Project
Analyzing trends of the hour

Specifications

Naeem A. Mahoto

Specifications

- Crawls newspapers (English) national and international

- Stores in textual files the crawled data

- Maintains database based on:
  - Newspaper timestamp (Date)
  - City
  - Country
  - Newspaper Name
  - Journalist Name (only for his/her articles)
  - Keywords (Most frequent ones i.e. Cricket, Football etc)
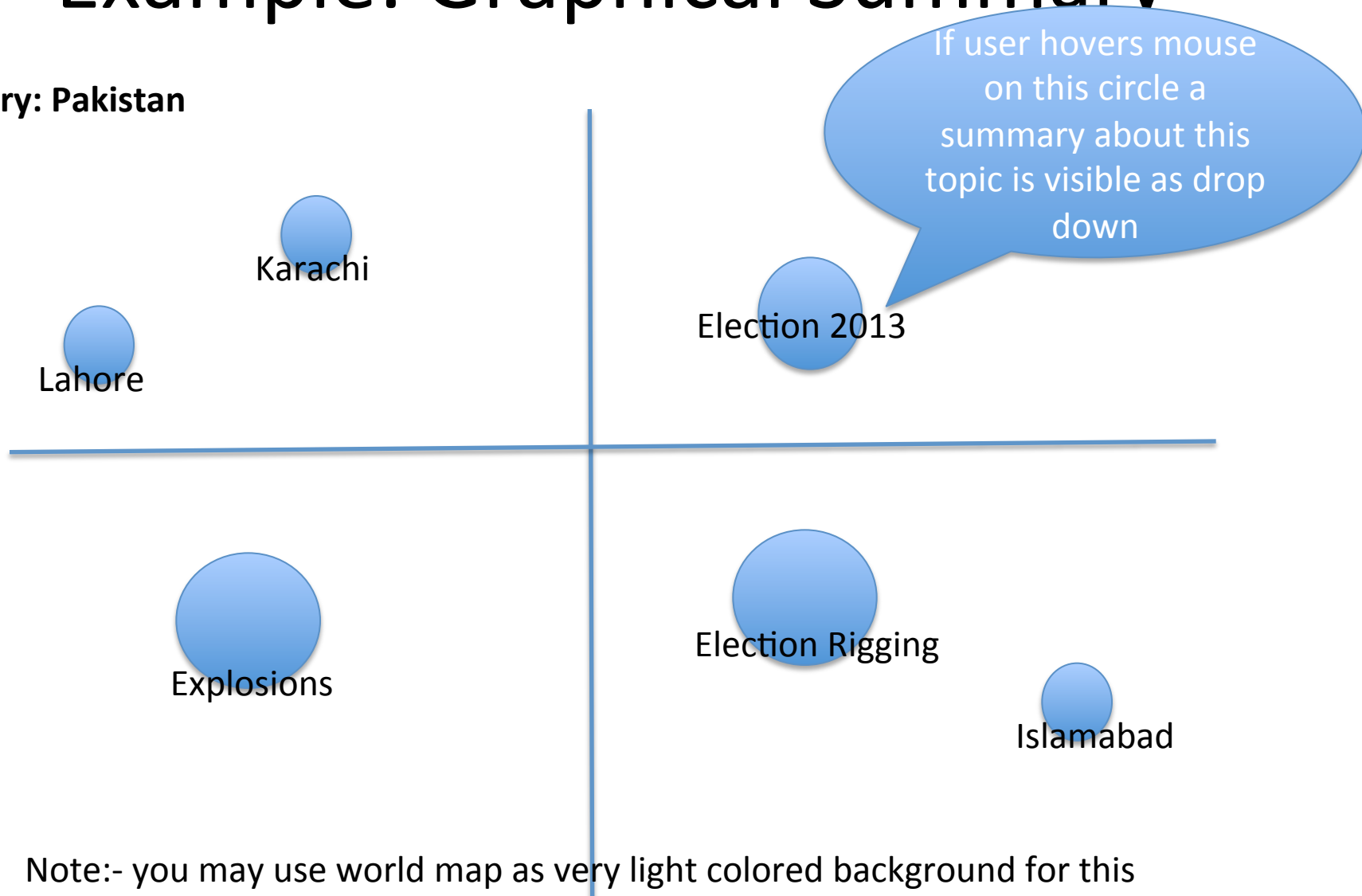
Specifications

- The application has user friendly user-interface

- Provides summary reports for printing and graphical reports

- The analysis takes place using data mining algorithms

- The data mining results are saved into databases for future info

Working Principle

- The crawled data is stored into databases as well as raw format on local disk

- Based on user specified query the data is extracted from databases

- If needed data mining techniques are used for prediction and other novel results

- The data mining results are also saved into separate database tables

- The results are shown in graphical format

Example: Graphical Summary

Graphical Summary results

- Similarly for other view-points like:
  - Topic
  - Newspaper name
  - Date
  - Person etc

  The results are displayed the same results can be printed in descriptive format (Table format)

What Next…???

- This is the beginning….

- you can add and improve the interface and results format… how to visualize in more sophisticated way….

- I have other plans also for this project… the application will also make summaries (1/3rd) of articles… for example: An article contains 1000 words… the application will provide main points in just 100 – 200 words as summary or abstract

What Next...???

- The articles are already saved in raw (unstructured) format on local disk...

- The summary making needs sophisticated algorithms to make summaries from articles...

- The application is just a base for more functionalities that a user or head of any organization may need...

Note:- Don't share this idea with others who may steal it... I plan to sale it to some reputed institution....