

QUESTION ANSWERING REALTIME EXERCISE

CLEF 2006

Fernando Llopis Pascual
Elisa Noguera Robles



Universitat d'Alacant
Universidad de Alicante



Departamento de
Lenguajes y Sistemas Informáticos



Index

Task Description

The task at clef 2006

Ideas for the future.

Clef 2007?

What is RealTime QA?

- Is a pilot task at CLEF 2006.
- Main goal : Evaluate the ability of QA systems to answer within a time constraint.
- Basis:
 - Target language : **Spanish**
 - Participants will be given a certain amount of time to answer a set of questions.
 - Number of questions : **20**
 - Time : **60 minutes** (until 3 minutes per question)
 - Each participating group may bring a laptop or use a computer with internet connection

Ideas

- The time is a very important aspect in QA process
 - What do you prefer efficiency or effectiveness?
- Is better to compare different systems using a test with restrictions of time
 - Clef 2006 main task 1 week. May be too time
- The systems work automatically. No manual process (There is no time)
- And we think that is funny to fight against the time

Clef 2006 Agenda

- The file format is the same that QA main Task
- Participating teams must return
 - one or more (up to a maximum of 3) exact answers per question.
 - up to two runs.
 - You can return a run in a short lapse of time and another later.

Evaluation

- The answers submitted by participants in the exercise were manually judged by native speaking assessors.
- We have proposed a new measure combining the precision of the system with the time in order to evaluate the system on both sides.
- This measure is:
 - MRR/t
 - t = answer time of the system calculated in seconds
- This way, we evaluate with this measure the amount of correct answers of the system per unit of time.



Index

Task Description

The task at clef 2006

Ideas for the future.

Clef 2007?

Participants

- Daedalous/ Universidad Carlos III. Madrid

Laptop

Pentium M 1,75

1 Gb RAM

Linux



Participants

- Tokio Institute. Japan

Internet conexión

Linux Server



Participants

The task at Clef-2006

- Priberam. Portugal

Laptop

Pentium M 1,86

1 Gb RAM

Windows



Participants

- University of Alicante. Spain

Desktop

Pentium IV 2,8

1 Gb RAM

Linux



The task at Clef-2006

Participants

- Inaoe. Mexico

Laptop

Pentium M 1,2

512 Kb RAM

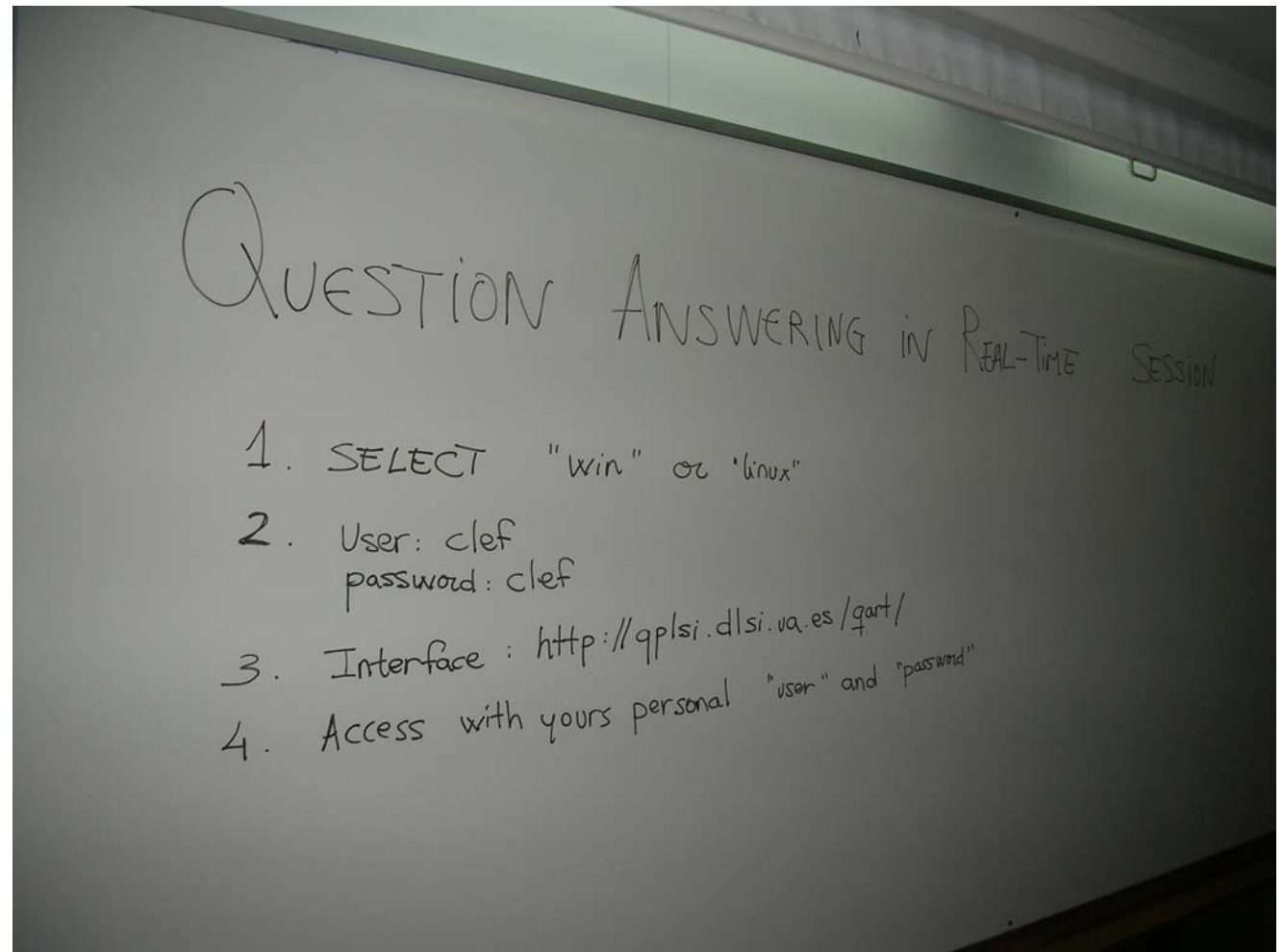
Linux



The task at Clef-2006

9:00. The room was opened

- The task on fifty words



The task at Clef-2006

9:00- 10:00

- Assistants show the interface to each participant



The task at Clef-2006

10:00

- Coffee before the competition



The task at Clef-2006

11:00-12:00 The exercise



Task Description

Preliminary Results

Participant	Run	MRR	Time (seconds)	$(MRR * 100) / \text{time}$
Daedalus	1	0.41	549	0.0744
	2	0.33	198	0.1684
Priberam	1	0.35	56	0.6250
	2	0.35	85	0.4118
Inaoe	1	0.3	1966	0.0153
Tokio Ins	1	0.38	5141	0.0073
U. Alicante	1	0.24	76	0.3180

Conclusions

- Participants use similar hardware
 - 3 laptop with pentium mobile
 - 1 desktop with pentium IV
 - 1 internet connexion to server
- Number of runs
 - 2 participants sent two runs
 - 3 participants sent only one run
- What is better?
 - Daedalous1 598 seconds with MRR 0.41
 - Or
 - Daedalous2 198 seconds with MRR 0.33
 - Or
 - Priberam1 56 seconds with MRR 0.35

Conclusions (II)

- More time does not implies better results
 - Priberam two runs with MRR 0.35
 - 1° run 56 seconds
 - 2° run 85 seconds
 - The two runs have not the same correct answers

Task Description

Preliminary Results with Inexact

Participant	Run	MRR	MRR (Inexact)
Daedalus	1	0.41	0.41
	2	0.33	0.36
Priberam	1	0.35	0.35
	2	0.35	0.35
Inaoe	1	0.3	0.3
Tokio Ins	1	0.38	0.48
U. Alicante	1	0.24	0.28



Index

Task Description

The task at clef 2006

Ideas for the future.

Clef 2007?

Conclusions

- Good
 - It was a good way to compare QA systems
 - You can know QA systems that works in real time
 - We think that has been a amazing experience
- Problems
 - We have few time to evaluate the runs



Ideas for the future. Clef 2007?

Ideas for Clef-2007

- We could
 - Study another evaluation measures that use the time.
 - Use another languages than Spanish (may be French and English)
 - Another ideas ??????????????????????




Acknowledgements

- All the participants
 - Alicante, Daedalous, Tokyo, Inaoe and Priberam
- The assistants
 - Oscar Ferrández, David Tomás y Antonio Toral
- Uned group for their support
 - Anselmo and Alvaro
- Clef organization for doing this possible



Question Answering RealTime Exercise



Clef 2006

Fernando Llopis Pascual
Elisa Noguera Robles