

UAIC: Participation in LAGI Task

Adrian Iftene

(UAIC: Faculty of Computer Science,
“Alexandru Ioan Cuza” University, Romania)

Track: LogCLEF

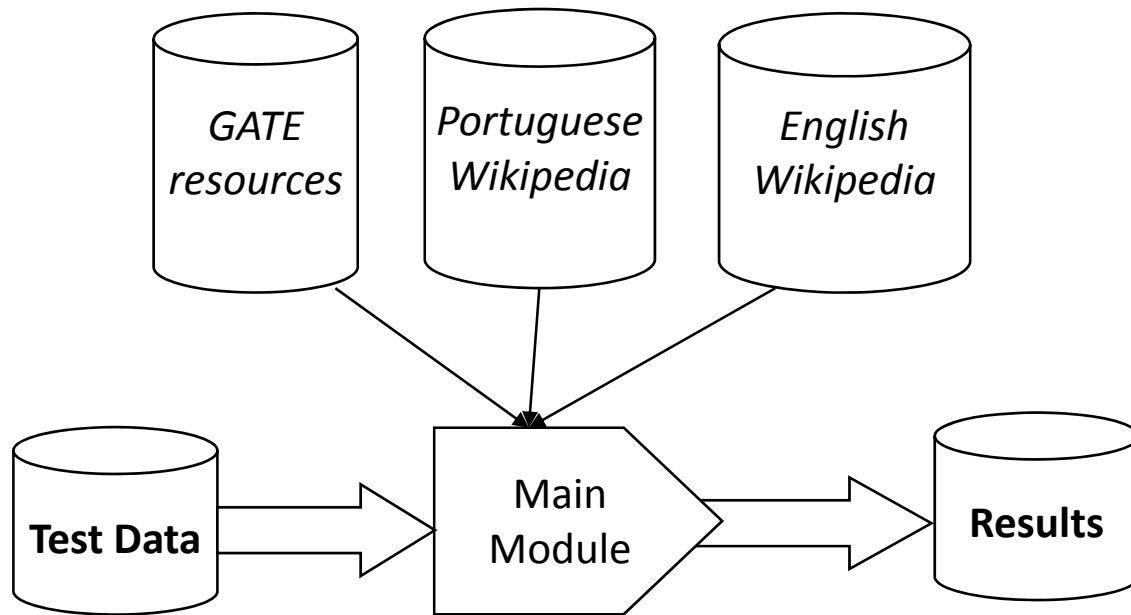


Content

- UAIC System
- Resources
- Main Module
- Submitted Runs
- Conclusions



UAIC System



Resources

- ✦ GATE (Cunningham et al., 2001) cities, countries, small regions, regions, mountains and provinces
- ✦ Internet search

Type	Examples	Entities Number
City	Aaccra, Aalborg, Aarhus, Ababa, Abadan, Abakan, Aberdeen, Abha, ...	143.487
Country	Afghanistan, Afrique, Albania, Albanie, Alderney, Algeria, ...	465
Mountain	Alps, Andes, Himalaya, Pyrenees, Snowdonia	5
Province	Aana, Aargau, Abaco, Abruzzi, Abyan, Aceh, Acores, Acquaviva, ...	2.411
Region	Africa, Algarve, Antarctica, Ashmore and Cartier Islands, Asia, ...	213





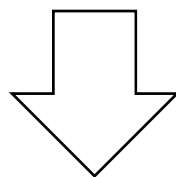
Main Module

- ✦ Test Data Pre-Processing
- ✦ Resources Loading in Cache
- ✦ Geographical Entities Identification

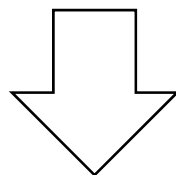


Test Data Pre-Processing

12936 & 16250 & publisher all "london : merton music, 1996."



publisher all "london : merton music, 1996."



publisher all " london : merton music , 1996 . "



Resources Loading in Cache

- ✦ Initial named entity and individual words
- ✦ Example: “Ashmore and Cartier Islands” => 4 entries in cache

Key	Value
ashmore and cartier islands	1
ashmore	3
cartier	3
islands	3



Geographical Entities Identification

- ✦ Main idea is to reduce the search space

Case	Text	In Cache
1	portugal	(portugal, 1)
2	climbing on the Himalaya and other mountain ranges	(himalaya, 1)
3	peter woods	(peter, 2), (peter, 3), (woods, 2), (woods, 3)

- ✦ And to identify sequences of maximal length with geographical entities that have the same values in hash map



Submitted Runs

Test Data	Resource	Rcount	Hcount	Match	Prec	Rec	Fmeas
TEL	GATE	21	29	7	24.14	33.33	28.00
TEL	Wikipedia	21	99	16	16.16	76.19	26.67
Tumba	GATE	35	69	18	26.09	51.43	34.62
Tumba	Wikipedia	35	147	13	8.84	37.14	14.29



Conclusions

- ✦ The system uses like external resources GATE files
- ✦ Main components are data pre-processing, cache loading and geographical entities identification
- ✦ For the future we want to include two new components:
 1. Automatic language identification for the query
 2. Named entities translation with Google Translate



THANK YOU!

