

Applying Logic Forms and Statistical Methods to CL-SR Performance

**Rafael M. Terol, Patricio Martínez-Barco
and Manuel Palomar
GPLSI**

Natural Language Processing and Information Systems Group

CLEF 2006

Alicante, Spain, 20-20 September 2006



**Universitat d'Alacant
Universidad de Alicante**

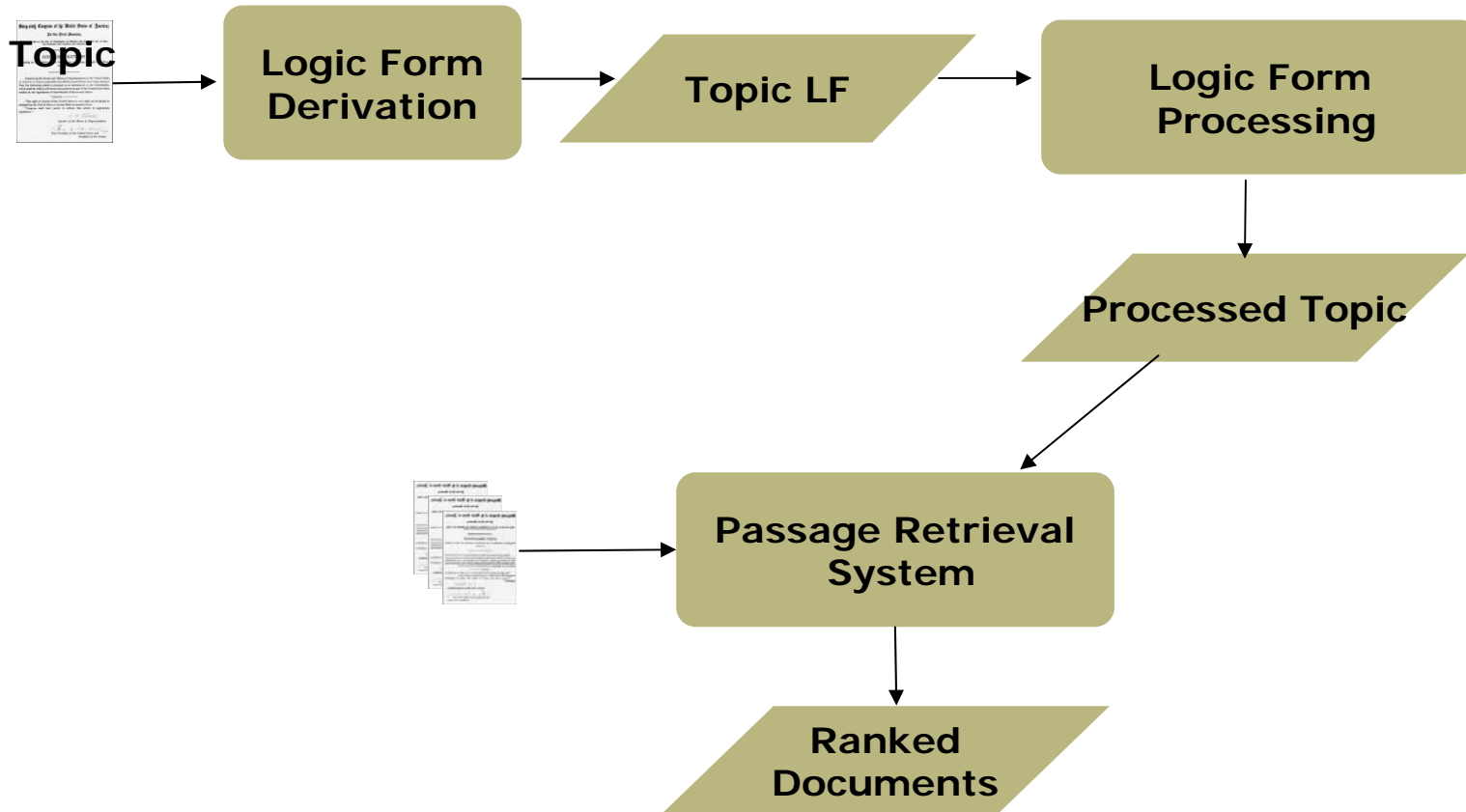


Index

- **System Architecture**
 - Topic Processing
 - Information Retrieval
- **Results**
- **Conclusions**



System Architecture



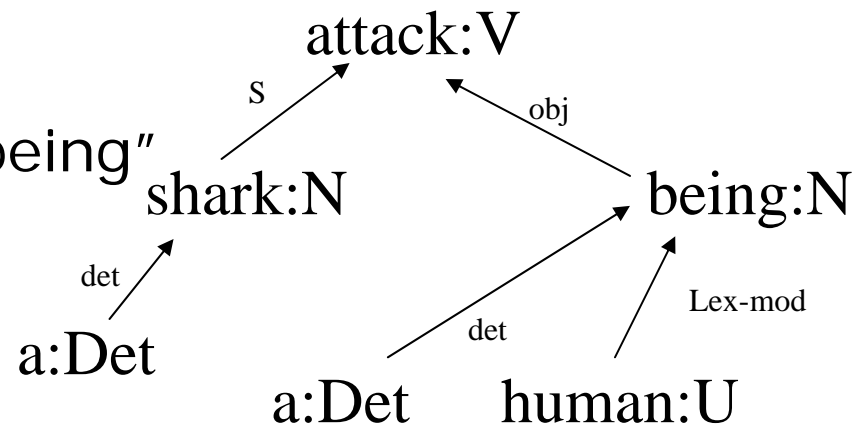
System Architecture

- **Logic Forms Derivation**

- Similar to the eXtended WordNet format
- For each sentence:
 - Obtain dependency relations between words
 - Analyse dependency relations

- **An example**

“A shark attacked a human being”



- Dependency tree
- Logic Form

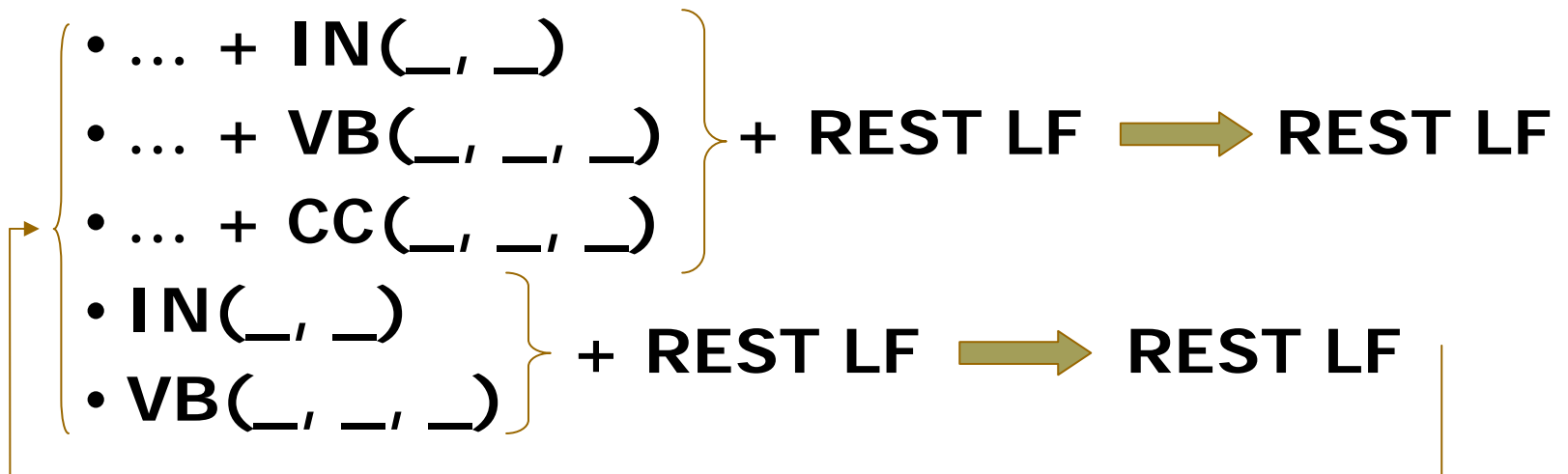
shark:NN(x1) attack:VB(e1,x1,x3)

human:NN(x2) NNC(x3,x2,x4) being:NN(x4)



System Architecture

- Logic Form Processing
 - Applied to Description and Narrative Topic Fields
 - Goal: To remove 'non-relevant' topic terms according to the logic structure of the field



* without NN in the topic field

System Architecture

- **Example**

- **Field:**

- Describe survival mechanisms of children born in 1930-1933 who spend the war in concentration camps or in ...

- **Logic Structure:**

- describe:VB(e2, x11, e1) survival:NN(x1) NNC(x8, x1, x9)
mechanism:NN(x9) of:IN(x8, x2) child:NN(x2)
bear:VB(e1,x8, x10) in:IN(e1, x4)

- **Updated field:**

- children born in 1930-1933 who spend the war in concentration camps or in ...

System Architecture

- **Passage Retrieval System**
 - IR-n System
 - Fernando Llopis and Elisa Noguera
 - University of Alicante
 - 30 words per sentence
 - 4 sentences per passage
 - Overlapping of 1 sentence

Results

run	map	R-prec	bpref	rr	p5	p20	p100	p1000
UA_TDN_ASR06BA1A2LF	0.0369	0.0757	0.0651	0.1800	0.0882	0.1029	0.0821	0.0262
UA_TDN_ASR06BA1A2	0.0365	0.0714	0.0640	0.2255	0.0882	0.0956	0.0785	0.0260
UA_TD_ASR06BA2	0.0328	0.0727	0.0660	0.1681	0.1000	0.0868	0.0700	0.0264
UA_TDN_ASR06BA2	0.0345	0.0756	0.0691	0.1671	0.0765	0.0926	0.0774	0.0278
UA_TD_ASR06B	0.0339	0.0736	0.0792	0.2010	0.1235	0.1088	0.0709	0.0297

Conclusions

- CL-SR System combines
 - NLP methods (topic pre-processing)
 - Statistical methods (IR engine)
- Topic pre-processing by way of logic forms improves IR scores

Applying Logic Forms and Statistical Methods to CL-SR Performance

Thank you very much!



Universitat d'Alacant
Universidad de Alicante

