MET - Overview
FIRE 2012

---Rashmi Sankepally

Introduction

Morpheme Extraction Task

- Introduced for the First time in FIRE 2012
- Task offered in ***six*** languages
- Bengali, Gujarati, Hindi, Marathi, Odia, Tamil

Objective

- Objective is to encourage development of systems which discover morphemes in Indian languages
- Many Indian Languages -- morphologically rich
- Morphological Analysis -- important for IR experiments in Indian Languages

Number of Participants

Language	Number of Systems	Participants
Bengali	5	JU, DCU, IIT-KGP, CVPR-1,2
Hindi	2	DCU, IIT-B (team-1)
Marathi	1	IIT-B (team2)
Odia	1	IIIT-Bh
Tamil	1	AUCEG
Language Independent	1	ISM

Total 10 teams from 8 institutes

The Task

The participating systems were asked to submit their Morpheme extraction systems.

System should be such that:

Input - large lexicon (already provided as test data to the participants)

Output - bicolumn file containing tab separated word \t morpheme list

Evaluation methodology

Runs-Information:

- Terrier-3.5
- Corpora, Queries, Qrels: Adhoc FIRE 2011
- ranking model: In_expC2
- Stopwords: FIRE data
- TrecQuery tags: TITLE,DESC (T,D)

Evaluation:

Trec-Eval 8.1 (Metric used: MAP)

MET Results

Bengali:

Team	Language	MAP obtained
Baseline	Bengali	0.2740
JU	Bengali	0.3307
DCU	Bengali	0.3300
IIT-KGP	Bengali	0.3225
CVPR-Team1	Bengali	0.3159
ISM	Bengali	0.3013
CVPR-Team2	Bengali	NA

MET Results

Gujarati:

Team	Language	MAP Obtained
Baseline	Gujarati	0.2677
ISM	Gujarati	0.2824

Hindi:

Team	Language	MAP Obtained
Baseline	Hindi	0.2821
DCU	Hindi	0.2963
ISM	Hindi	0.2793
IIT-B	Hindi	NA

MET Results

Marathi:

Team	Language	MAP Obtained
Baseline	Marathi	0.2320
ISM	Marathi	0.2797
IIT-B	Marathi	0.2684

Odia:

Team	Language	MAP Obtained
Baseline	Odia	0.1537
IIIT-Bh	Odia	0.1537
ISM	Odia	0.1537

Final Remarks

- Tamil systems were not evaluated because Qrels are not available
- Good response in the form of participation
- Some promising results
- Future Plan:
- Develop Gold standard data for each language.
- Choose 30,000 surface words from this gold standard data and evaluate systems for those words.

Acknowledgements

- Prof. Prasenjit Majumder (DA-IICT)
- Harsha Kokel and IRLab@DA-IICT
- Prof. Mandar Mitra (ISI, Kolkata)
- Somnath Chandra (DIT, Govt of India)
- All the organizers of FIRE-2012
- All the esteem research personnel who have contributed to the FIRE data
- All the participants of MET
- The developers of Terrier (Univ of Glassgow)

Thank you!