CHAPTER 10

CONCLUSION AND FUTURE DIRECTIONS

This chapter provides the overall conclusion of the whole thesis. It helps to understand work carried out in previous chapters at a glance. Future directions have been presented in this chapter to extend the presented work for providing more fruitful results.

This chapter is organized into 2 sections, Section 10.1 provides the conclusion of the whole thesis followed by the future directions in Section 10.2.

10.1 CONCLUSION

- The research work presented in this thesis provides different approaches for improving the efficiency of Web data extraction. The presented approaches can be applied to meet the challenges and issues of Web data extraction like large scale of data available over the Websites, nature of data (semi structured, structured or unstructured) over the Websites, noisy data in the Websites etc. The presented models are applied over a number of Websites developed and provide efficient results in terms of parameters used for Web data extraction.
- The research work is applied over the real world scenario i.e. over the self-developed Websites.
- MCMM-LSW (Multi Level Content Mining Model for Large Scale Websites) provides extraction of data from large scale Websites and hence gives a solution for the issue of rate of growth of information over the period of time on websites. It provides solution for content matching over large scale Websites instead of traditional approach i.e. keyword matching based Web data extraction.
- UMC (Usage mining using Clustering approach) use partition based clustering approach for Web Usage Mining which extracts the data according to user choice.
- n x 1 model is used to extract noise free data in efficient manner by reducing the extraction time. It provides a template based approach for Web wrapper generation.
and also show the use of XML in better way. It provides better results in terms of web data extraction parameters than the existing image method.

- RWW (Relational Web wrapper) provides an example of developing program for Web wrapper and helps other developer to know the way web wrappers are developed, works and tested. Results shows that a web wrapper generally takes more time when it runs for the first time than the subsequent rounds.

- Performance comparison of Web data extraction techniques provides comparative study of different techniques used for efficient Web data extraction. The results help the users to know which technique to use in a particular scenario?

- Finally the combination of Web content mining and Web usage mining (CCMUM) provides more fruitful results for improving the efficiency of Web data extraction. Web usage mining provides results in terms of user choice and Web content mining help to extract the data of the WebPages or Websites. The proposed approach combines the already developed approaches i.e. MCMM-LSW, UMC and n x 1 model.

10.2 FUTURE DIRECTIONS

- Presented approaches can be applied over to other real world datasets or Websites available over the Internet with more enhancements to have better impact on Web data extraction system for overall Web.

- The approaches may be enhanced to include and improve ranking of the extracted data. Hence some approaches for Web Structure Mining may be proposed and developed.

- Security is a major concern in Web data extraction hence, security approaches may be applied over the developed approaches to provide secure and efficient web data extraction.

- Web data extraction techniques can be developed for Semantic Web scenario to extract data from the ontologies. By extracting data in machine understandable form may provide more fruitful results.
• Today the field of Big data is also moving towards Web scenario and hence some approaches may be developed for the extracting the data from such situations. More models can be proposed for the characteristics of Big data like volume, velocity and versatility.