FACULDADE · DE · CIÊNCIAS UNIVERSIDADE · DE · LISBOA

The University of Lisbon at GeoCLEF 2006

Bruno Martins, Nuno Cardoso, Marcírio Chaves,
Leonardo Andrade, Mário J. Silva

{bmartins, ncardoso, mchaves, leonardo, mjs}
@xldb.di.fc.ul.pt

Objectives for GeoCLEF 2006

- Compare 2 GIR strategies against a standard IR approach:

  1) Geographic text Mining

  2) Augmentation of geographic terms

Geographic text mining

1) Mining geographic references from text.

2) Assign a scope to each document.

   (“*One sense per discourse*” assumption.)

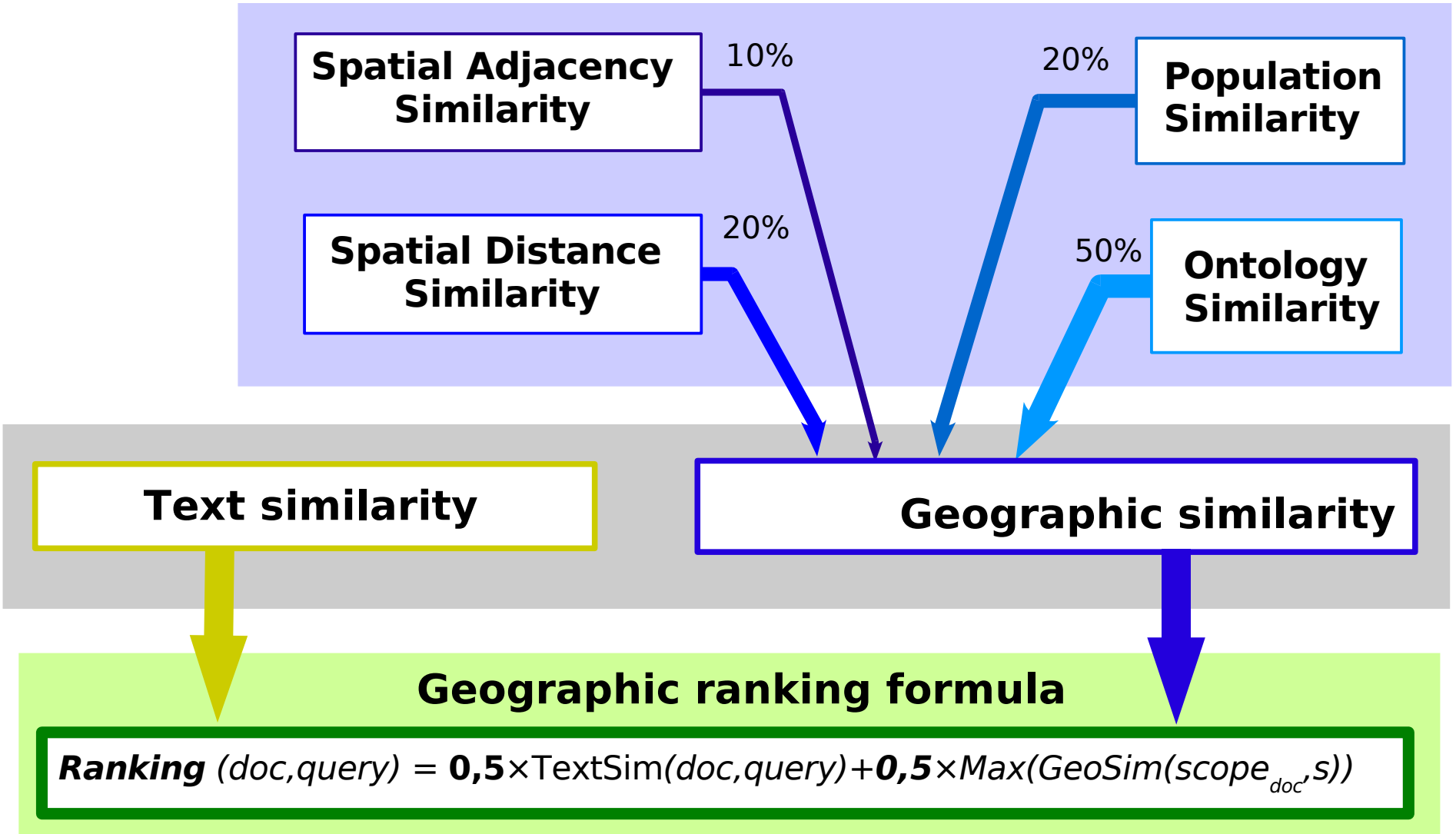
3) Convert CLEF topics into

   <*what*, *relation*, *where*> triples.

4) Assign scopes to **<*where*>** terms.

5) GeoIR using ***Geographic ranking formula***.

Geographic text mining



Geographic text mining

Topic

Forest Fires in Northern Portugal

<*what, relation,where*>

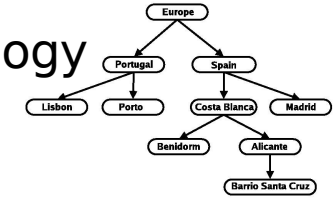
Forest Fires | in | Northern Portugal

Term expansion

Query expansion module

forest fires burn flame fireman water tree shrubs green area ...

Ontology



Assigning scopes to **<where>**

Northern Portugal
GeoscopeID: 12345

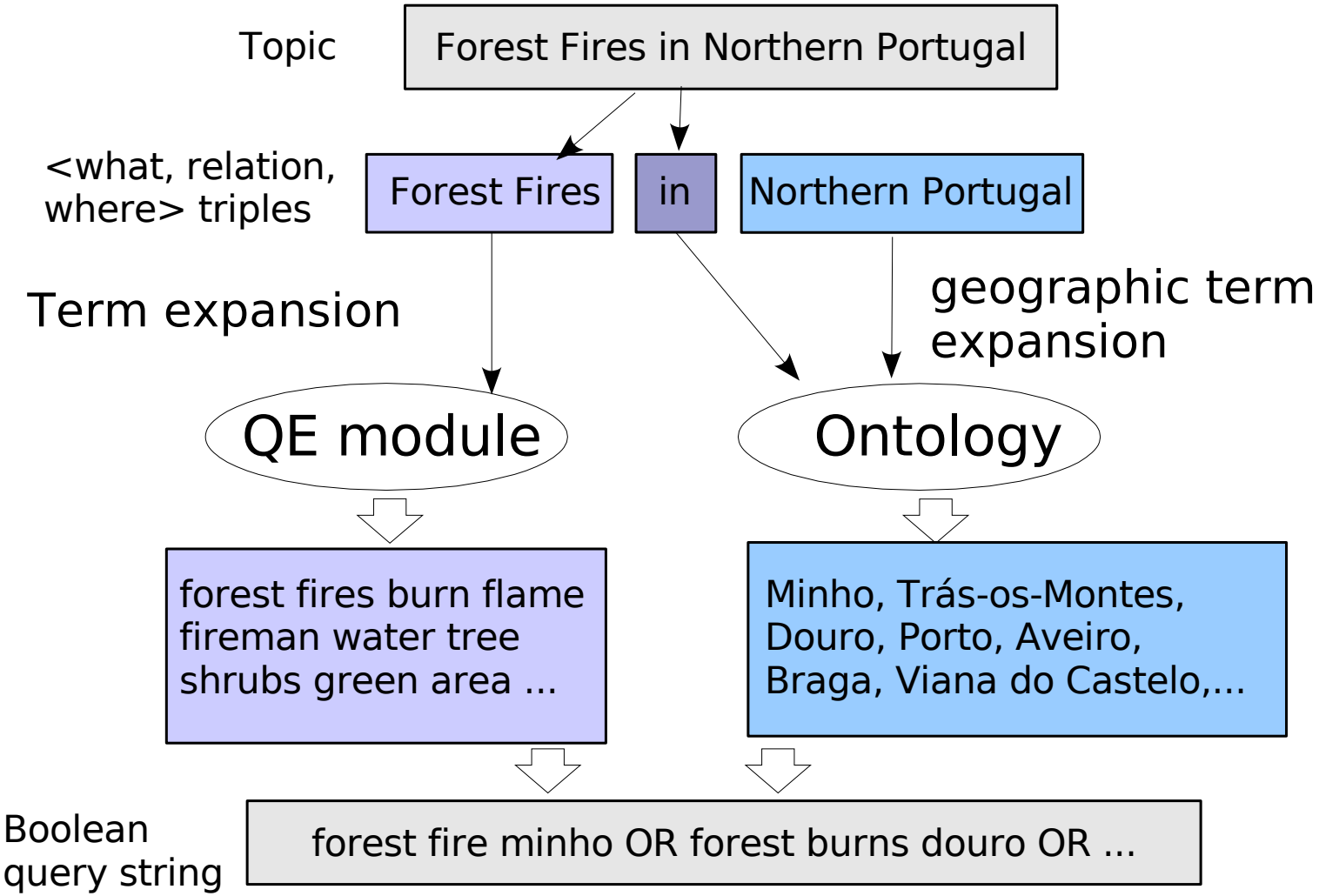
GIR using geographic ranking formula

forest fire@GID:12345 OR forest burn@GID:12345 OR ...

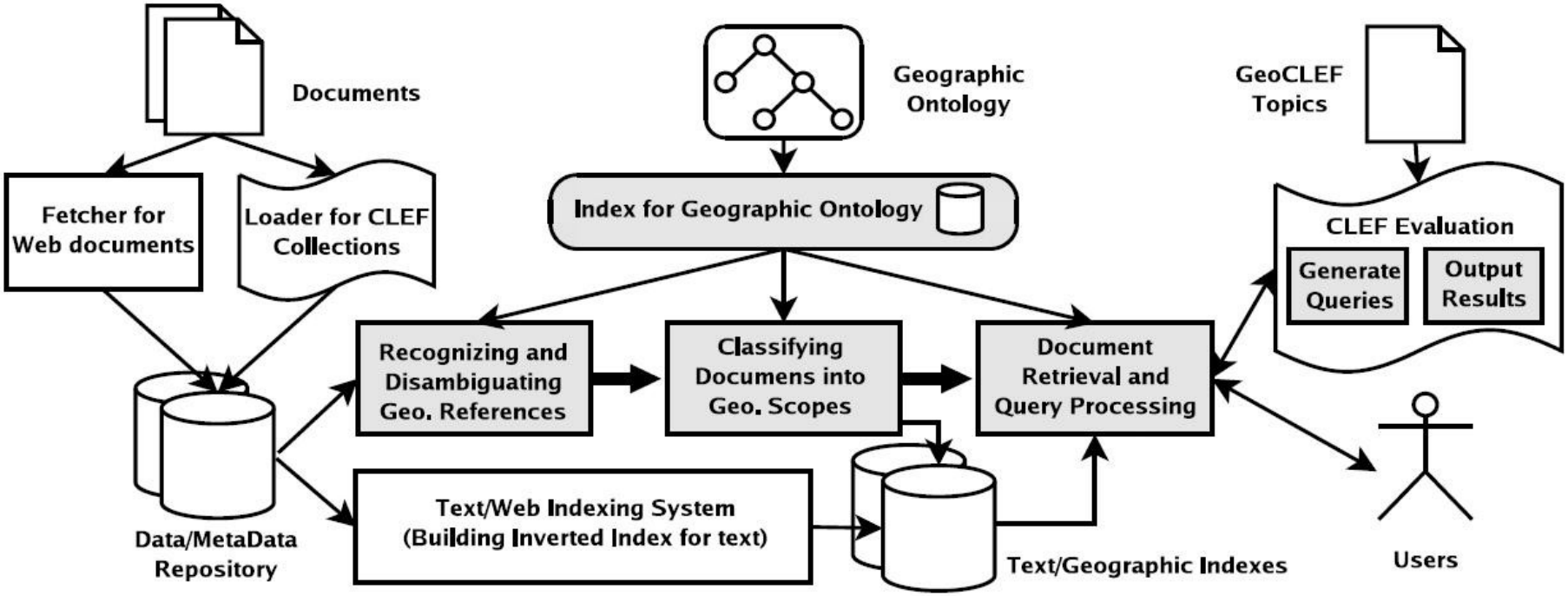
#2: Augmenting geographic terms

1) Convert CLEF topics into <*what, relation, where*> triples
2) Augmentation of <*where*> terms, i.e., expanding terms from our geographic ontology, using <*relation*> criteria.
3) Use BM25 term ranking for <*what*> and <*where*> terms

#2: Augmenting geographic terms



Geo-IR system architecture



Stage 1: Data Loading

Stage 2: Indexing and Mining

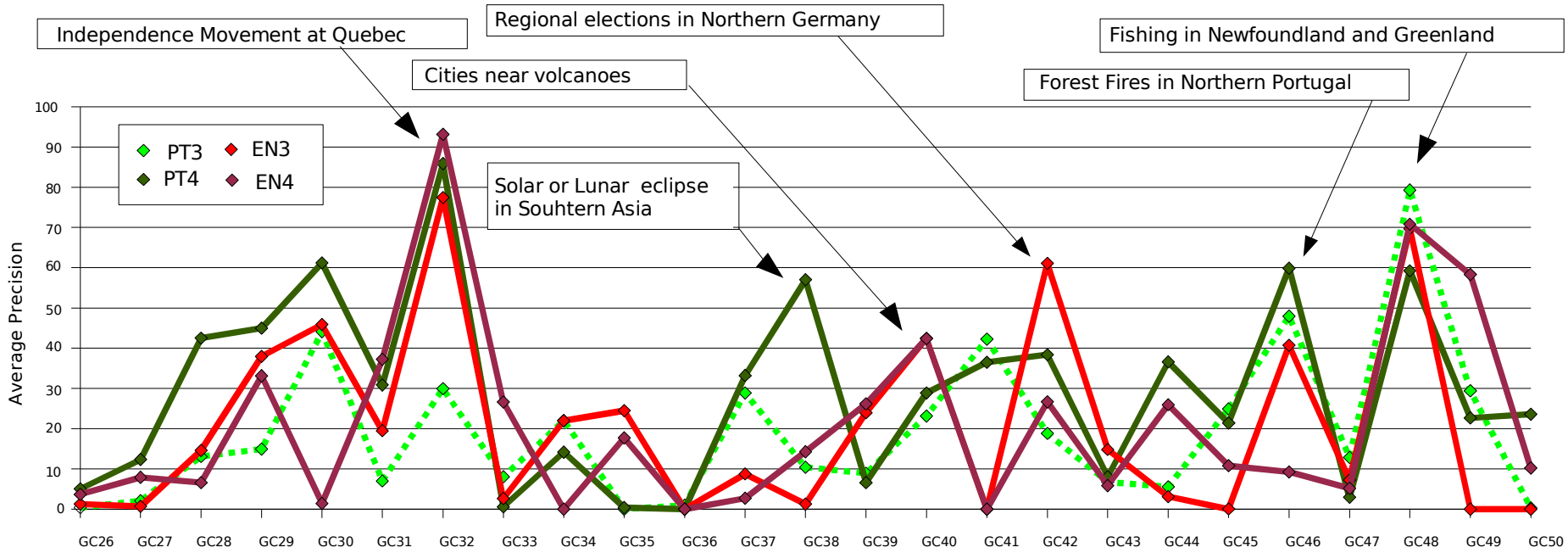
Stage 3: Geo-Retrieval

Runs

#	Description
1	Baseline using manually created queries. BM25 text retrieval.
2	BM25 text retrieval. Queries generated by QE of <*what*> terms, together with original <*where*> terms.
3	Geographic relevance ranking using geographic scopes. Queries generated by QE of <*what*> terms, <*where*> terms matched the scopes (***Strategy #1: Geographic text mining***)
4	Augmentation of <where> terms, using top-10 related concepts from the geographic ontology. Queries generated by QE of <*what*> terms, together with augmented <where> terms (***Strategy #2: Augmenting geographic terms***)

Results

	Monolingual PT				Monolingual EN			
	PT1	PT2	PT3	**PT4**	EN1	EN2	EN3	**EN4**
num_ret	5232	23350	22617	10483	3324	22483	21228	10652
num_rel	1060	1060	1060	1060	378	378	378	378
num_ret_rel	607	**828**	519	624	192	**300**	240	260
MAP	**0,301**	0,257	0,193	**0,293**	**0,303**	0,158	0,208	**0,215**
R-Prec	**0,359**	0,281	0,239	**0,346**	**0,336**	0,153	0,215	**0,220**
bpref	**0,321**	0,254	0,208	**0,306**	**0,314**	0,140	0,191	**0,199**
gm-ap	**0,203**	0,110	0,074	**0,121**	**0,065**	0,027	0,024	**0,047**
P5	0,488	0,416	0,432	**0,536**	**0,384**	0,208	0,240	0,288
P10	**0,496**	0,392	0,372	0,480	**0,296**	0,180	0,228	0,240
P100	**0,218**	0,193	0,162	**0,218**	0,072	0,073	0,068	**0,084**



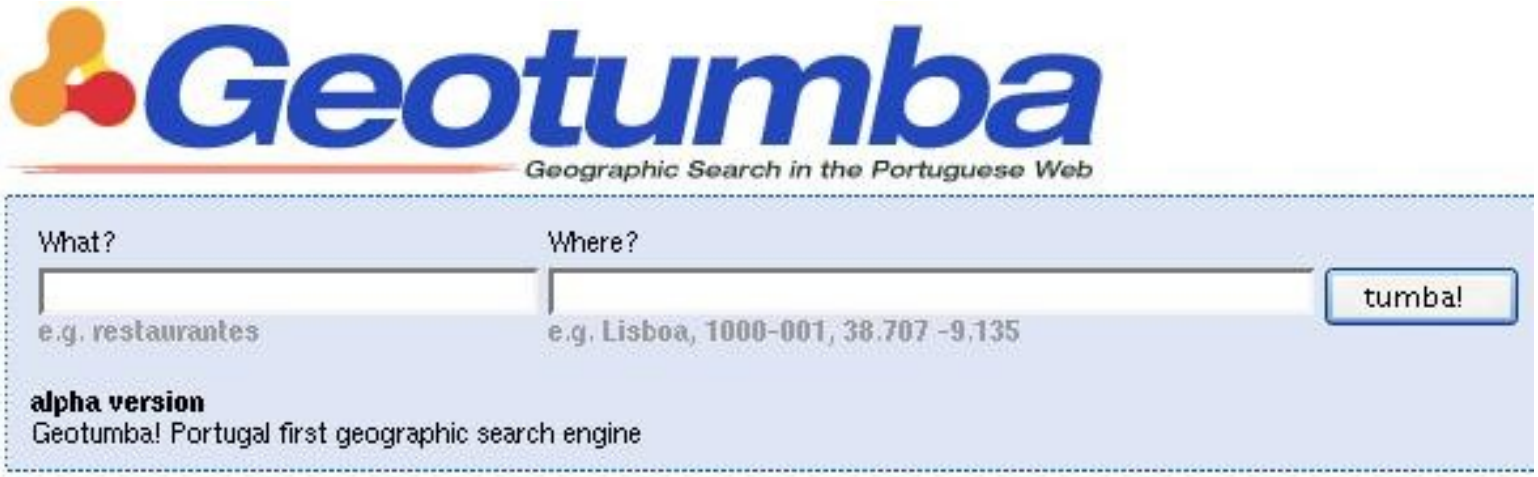
Conclusions

- Strategy #2 runs performed better than strategy #1 runs.
- Text mining approach had some problems.
  - errors in scope assignment.
  - single scope limitations.
- Additional experiments are underway.

The end.

- Questions?

The University of Lisbon at GeoCLEF 2006



- Portuguese local Web search engine
- Project **GREASE** – Geographic Reasoning for Search Engines
- **http://local.tumba.pt**