



# Faculty of Computer Science CS6999 Reading Course Presentation

Web data and Text Mining Course

By

Amin Jashki, Hamid Baghi and Reza Zafarani

*Wednesday October 24<sup>th</sup>, 2007*

*3:35pm*

*ITC317*

The emerging of Blog technology and Blogosphere has created large sources of news which are highly dependent on socio-political daily topics. Analyzing these sources has been an appealing topic in the area of text mining and various online systems are dedicated to this purpose (Google Trends and BlogPulse).

Blogizer, is a web based system designed in order to track different topics discussed in blogs. These topics are evaluated and analyzed based on some important factors such as quality and frequency of textual data. Blogizer combines these values in a utility function and incorporates different methodologies in the area of text mining such as stemming, stop word removal, pruning, clustering and well known techniques such as PageRank, to name a few.

\*\*\*\*\*  
STUDENTS ARE ENCOURAGED TO ATTEND  
\*\*\*\*\*