

CLEF Interactive Track Overview

iCLEF

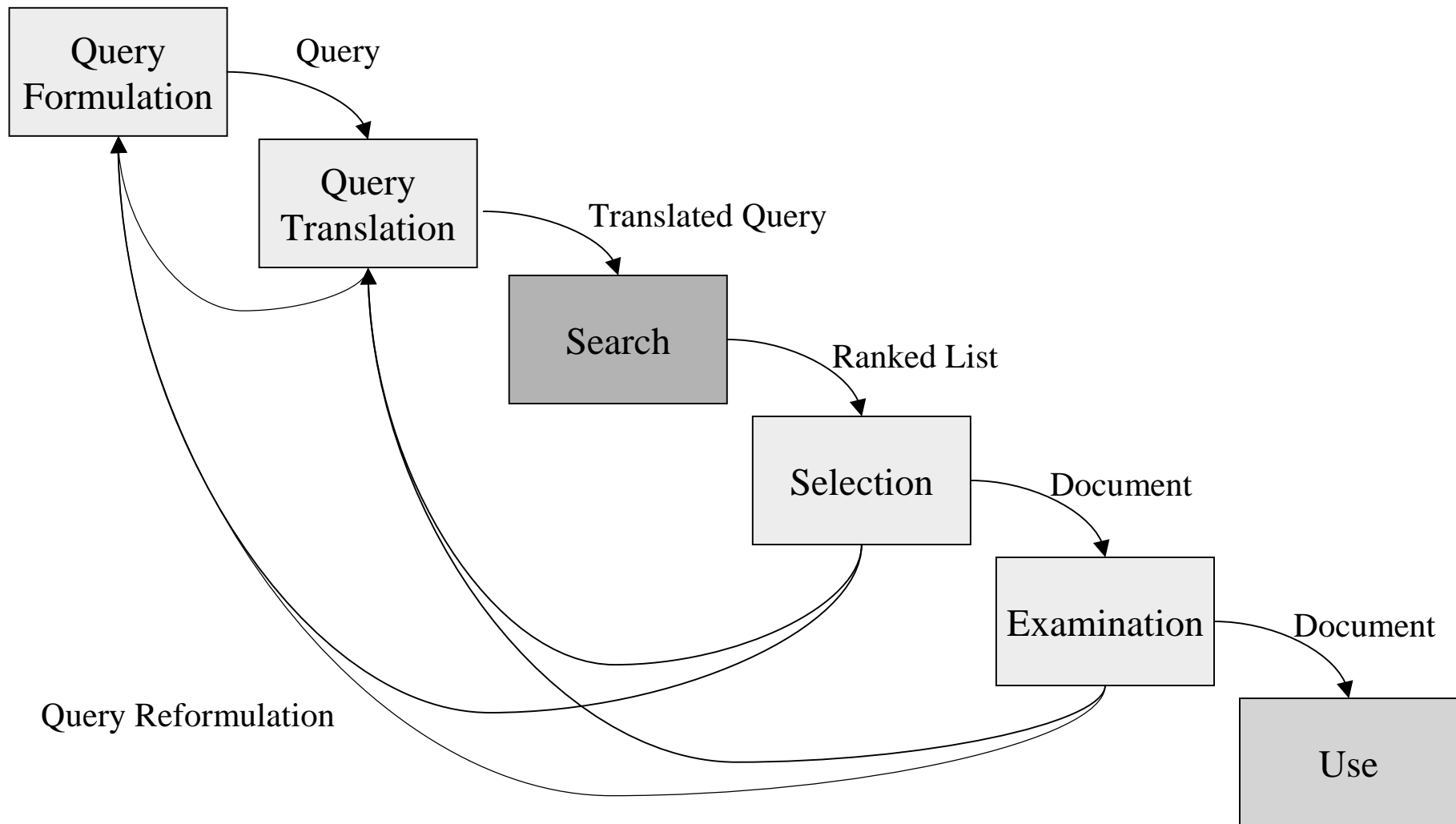
Julio Gonzalo, UNED, Spain

Douglas W. Oard, UMD, USA

Outline

- Goals
- Track design
- Participating teams
- Results

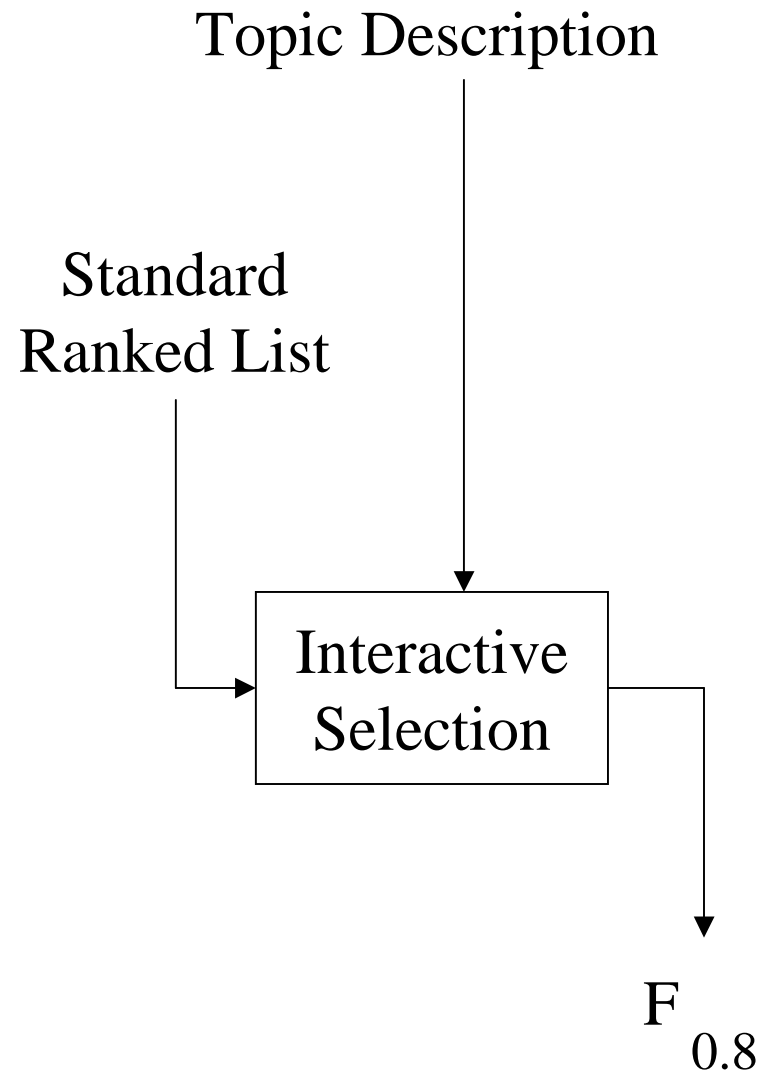
An Interactive CLIR Process Model



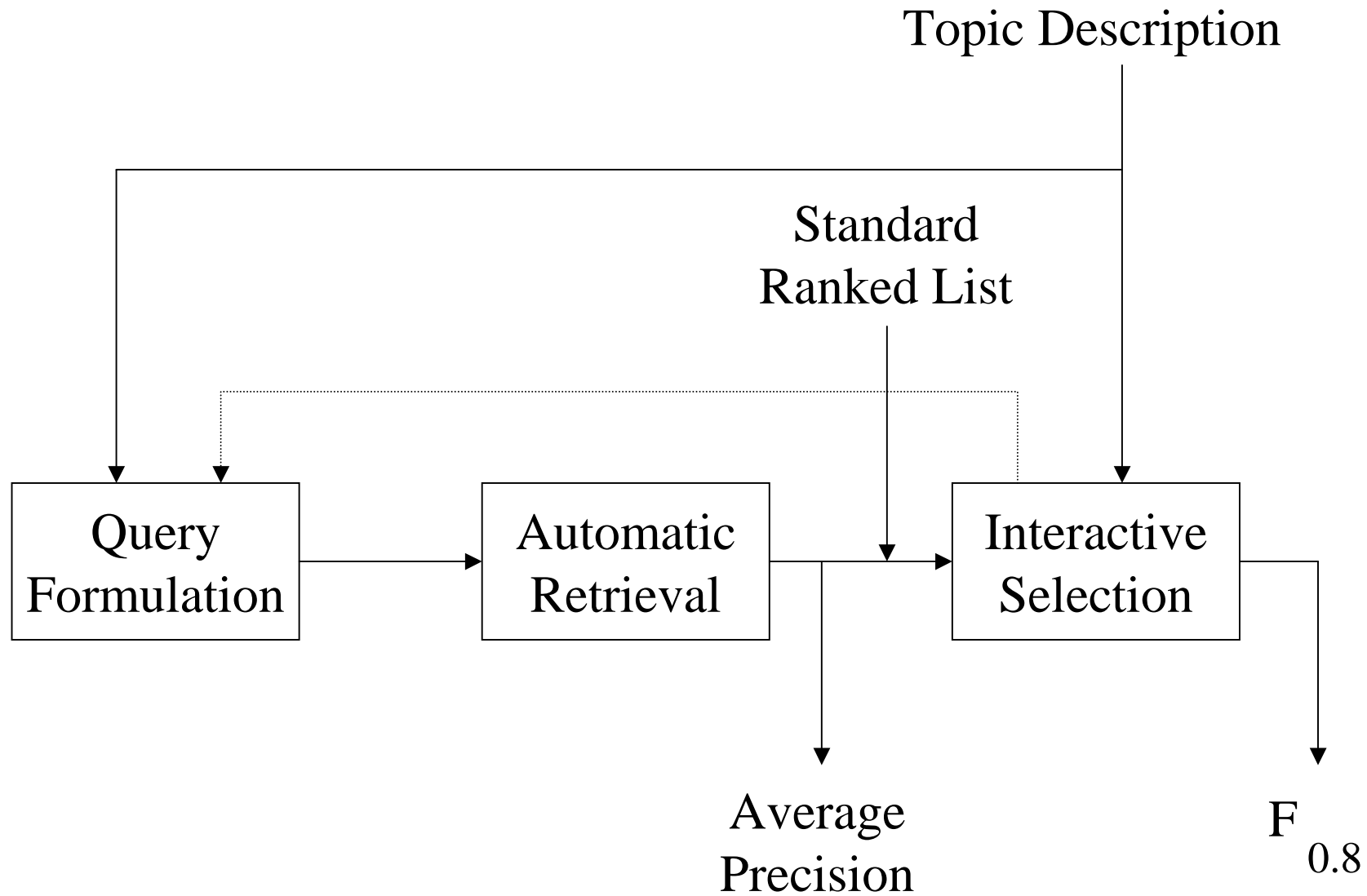
iCLEF Goals

- Track design
 - Component evaluation (iCLEF 2001)
 - End-to-end evaluation (iCLEF 2002)
- System evaluation
 - Support for interactive document selection
 - Support for query creation
 - Support for interactive query reformulation

iCLEF 2001 Experiment Design



iCLEF 2002 Experiment Design



Task Menu

- Formulate a query
- Examine retrieved documents
- Iterate as desired
- Record judgments
 - Degree of relevance: Highly, somewhat, not
 - Judgment confidence: High, medium, low
- Allowed 20 minutes per topic
 - About 3 hours w/training, questionnaires, breaks

Topics

- Four “broad” (multifaceted) topics
 - 1: Genes and Diseases (C053)
 - 2: Treasure Hunting (C065)
 - 3: European Campaigns against Racism (C056)
 - 4: Hunger Strikes (C080)

Example: Topic 3 in English

Title: European Campaigns against Racism

Desc: Find documents that talk about campaigns against racism in Europe.

Narr: Relevant documents describe informative or educational campaigns against racism (ethnic or religious, or against immigrants) in Europe. Documents should refer to organized campaigns rather than reporting mere opinions against racism.

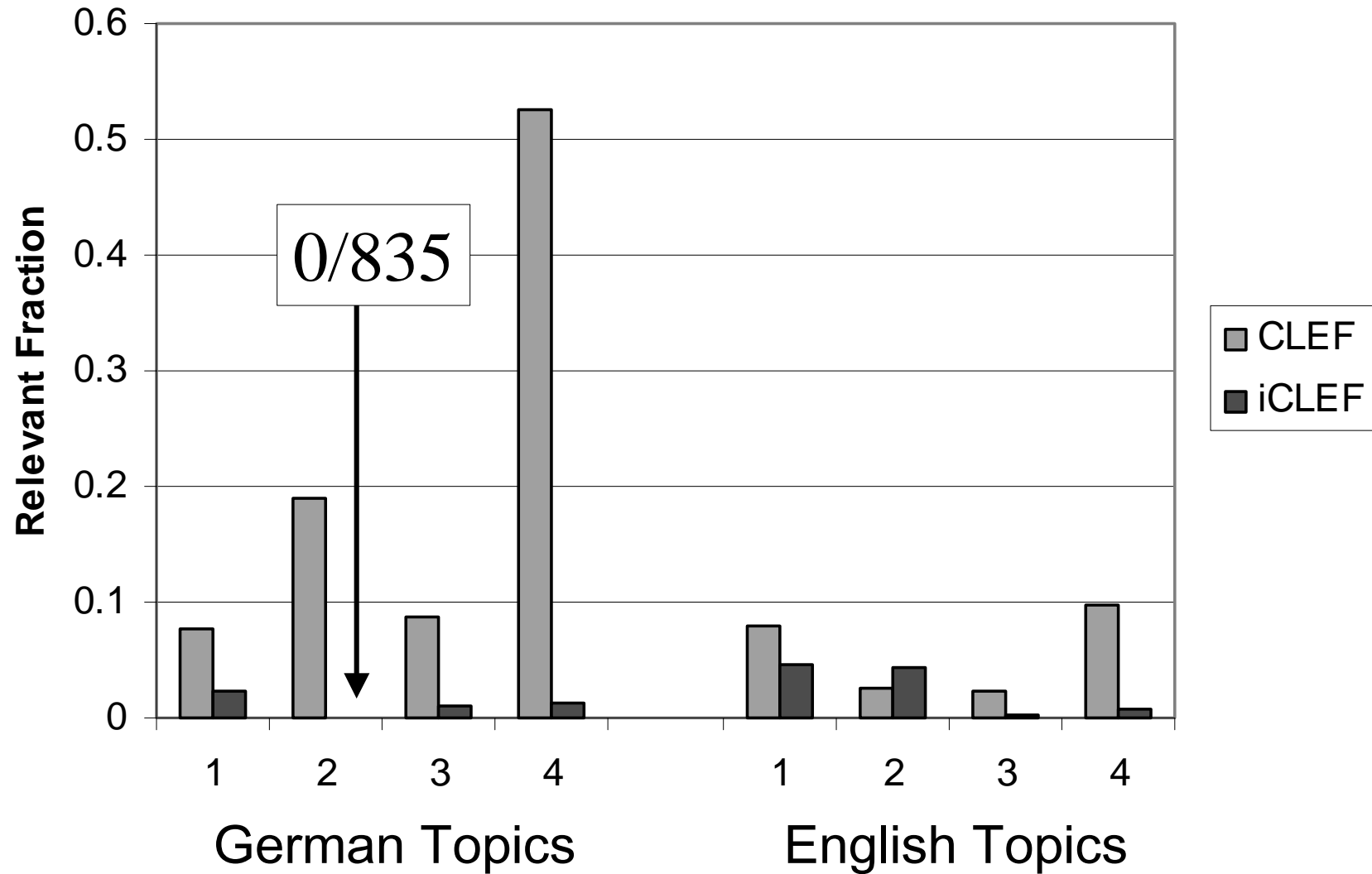
Document Collections

- CLEF-2001 collections
 - German (Der Spiegel and SDA only)
 - English
- Systran baseline translations
 - German to English
 - English to Spanish
- CLEF-2002 collections
 - Finnish
 - Swedish

Relevance Judgments

- Start with CLEF-2001 judgments
 - No judgments for Finnish or Swedish
- Enrich pools with
 - Top 20 documents from every iteration
 - Every document judged by a user
- Judge all additions to the pools

Pool Density



Measures of Effectiveness

- **Query Formulation: Uninterpolated Average Precision**
 - Expected value of precision [over relevant document positions]

$$AP = E_j \left[\frac{j}{r(j)} \right]$$

- Interpreted based on query content at each iteration

- **Document Selection: Unbalanced F-Measure:**

- P = precision
- R = recall
- $\alpha = 0.8$ favors precision
 - Models expensive human translation

$$F_{\alpha} = \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}}$$

Effects on CLEF Measures

- System
 - What we seek to measure
- Topic
 - Sample topic space, compute expected value
- Topic+System
 - Pair by topic and compute statistical significance
- Collection
 - Repeat the experiment using several collections

Additional Effects in iCLEF

- Learning
 - Vary topic presentation order
- Fatigue
 - Vary system presentation order
- Topic+User (Expertise)
 - Ask about prior knowledge of each topic

Participant


Task Order

Presentation

Order

1	1 - 4	3 - 2
2	2 - 3	4 - 1
1	1 - 4	3 - 2
2	2 - 3	4 - 1
1	4 - 2	1 - 3
2	3 - 1	2 - 4
1	4 - 2	1 - 3
2	3 - 1	2 - 4

System Key

System A: 

System B: 

iCLEF 2001 Research Questions

Document Selection

- Maryland (USA)
 - Are word-by-word translations useful?
- Sheffield (UK)
 - How good is Systran?
- UNED (Spain)
 - Can phrase translation beat Systran?

iCLEF 2002 Research Questions

Document Selection

- Alicante/Jaen (Spain)
 - Can summaries beat full documents?
 - SICS (Sweden)
 - What happens when Swedes read English?
-

End-to-End

- Maryland (USA)
 - Is user-assisted translation helpful?
- Sheffield (UK)
 - Is user-assisted translation helpful?
- UNED (Spain)
 - How useful are phrase translations?

Official Results

