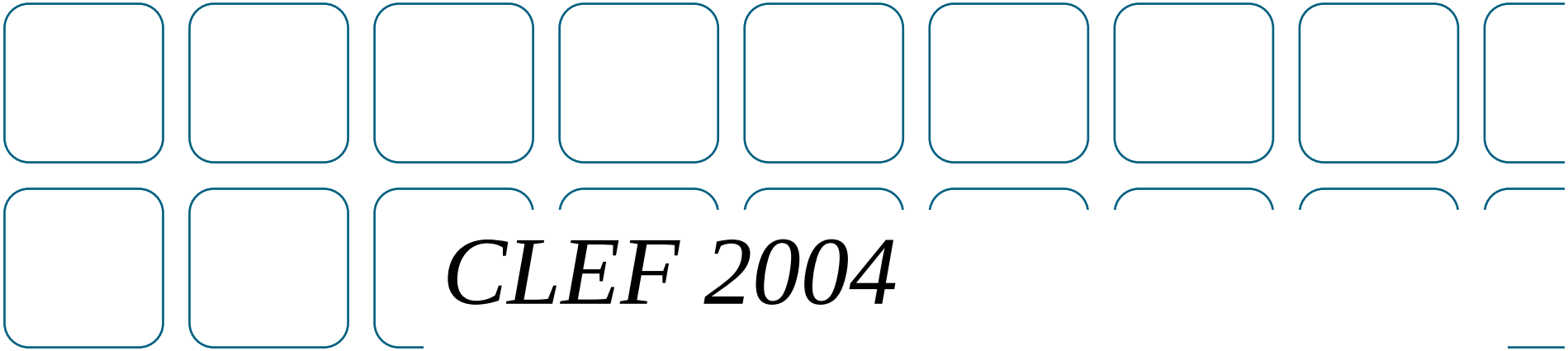


CLEF 2004 Overview of Results: Adhoc Tracks

26.8.2004, Martin Braschler

eurospider
relevancy retrieval

Outline

- Participants
- Collection
- Track Definition
- Details of experiments
- Pooling
- Preliminary Trends
- CLEF 2004 vs. 2003
- Challenges
- Results
- Conclusions

Participants in Adhoc Tracks

CEA/LIC2M (FR) *	SUNY Buffalo (US) *
CLIPS (FR) *	Thomson Legal (US) ***
Daedalus (ES) *	U Alicante (ES) ***
Dublin City U (IE) ***	U Amsterdam (NL) ***
FCUL Lisbon (PT)	U Chicago (US) *
FU Hagen (DE) *	U Glasgow (UK) *
Hummingbird (CA) ***	U Hildesheim (DE) **
ILTG (CA)	U Jaen/SINAI (ES) ***
IRIT (FR) ***	U Oviedo (ES) *
JHU/APL (US) ****	U Padua (IT) **
NII Group (JP) *	UC Berkeley (US) ****
Ricoh (JP) *	U de Evora (PT)
SICS (SE) ***	U de Neuchâtel (CH) ***

26 participants, 13 different countries

(*-**** = number of previous participations to CLEF)

The CLEF multilingual collection

	# part.	# lg.	# docs.	Size in MB	# assess.	# topics	# ass. per topic
CLEF 2004	26	5	273,863	779	114,346	50	2287
CLEF 2003	33	9	1,611,178	4124	188,475	60 (37)	~3100
CLEF 2002	34	8	1,138,650	3011	140,043	50 (30)	~2900
CLEF 2001	31	6	940,487	2522	97,398	50	1948
CLEF 2000	20	4	368,763	1158	43,566	40	1089
TREC8 CLIR	12	4	698,773	1620	23,156	28	827
TREC8 AdHoc	41	1	528,155	1904	86,830	50	1736
TREC7 AdHoc	42+4	1	528,155	1904	~80,000	50	~1600

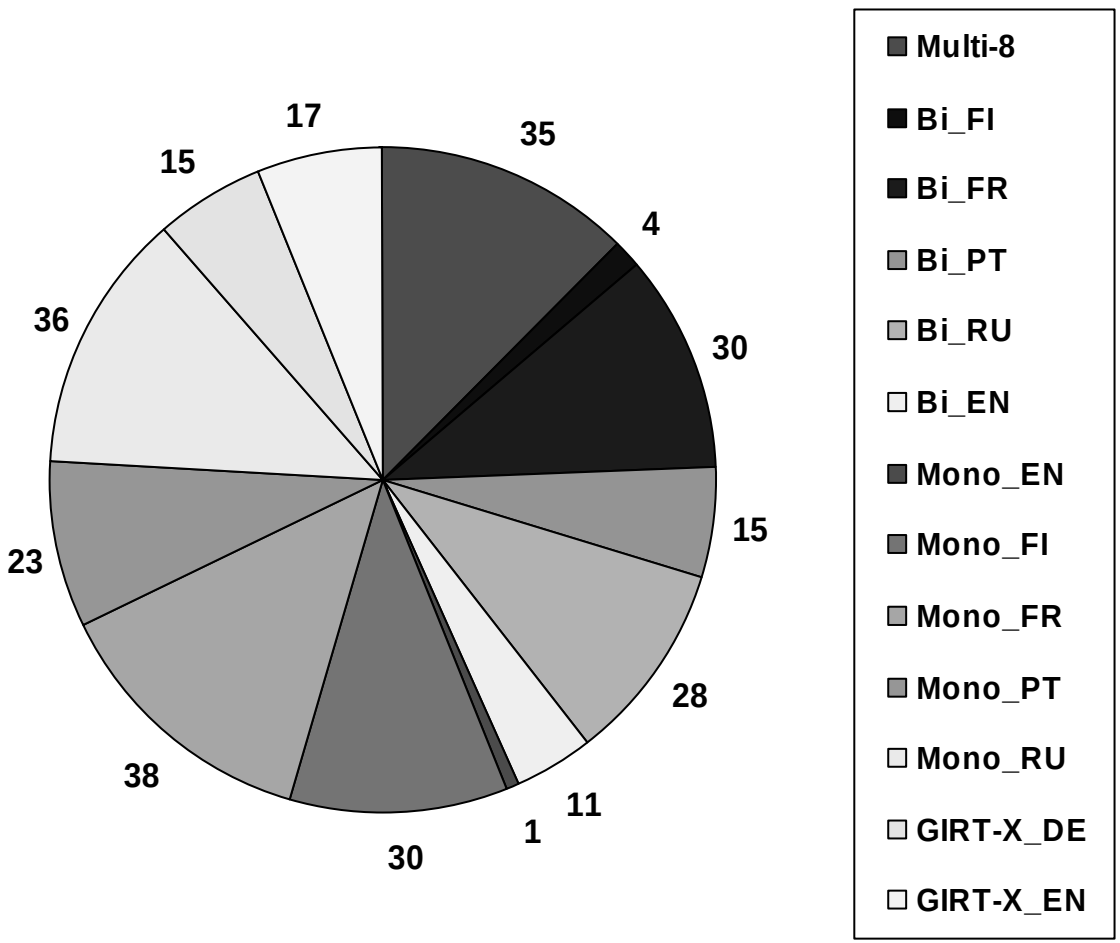
Adhoc Tasks in CLEF 2004

- Multilingual using 4 different target languages: X-> EN, FI, FR, RU. Intention was to offer a diverse set of target languages, some of which have been explored less deeply in previous CLEF campaigns.
- Bilingual: X->FI, X->FR, X->PT, X->RU, (X->EN)
- Monolingual: Monolingual retrieval on FI, FR, PT, RU collections
- GIRT: bilingual & monolingual domain-specific retrieval
- Other tracks (see different overviews)

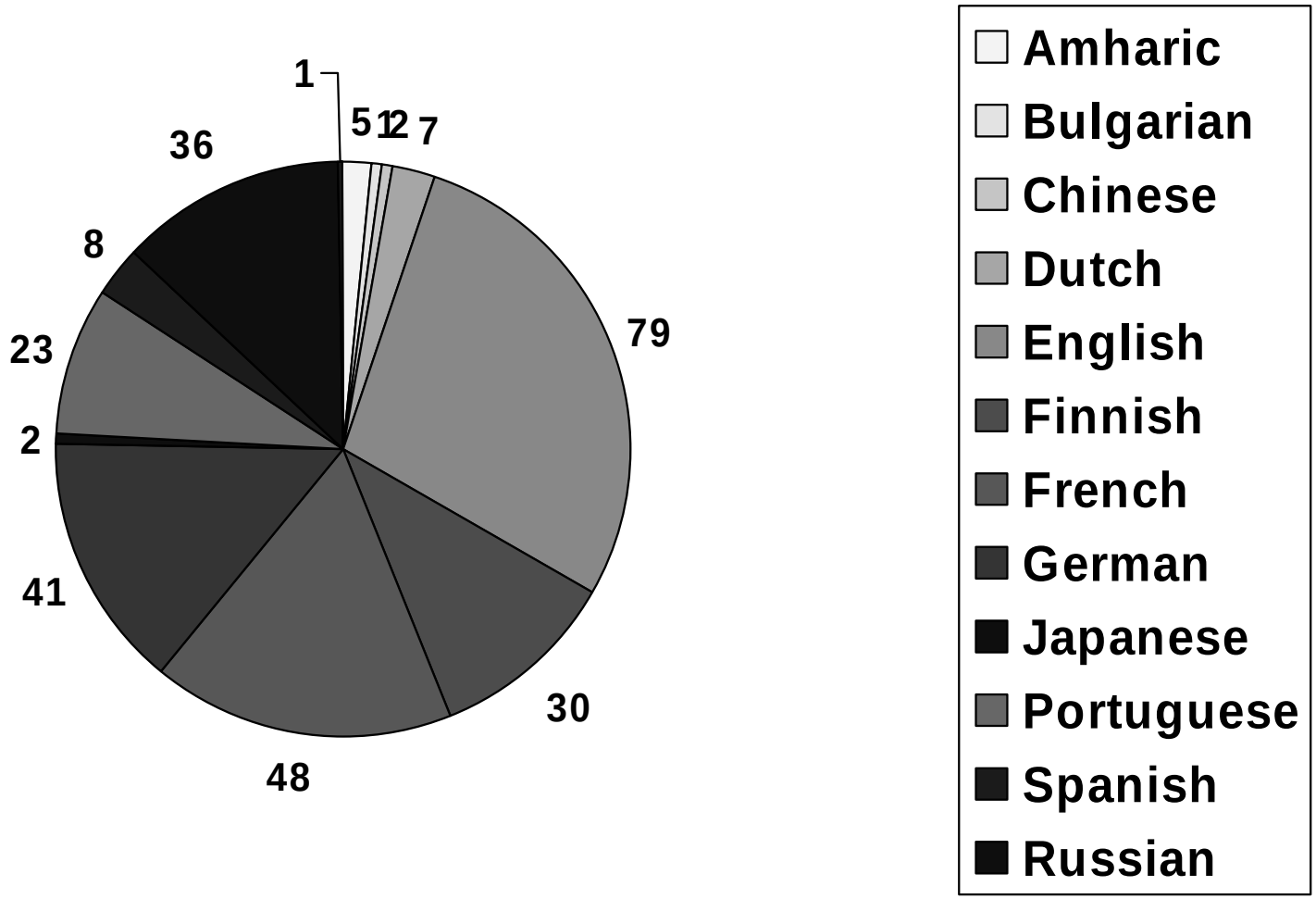
Details of Experiments

Track	*# Participants*	*# Runs/Experiments*
Multilingual	9	35
Bilingual X → FI	2	4
Bilingual X → FR	7	30
Bilingual X → PT	4	15
Bilingual X → RU	8	28
Bilingual X → EN (restricted)	4	11
(Monolingual EN)	(1)	(1)
Monolingual FI	11	30
Monolingual FR	13	38
Monolingual PT	8	23
Monolingual RU	14	36
Domain-specific GIRT → DE	3	15
Domain-specific GIRT → EN	3	17

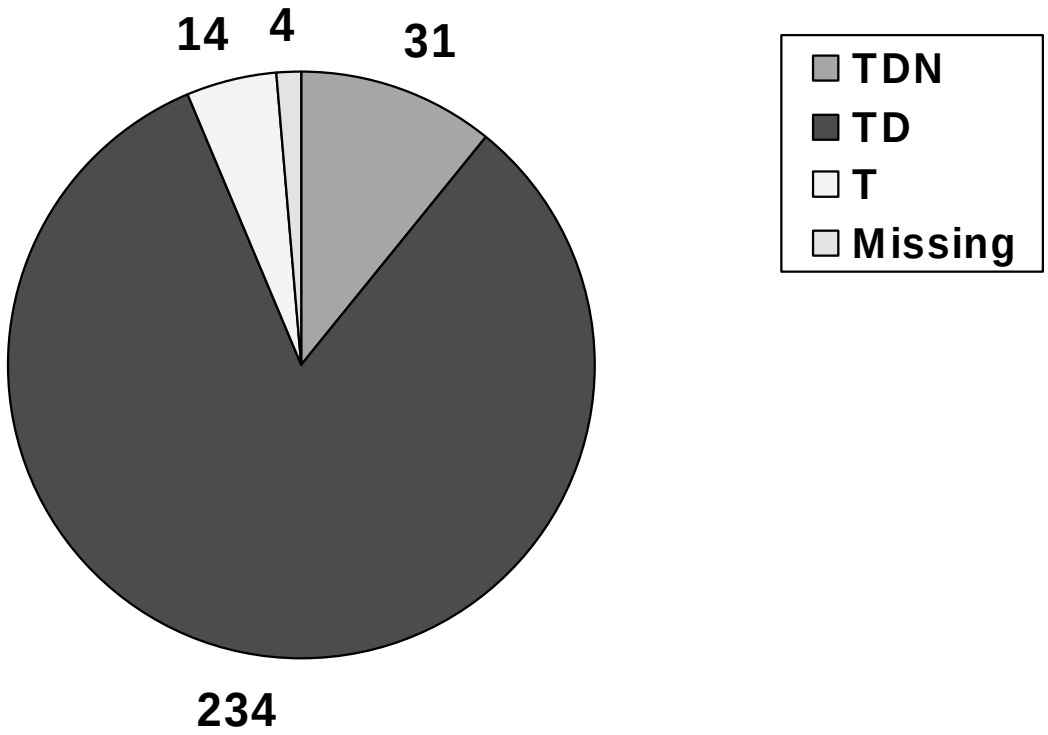
Runs per Task (Adhoc Tracks)



Runs per Topic Language (Adhoc Tracks)



Topic Fields (Adhoc Tracks)



Pooling

- "Tool" to handle the size of relevance assessment work
- 155 of 283 runs assessed
- GIRT task had all runs assessed
- Runs are pooled respecting nearly a dozen criteria:
  - participant's preferences
  - "originality" (task, topic fields, languages, ..)
  - participant/task coverage
  - ..

Preliminary Trends for CLEF 2004 (1)

Machine translation is a popular choice to bridge the language gap

„New(er)“ weighting schemes are gaining support: deviation from randomness, LM, others

Experiments with Finnish decompounding

Experiments with adaption to „new languages“: Finnish & Portuguese

More merging experiments

Experiments with commercial and open source „off-the-shelf“ retrieval systems

Query translation is dominant

Stemming and decompounding less debated – generally accepted?

The most popular bilingual task (X->FR) was hotly contested. Likewise, for the monolingual FR task, covering a language used for a long time in CLEF, very similar performance was achieved by a number of groups

CLEF 2004 vs. 2003

Many participants were back, but as the adhoc tracks were not as „dominant“ anymore, some of these participants concentrated on other tracks

The multilingual track was radically redesigned. Still, there was an advantage for „veteran“ groups, as in earlier CLEF campaigns. Some „new“ veterans have achieved very good results

People think about free resources, a trend introduced last year

Challenges

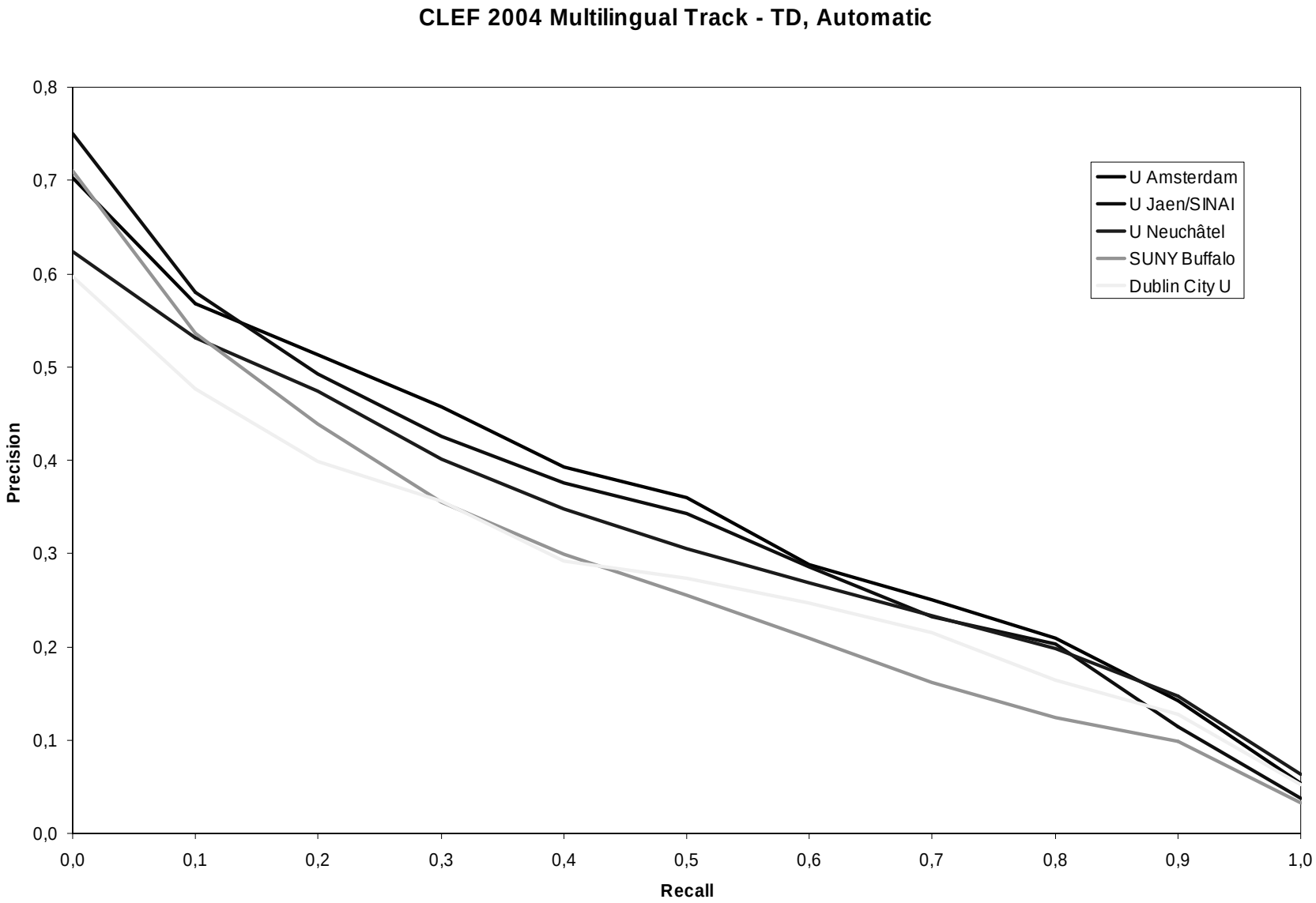
How can we understand the interaction between (black box) MT systems and CLIR better?

How can we systematically bring light into the „jungle“ of merging approaches? How can we systematically test them?

Is the blueprint for successful MLIR stifling innovative alternatives?

How do we avoid over-tuning to the CLEF collections (especially for the monolingual tracks)

How can we work on better transfer of our results to operational systems?


CLEF 2004 Multilingual Track - TD, Automatic
Precision
Recall
0,0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1,0
U Amsterdam
U Jaen/SINAI
U Neuchâtel
SUNY Buffalo
Dublin City U

Bilingual Tasks

Task	Top Perf. (TD)	Diff. To 5th Place
Bilingual X->FI	JHU/APL	-
Bilingual X->FR	JHU/APL	+12.4%
Bilingual X->PT	U Neuchâtel	-
Bilingual X->RU	U Alicante	+138.6%
Bilingual X->EN	U Amsterdam	-

Monolingual Tasks

Task	Top Perf. (TD)	Diff. To 5th Place
Monolingual FI	Hummingbird	+22.4%
Monolingual FR	Hummingbird	+7.5%
Monolingual PT	U Neuchâtel	+19.9%
Monolingual RU	U Alicante	+26.9%

Conclusions

Interesting new experiments due to the language restrictions

Challenges to understand the „black box“ parts of the systems better

Merging, the everlasting issue

-> Where do we go with the adhoc tracks?