

---

# MSc Project — Image Retrieval Using Natural Language and Content-Based Techniques

Kate Byrne

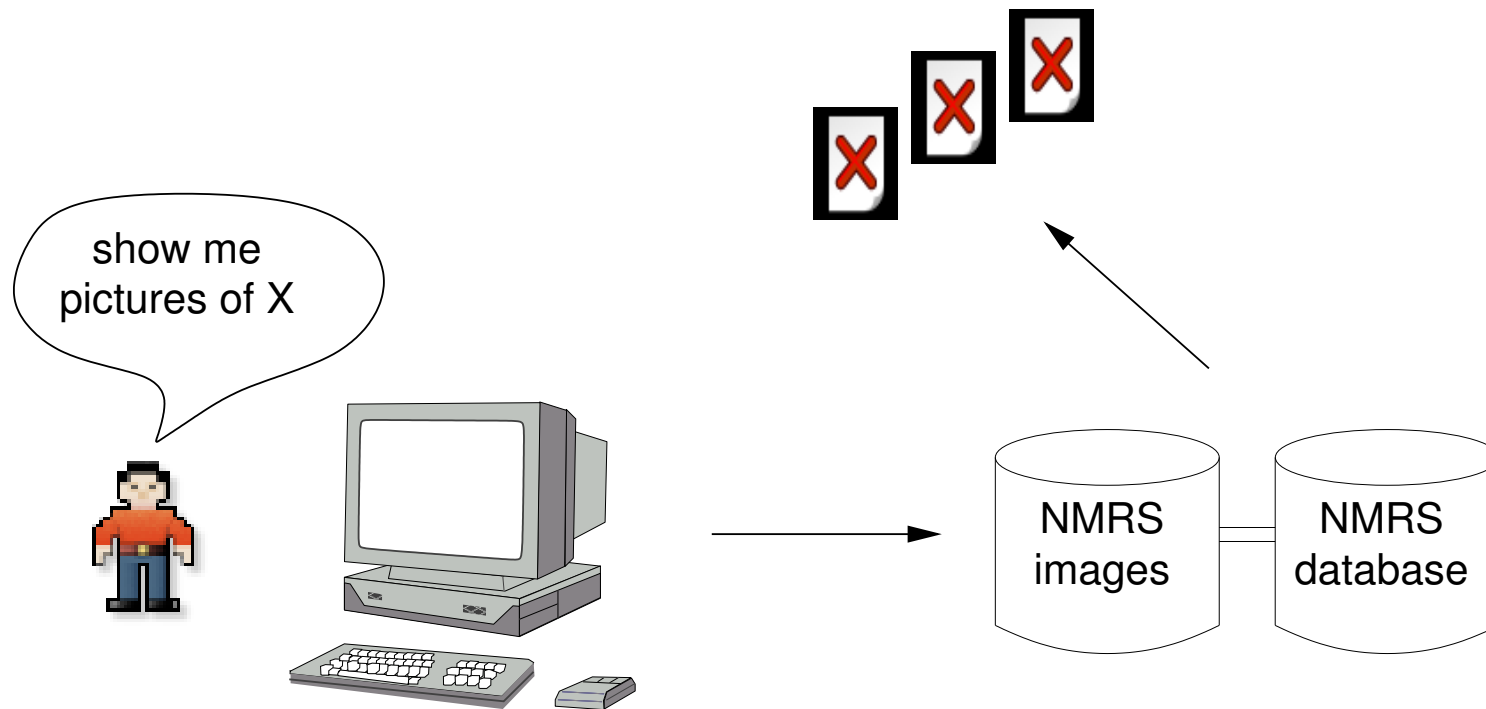
September 21, 2003



## Agenda

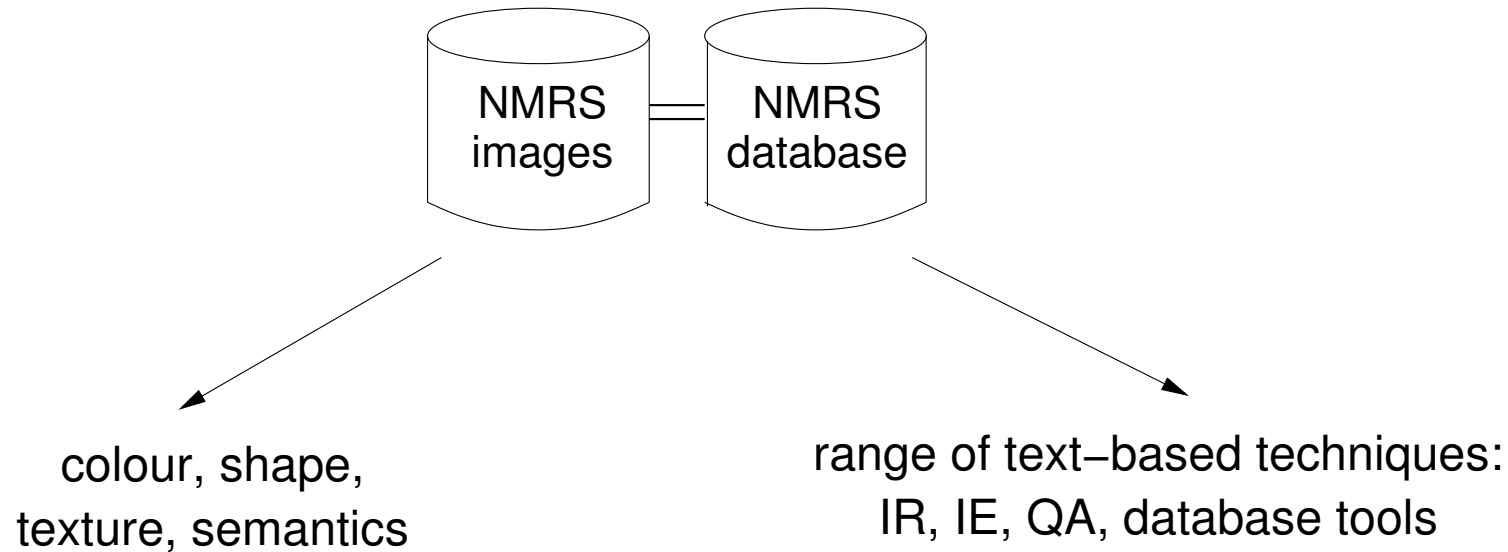
- The Task — what was the aim of the project?
- The Tools — combining and comparing methods
- Demo — CANTRIP in operation
- Evaluation — summary of results

# Aim of the Project: Image Retrieval



## Variety of Approaches

Start with physical image content or with text descriptions?



# Text-based Methods 1



- IR — inverted indexes with TF-IDF weighting, as in Web search engines
  - *TF-IDF weight = term frequency × inverse document frequency*
  - high TF-IDF weight  $\Rightarrow$  several times in each document, but rarely in the corpus as a whole  $\Rightarrow$  word occurs in “clumps”
  - such words are good for discriminating between different documents
- IE — extracting structure from data, eg by finding Named Entities
  - people, places, organisations, dates, numeric amounts
  - domain-specific phrases like “unenclosed settlement” or “cruck-framed”

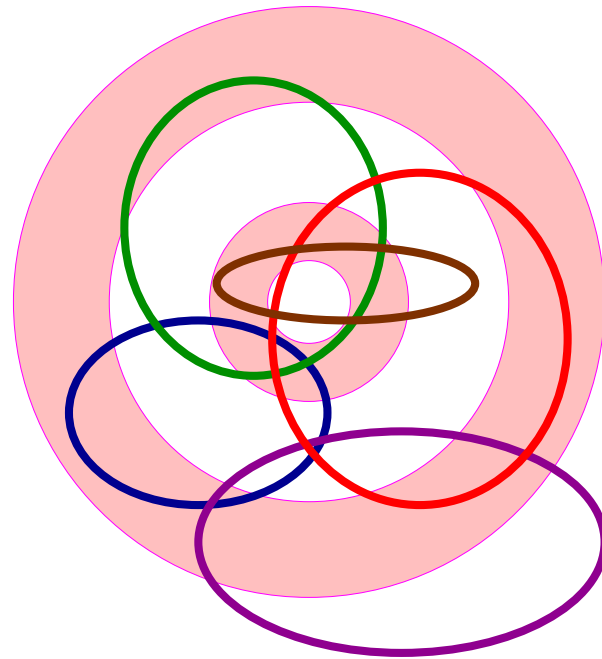
## Text-based Methods 2

- QA — question answering tools, eg query expansion
  - find preferred or related terms from thesaurus
  - user asks for “greenhouse”, system also searches for “glasshouse” and “botanic garden”
  - sometimes helpful, sometimes not
- Database tools — structured queries against fields, with SQL
  - distinguish between type of site and name of site etc
  - use inverted indexes on individual text fields in database



## Combining and Evaluating

TF-IDF  
+  
NER

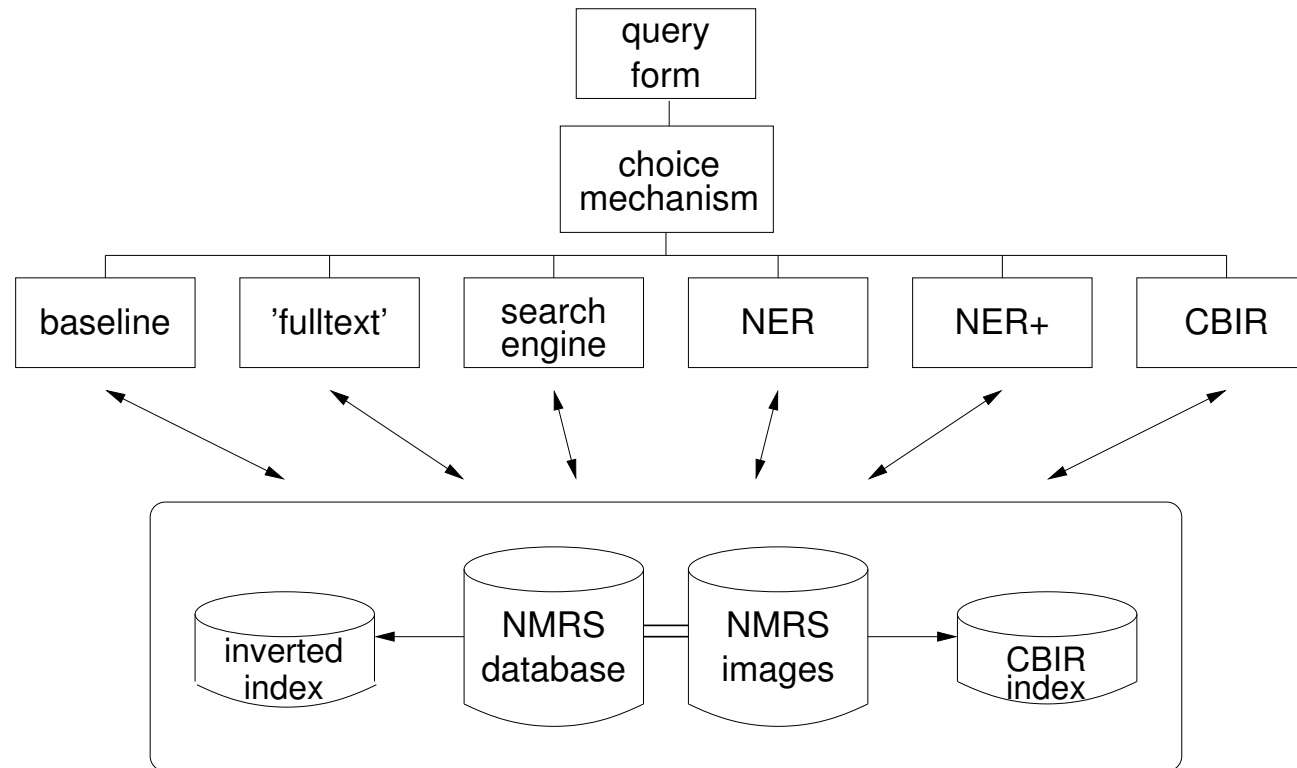


CBIR +  
text-based IR

relational structure + unstructured text

# Implementation

Six versions of CANTRIP, to test different approaches





Demo time...

## Results

	<b>Time</b>	<b>Precision</b>	<b>Recall</b>	$score_K$	<b>Accuracy</b>
NER+	5.77	66.88%	100.00%	100.25	52.28%
fulltext	1.96	48.13%	100.00%	72.69	40.30%
search engine	3.17	29.38%	75.00%	48.63	29.84%
baseline	1.63	37.50%	62.50%	40.13	26.14%
NER	5.00	31.56%	75.00%	39.06	25.68%
CBIR	7.98	60.28%	36.11%	23.63	18.97%

