



The CLEF 2005 interactive track (iCLEF)

Julio Gonzalo¹, Paul Clough² and Alessandro Vallin³

¹Departamento de Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia

²Department of Information Studies, University of Sheffield, UK

³ITC-irst, Trento, Italy

Overview

- Consolidation of two pilot user studies at iCLEF 2004
 - Interactive question answering task
 - Interactive image retrieval task
- iCLEF provides resources and experiment design
- Participants select a research question to investigate (by comparing the behaviour and search results of users with a reference and a contrastive system)
- Five research groups submitted results
 - 2 groups for image retrieval
 - 3 groups for QA

Agenda

- Image retrieval task
- Question Answering task
- Ideas for 2006: the flickr task

Cross-language image retrieval

Overview

- Limited evaluation with ranked lists
 - Image retrieval systems highly interactive
- Appealing for CLIR research
 - Language-independent: object to be retrieved is an image
- Image Retrieval
 - Purely visual (QBE) e.g. “find images like this one”
 - Text-based e.g. Web image search
 - Combination

Interactive task

- Areas of interaction to study include
 - Query formulation (visual and textual)
 - Query re-formulation (relevance feedback)
 - Browsing/navigating results
 - Identifying/selecting relevant images
- Based on iCLEF methodology
 - Participants require minimum of 8 users
 - Within-subject experimental design
 - 16 search tasks (5 mins per task)
- Participants select area to investigate

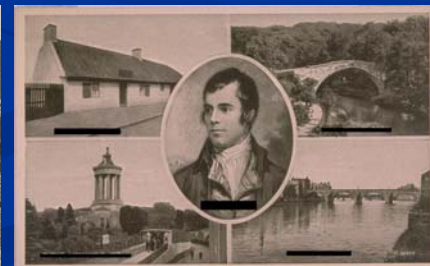
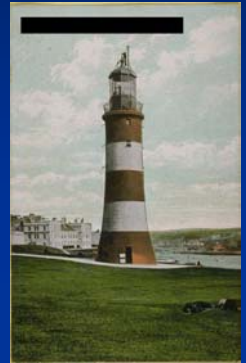
Target search task



海 (sea)

- Clear goal for the user (easy to describe task)
- Can be achieved without knowledge of the collection
- Clearly defined measures of success
- Invokes different searching strategies

Search tasks



Participants

- 11 signed up; 2 submitted
- University of Sheffield
 - Compared Italian version of the same system
 - Aimed to test whether automatically-generated menus better for presenting results than ranked list
- Miracle
 - Compared Spanish versus English query formulation
 - Aimed to test whether Boolean AND or OR better

Results

- Miracle
 - 69% of images found English; 66% Spanish
 - Domain-specific terminology caused problems for users (and system)
- University of Sheffield
 - 53% images found using list; 47% menus
 - Users preferred the menus
- Comparison between groups (limited)
 - Miracle: 86/128 images found overall
 - Sheffield: 82/128 images found overall

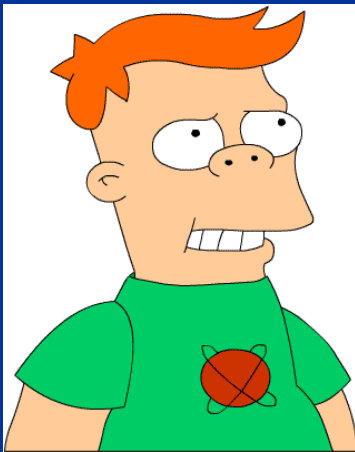
Interactive CL Q&A

Q&A task

Question
(native language)

Answer
(native language)

Text collection
(foreign language)



Q&A search assistant



Q&A vs. interactive Q&A

- People know some of the answers
 - Questions must be carefully selected
- People can draw inferences
 - Answer from multiple documents: considered in 2004, problems with assessment
 - Combination of document evidence with user knowledge: avoid definition and other open questions.
- People answer in the question language
 - Need to provide high-quality manual translations for assessment.
- People get tired
 - Exclude nil questions, limit question types

Experimental Design

- 8 users (native query language)
- 16 evaluation questions (+ 4 for training)
- 5 minutes per search (~3 hours per user)
- Independent variable: CLIR system design (reference/contrastive)
- Dependent variable: accuracy
- Latin square to block user/question effects

Evaluation measures

- Official score: accuracy (= Q&A track)
- Additional quantitative data: searching time, number of interactions, log analysis in general.
- Additional data: questionnaires (initial, 2 post-system, final), observational information.

Experiments

- **Alicante**: How much context users need to correctly identify answers? (clauses vs full paragraphs in a QA-based system)
- **Salamanca**: How useful is MT for the task? (with/without MT) X (poor/good target language skills) X (EN/FR as target language)
- **UNED**: Is it better to search paragraphs than full documents?

Official results

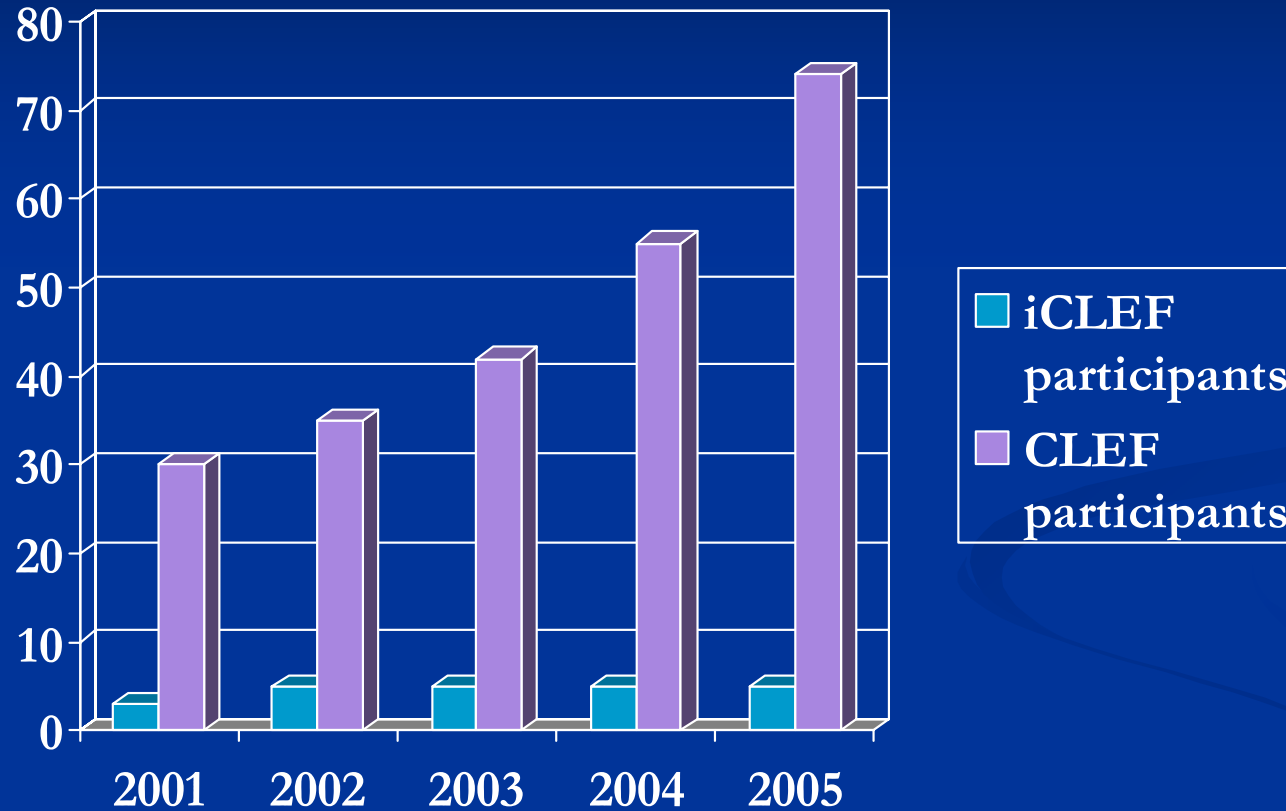
Group	Users	Docs	Experiment Condition	Accuracy	
				Strict	Lenient
Alicante	ES	EN	full passages	.44	.45
Alicante	ES	EN	clauses	.34	.34
Salamanca	ES	EN	good lang. skills / no translation	.50	.53
Salamanca	ES	EN	good lang. skills / translation	.56	.56
Salamanca	ES	EN	poor lang. skills / no translation	.36	.42
Salamanca	ES	EN	poor lang. skills / translation	.39	.45
Salamanca	ES	FR	good lang. skills / no translation	.66	.67
Salamanca	ES	FR	good lang. skills / translation	.69	.73
Salamanca	ES	FR	poor lang. skills / no translation	.63	.70
Salamanca	ES	FR	poor lang. skills / translation	.61	.66
UNED	ES	EN	documents	.53	.53
UNED	ES	EN	sentences with answer type filter	.45	.45

Remarkable facts

- UNED & Alicante: accuracy increases with larger contexts.
- Salamanca: MT is not very helpful!
- Implications for CL-QA systems?

Ideas for 2006

Participation



Conclusion: terminate the track!

Failure analysis (1)

- High cost of entry
 - Long, boring guidelines.
 - User recruitment, scheduling, training, monitoring.
 - Can't really do experiment variations.
 - Made a programming mistake? Start recruiting volunteers again.

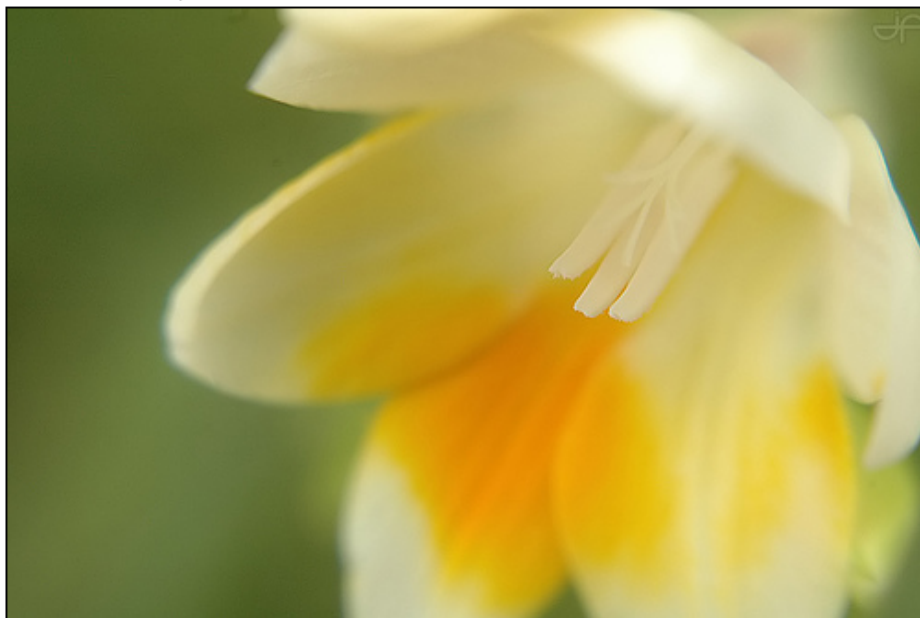
Failure analysis (2)

- “users screw everything up” (XXX, IR competition organizer). Recruiting, training, monitoring, sometimes even paying... just to see how users ruin your hypothesis.
- Is your search assistant at least good for demonstration purposes? No, because
 - 1) **Cross-Language Search \Leftrightarrow cross-cultural need**
 - 2) **But cross-cultural need is unfrequent!**

(Experiment: show your mother)

Lirio..

title

ALL
SIZES

O cheiro é bom..
Smells good..

Ijuí - RS

Comments

**Dricota! pro** says:

Adoro o cheiro e o macro ficou lindíssimo!
Posted 20 hours ago. ([permalink](#))

**corbata1982** says:

tava sumido hein?

baita foto!
Posted 19 hours ago. ([permalink](#))

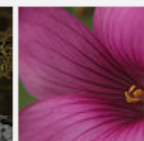
**mrsdalloway pro** says:

Uploaded on September 19, 2005
by [Arzivenko](#)

Arzivenko's photostream



< prev



next >

700
photos
[View as
slideshow](#)

This photo also belongs to:

Macros / Nature (Set)



< prev



next >

148
photos
[View as
slideshow](#)

FLOWERS (Pool)



< prev



next >

50230
[View as
slideshow](#)

+ Brasil (Pool)

+ Macro (Pool)

+ Closer and Closer Macro
Photography (Pool)

+ Catchy Colors (Pool)

+ Nature (Pool)

+ DoF (Pool)

description

O cheiro é bom..
Smells good..

Ijuí - RS

Comments



Dricota! pro says:

Adoro o cheiro e o macro ficou lindíssimo!
Posted 20 hours ago. ([permalink](#))



corbata1982 says:

tava sumido hein?

baita foto!

Posted 19 hours ago. ([permalink](#))



mrsdalloway pro says:

Poesia visual... :)

Posted 19 hours ago. ([permalink](#))



lil aNNa pro says:

lovely color!

Posted 18 hours ago. ([permalink](#))



FreeMySoul pro says:

This is really awesome!

Posted 6 hours ago. ([permalink](#))



DropsOfRuby says:

another great shot :D
fantastic

Posted 5 hours ago. ([permalink](#))

comments

sets

Tags
(folksonomies)

< prev

next >

- + FLOWERS (Pool)
- + Brasil (Pool)
- + Macro (Pool)
- + Closer and Closer M
Photography (Pool)
- + Catchy Colors (Pool)
- + Nature (Pool)
- + DoF (Pool)
- + Nikon D70 Users (Pool)
- + Rio Grande do Sul (Pool)
- + flickritis: invite EVERYONE (Pool)
- + Naturalight (Pool)
- + Top 20 F
- + Macro

Tags

- JP
- nikon
- d70
- macro
- nature
- flower
- top20flower

Additional Information

Abrete Sesamo

ADD TO FAVES BLOG THIS ALL SIZES



Spanish

Comments



ramond says:

Muy buena Lou Rouge, a ver si, si me dejas tu camara que la mia no tiene tantos aumentos, tu sacas afotos buenas por tu camara, guapa, por eso

Posted 3 months ago. (permalink)



Amelie -- pro s

Bella foto
mi piace molto
what it opens?

Posted 3 months ago. (permalink)

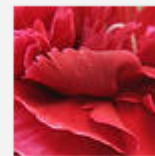
Italian



Uploaded on June 16, 2005

by [Lou Rouge](#)

Lou Rouge's photostream



< prev



next >

236
photos

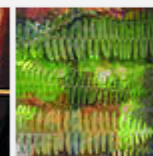
[View as
slideshow](#)

This photo also belongs to:

Curiosidades (Set)



< prev



next >

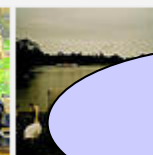
52
photos

[View as
slideshow](#)

Verde (Set)



< prev



29

English

+ FlickrCentral (Pool)

+ Green is Beautiful (Pool)

+ * Vanishing Points

+ Catchy Colors (Pool)

+ Color and Colors (Pool)

Japanese

flickling
multilingual photo search at flickr

search: sunset

Flickr Search

I'm writing in... ☒ eng ☐ esp ☐ fra ☐ ita ☐ deu

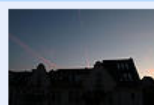
☐ Don't add translations to my search

Advanced search
Other great flickr tools

Searching: sunset or (deu) **Abendrot** or (ita) **tramonto** • **Results:** 74016 photos, page 1 of 6170 (24.29 seconds)



sunset
ferropolis • festival • melt2005 • germany • analog fisheye • **sunset** • aufbagger • ooo



Der Himmel über Berlin 2
2005 • berlin • sky • himmel • cagliari • sardegna • sardinia
sunset
moon • mond • abendrot • fujifilm • finepix s2pro • 1735mm



cattedrale
• sardinien • tramonto panorama • skyline



sky • **sunset** • bw • cloud • overhead house • residential • roof • pack



Der Himmel über Berlin
berlin • sky • himmel • **sunset** • cross abendrot • 2005 • fujifilm • finepix • s2pro 1735mm



bastione
cagliari • sardegna • sardinia • sardinien • tramonto panorama • skyline



Fire's Edge
prayingmantis • mantis • fire • **sunset** • silhouette insect • bug • garden • nature • macro



Abendrot
abendrot

Advantages of flickr

- Naturally multilingual, new IR challenge (folksonomies)
- You can show your mother! (it is cross-language but it is not cross-cultural)
- Can avoid recruiting users: study behaviour of real web/flickr users (log analysis)
- Challenges of web scenarios (social network effects) plus advantages of controlled scenarios (unlike Google or Yahoo image search)

Interactive Flickr task (2006)

- Target language: Portuguese
- Data: Flickr images (local or via Flickr API)
- Search task:
 - Illustrate this text (open)
 - What's behind this house? (focused, Q&A type)
 - Sunsets in Mangue Seco (ad-hoc type)
 - Pictures where the Nike logo appears (ad-hoc, content oriented)
 - Track real users w. real information needs (log analysis!)
- Experiment design: open!! (let's also compare evaluation methodologies!)

Plans for 2007

- Make the task compulsory for CLEF participants
- Terminate all other tracks
- Task coordinators hired by Yahoo!

Acknowledgments

- People who helped organizing iCLEF 2005: Richard Sutcliffe, Christelle Ayache, Víctor Peinado, Fernando López, Javier Artiles, Jianqiang Wang, Daniela Petrelli
- People already helping us shape flickr task: Javier Artiles, Peter Anick, Jussi Karlgren, Doug Oard, William Hersch, Donna Harman, Daniela Petrelli, Henning Müller
- All participant groups