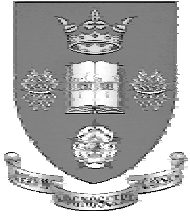


Image CLEF

Daniela Petrelli

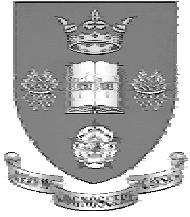
Mark Sanderson & Paul Clough
[m.sanderson|p.d.clough]@shef.ac.uk

ir.shef.ac.uk



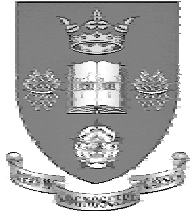
Why do CLIR for images?

- Images can be described and searched by text
 - Captions but often also author, title, keywords
 - Text can be searched cross-language
- Images can be judged relevant without reading
 - No need to know the language of image description
- Potential for other objects: video, source code...



Why do images need CLIR?

- Content-based image retrieval isn't going to solve things soon
 - Need text/metatag information
 - Cheaper to use CLIR than translate captions
- Images is everyday business
 - The media buy/sell images daily
 - More searchers, more money



Research challenges

- Genre
 - Image captions are short
 - Need for more language processing?
 - Image queries seem to often hold verbs
- User interaction
 - Potential for extensive use and study of user feedback
- Opportunities for MT
 - Translation of captions
- Links with content-based searching
 - Images as cognates?

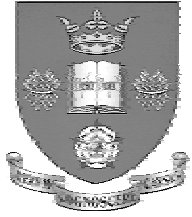
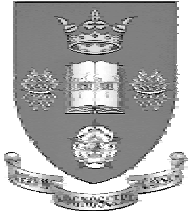


Image CLIR pre-track

- St. Andrews collection
 - *In-principle* agreement to release a 40,000 image collection
 - Black & white historical images of Scotland
 - Large, fielded English text captions
 - Thumbnail images
 - Full size unlikely
 - Probably not good for content retrieval
 - Queries: 50, English, German?, Portuguese?
French and Italian to come as a Christmas gift ;-)
 - You can translate how you like



Proposed Timetable

- December
 - Register your interest
 - Sheffield confirms track running
- January
 - Collection released
 - TREC SGML format
- March
 - Your rankings back to Sheffield
- July
 - Results announced