

TimeMiner Proposal

Alok Hota, Mohammad Ahmadzadeh
<https://github.com/fdac/TimeMiner>

October 29, 2014

1 Introduction

The idea behind this project is to find world events by analyzing clusters of Wikipedia articles as they change over time. The results of this project will be visualizations that depict the occurrence and mutation of a world event.

2 Article clustering

Clustering articles is done by defining a distance measure between articles. The distance between articles may be affected by two types of parameters:

- *Static strength*: These are parameters that are relatively stable over time, such as link backs, number of common editors, number of common references and syntactic correlation between articles. These parameters will be used as a connection weight between articles for clustering.
- *Dynamic strength*: These are parameters that change rapidly over time, such as the number of edits and the number of page views. Measuring the temporal similarity of these parameters helps visualize the occurrence and importance of events.

3 Visualizations

- A timeline of historically notable world events.
- A graph cluster that shows the connections between articles over time.