)	0,3637
Arithmetic Mean	0,3080
Standard Deviation	0,2340
Lower Outlier Threshold	0,0037
Upper Outlier Threshold	0,9267
Arithmetic Mean without Outliers	0,3080
Standard Deviation without Outliers	0,2340

Download PDF Version | Download PNG High Resolution Version



DIRECT Assessor Interface (1/2)

DIRECT

DIRETTORIO INFORMAZIONI INTERNAZIONALE CAMPANA TOC

for Cross Language Evaluation Forum (CLEF)

assessor



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	Not 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 непозволените 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> <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>

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>

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 приближаващата 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 [View](#)
 RELEVANT
 Document 54 / 436

SEGA20021228-041 [View](#)
 NOT_RELEVANT
 Document 55 / 436

SEGA20021228-043 [View](#)
 NOT_RELEVANT
 Document 56 / 436

DELUG Network of Excellence for Digital Libraries under the Sixth Framework Programme of the European Commission
 DIRECT - Copyright © 2009-2008 Information Management Systems (IMS) Research Group, Department of Information Engineering (DEI), University of Padua



Conclusions

- ❑ We have discussed the issues concerning the management, enrichment and curation of the scientific data produce during the evaluation

- ❑ We have proposed some possible extension to the current evaluation methodology

- ❑ We have designed and developed the DIRECT system, a DLS for IR scientific data, which has been tested in the context of CLEF 2005 and 2006