

*Exclusive Pad*

**MARK SUN<sup>TM</sup>**



A Product by : VERINDA ENTERPRISES  
An ISO 9001 : 2008 CERTIFIED CO.

## UgP2

## Session 2

### DATASET

UST: Dictionary start with binary

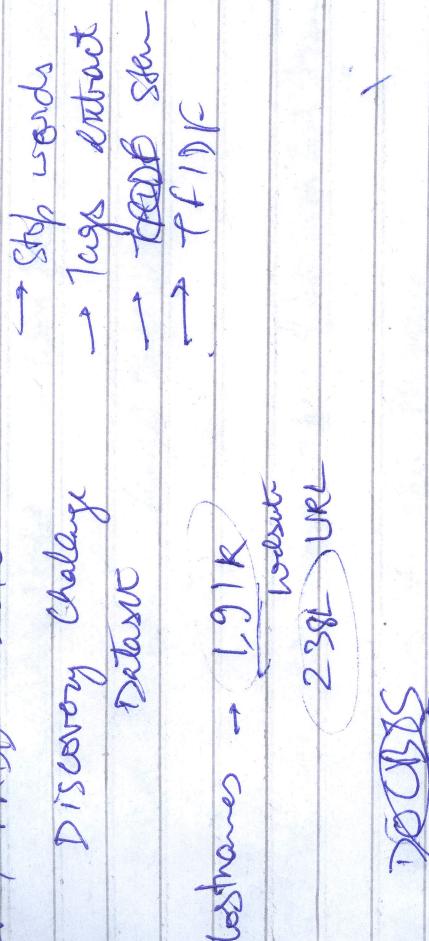
Bertl' PKDD 2010

Convert to binary & TFIDF (one-time effort)

- > Rest (see total labels)
- > clustering (single level link)
  - (link from one tree to another)
  - ↓
  - Bertl' Bertl'

word by looking at top n freq. words in doc

Done



DECISION  
+ keyword

multi-level dataset not found (Google)

ta as keywords they gave

- Later on
- |                             |                  |
|-----------------------------|------------------|
| Total labels (full set)     | 1000             |
| -> 1 vs Rest SVM classifier | - link depending |
| - Random 50:50 lab          | - Independent    |
| many : testing              |                  |
| - Random 50:50 database     |                  |
| - Keyword overlap           |                  |
- Use Dictionary so that standardized

focus: Dataset

Dictionary

SVM (1 vs Rest)

↓

Clustering

↓

NAEPR

## VLSP2

### DATASET

MUST: Dictionary

start with binary

Berly PKDD 2010

DONE

- 1 → Convert to binary & Tfidf (one tree of sort)
- 2 → 1 vs Rest (see total labels)
- 3 → Clustering (Single Level like) (kick from one tree to another)

Same Performance

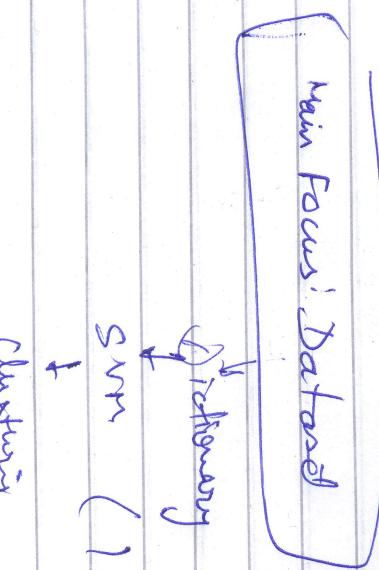
238k URLs

- 2 → Keyword by looking at top n freq. words in doc

TOP 2000  
Keywords

- 3 → keep note on responses they gave

multi level dataset not found (can



- Total levels (full set)
- 1 vs Rest SVM Classify
- Random 50:50 data many testing
- 10th topic database keyword overlap
- Use Dictionary so that standardization

KNN