UA in ImageCLEF 2005

Maximiliano Saiz Noeda

Index

- System
  - Indexing
  - Retrieval
- Image category classification
  - Building
  - Use
- Experiments and results
- R2D2 Joint participation
- Conclusion & future work

System

- Based on textual captions
  - 11(13) languages: Chinese (simplified and traditional), Dutch, English, French, German, Greek, Italian, Japanese, Portuguese, Russian and Spanish (europ. and latinoam.)
  - Web translator investigation for better performance on 2004 dataset
- Combination of probabilistic and automatic extracted knowledge
- Use of Xapian, a probabilistic IR engine.
  - Automatic feedback using first 23 relevant documents
- Two phases
  - Indexing (offline)
  - Retrieval (online)

System: Indexing

- Creation of three indexes
  - Stems
  - Words
  - Stem bigrams
- Token weighting depending on:
  - Weight of field where token is contained
  - Upper or lowercase in token first letter

System: Retrieval

- Three retrievals, one in each index and combine them
- In each retrieval:
  - First basic retrieval
  - Feedback → second enriched retrieval
  - Add category information
  - Combine three results and sort

System: Retrieval (II)

- Combination of three indexes to improve performance
  - Assigning weights to each retrieved list
  - Obtaining their mean average

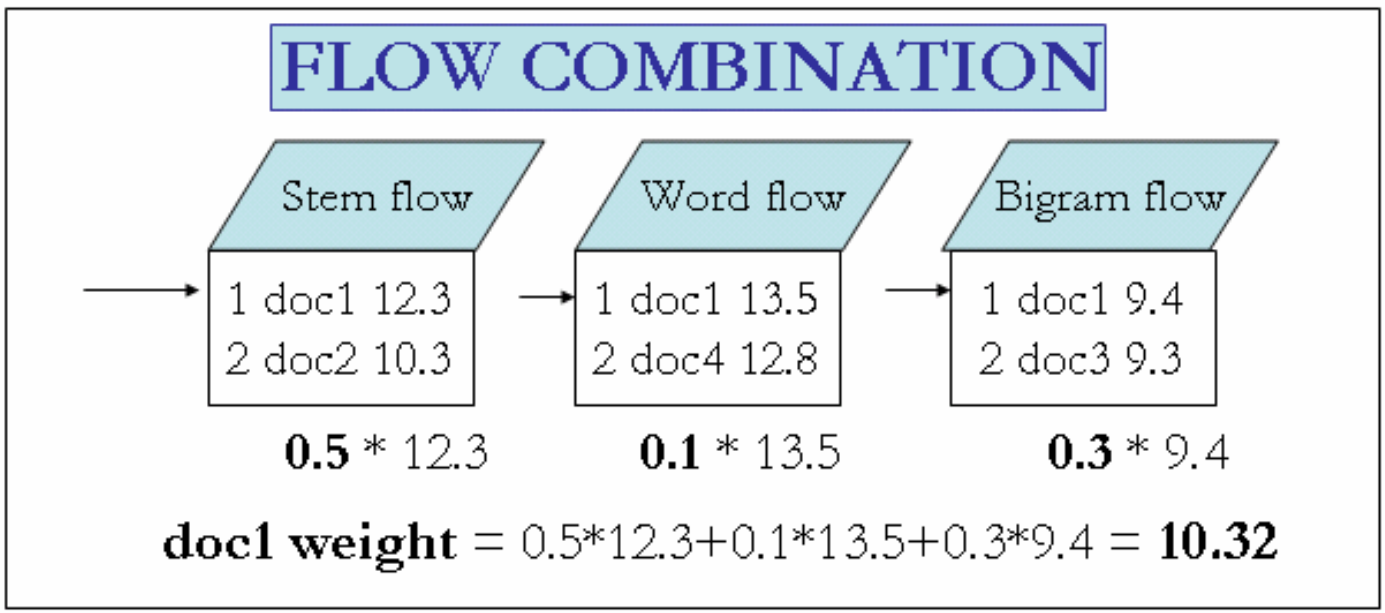


Image category classification

- Knowledge base to enrich IR
- Automatic created from St. Andrews Corpus
- Extract categories related with a query and retrieve documents with the same categories

Image category building

- Create a document with words that appear with a concrete category

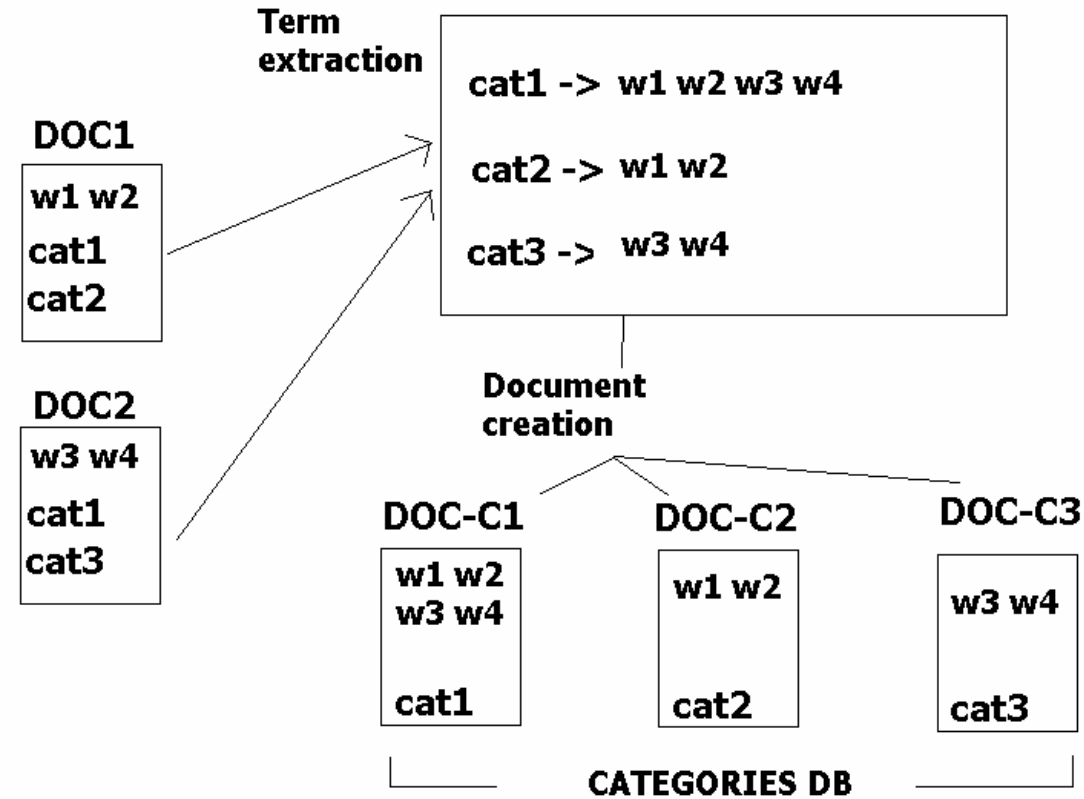
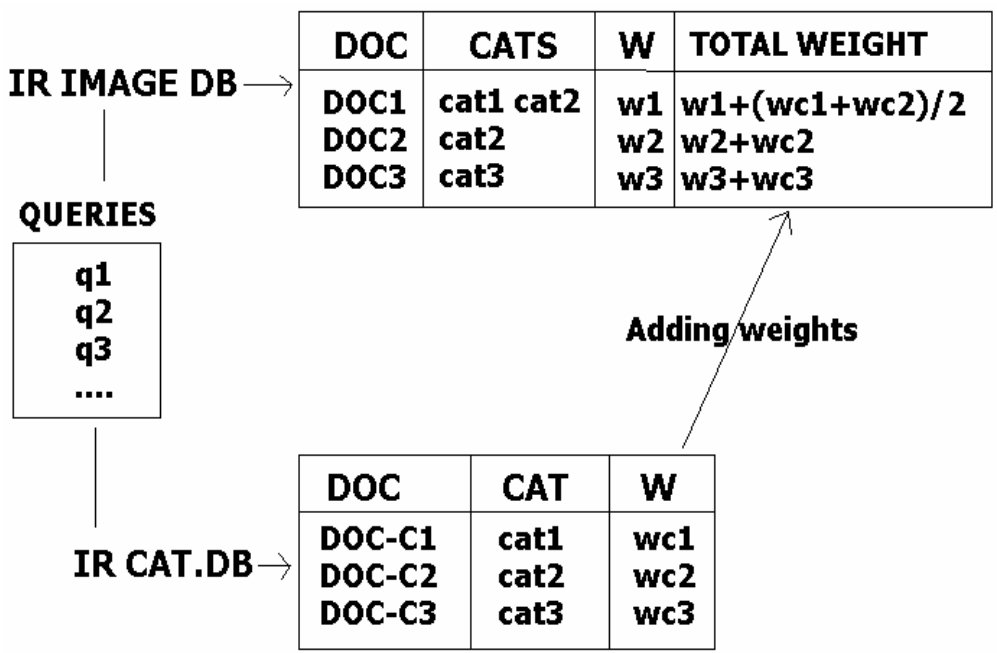


Image category use

- Queries are posed to category index to retrieve a list of "category-documents"



Experiments and Results

- Different features combined to create more than 100 experiments

	stem	word	bigram	cats	feed	2005 MAP	Prec. 10	Recall
Baseline	X					0.3944	0,5143	0,8377
Experiment1	X	X	X			0.3942	0,5536	0,8399
Experiment2	X	X	X		X	0.3909	0,5429	0,8246
Experiment3	X	X	X	X	X	0.3966	0,5429	0,8246
BEST 2005	-	-	-	-	-	**0,4135**	**0,55**	**0,8434**

R2D2 Joint participation

- UA, SINAI (Jaén) and UNED (Madrid)
  - UA+SINAI
    - English, Dutch, French, German, Italian, Russian and Spanish
  - UA+SINAI+UNED
    - Spanish
- Voting method based on combining document weights from three systems
  - weight normalization (document weight/max. weight)
  - weight adding (for each document in each list)
  - priority for documents in all the lists

Joint participation: results

Language	UA	SINAI	UNED	UA-SINAI	UA-SINAI-UNED
English	0.3966(14)	0.3727(30)	-	**0.4080(7)**	-
Dutch	0.2765(8)	0.3397(2)	-	**0.3435(1)**	-
French	0.2621(6)	0.2864(1)	-	0.2630(5)	-
German	0.2854(7)	0.3004(4)	-	**0.3375(1)**	-
Italian	0.2230(4)	0.1805(11)	-	**0.2289(2)**	-
Russian	0.2683(3)	0.2229(11)	-	0.2665(5)	-
Spanish (eur)	0.2105(12)	0.2416(5)	0.3175(1)	0.2668(4)	0.3020(2)
Spanish (lat)	0.3179(2)	0.2967(8)	0.2585(17)	**0.3447(1)**	0.3054(4)

Thank you very much!