Abstract

World Wide Web makes it a prominent source of online information as abundance of data is available on the web and lots of data gets uploaded on daily basis. Due to the presence of massive information on the web it seems easier and simpler to get any information at any time effortlessly, but it requires a lot of focus. Numerous web mining techniques have been studied like extractors, wrappers etc, that provide various methods to extract useful web content. In this paper a semi-supervised web news extraction technique is proposed that uses unsupervised clustering technique and supervised classification technique.

References

1. Zhong Ji, Member, Yanwei Pang, Senior Member, and Xuelong Li, “Relevance Preserving Projection and Ranking for Web Image Search Reranking”, VOL. 24, NO. 11, NOVEMBER 2015.
2. Debina Laishram and Merin Sebastian, “Extraction of web news from web pages using a


**Index Terms**

Computer Science  
Information Sciences

**Keywords**

Web Mining, Web News, Web News Extraction, Unsupervised Machine Learning, Classification