

## Cleaning various noise patterns in Web pages for Web data extraction

Thanda Htwe  
University of Computer Studies, Yangon  
tdhtwe80@gmail.com

**Abstract:** Cleaning Web pages before mining becomes critical for improving performance of information retrieval and information extraction. With the exponentially growing amount of information available on the Internet, an effective technique for users to discern the useful information from the unnecessary information is urgently required. So, we investigate to remove various noisy data patterns in Web pages instead of extracting relevant content from Web pages to get main content information. In this paper, we propose an approach NoiseEliminator that detect multiple noise patterns and remove these noise patterns from Web pages of any Web sites. Our approach is based on the basic idea of Case-Based Reasoning (CBR) to find noise pattern from mixture (data and noise together) patterns in current Web page by matching similar noise pattern kept in Case-Based. We also apply back propagation neural network algorithm to classify various noise patterns, data patterns and mixture patterns in current Web page. The classification result of neural network is used for removing noise patterns. We have implemented our method on several commercial Web sites and News Web sites to evaluate the performance and improvement of our approach. Experimental results show the effectiveness of the approach.

## Nonlinear Pseudorandom Sequence

Thin Thin Yu  
University of Computer Studies, Yangon, Myanmar  
thinthyu.ucsm@gmail.com

**Abstract:** Pseudorandom numbers are crucial ingredients for a whole range of computer usages. In this paper we propose the construction of nonlinear pseudorandom sequence from the group of points over an elliptic curve. This method is based on an elliptic curve (EC), a linear feedback shift register and a random box. One technique for destroying the linearity inherent in linear feedback shift register (LFSR) is to generate the key stream as some nonlinear function of the stages of a single LFSR. Random Box is used as nonlinear substitution. This generator generates key stream of encryption algorithm for cryptographic purposes.