With the start of the Internet, Information Retrieval (IR) became increasingly relevant and evident. The main aim of an IR system is to retrieve the documents efficiently and effectively relevant to a user query. Apparently, a good MLIR system should satisfy the information needs of the user. The performance evaluation of MLIR systems is a tiresome task and it is found that quality evaluation can be done effectively with the use of the metrics. Hence it is made as a standardized practice in research community of the MLIR domain to concentrate mainly on measures concerning the quality of the retrieval output.

On the other hand, metric-based evaluation schemes are very important for evaluation of MLIR systems because relevance assessment is purely subjective to the user. The performance evaluation of any MLIR system is carried out by assessing the documents retrieved in response to a query with respect to their relevance score and computing the appropriate set-based measures. Apart from the important evaluation initiatives in the IR domain, the MLIR is also becoming popular to envisage the effective evaluation measurement schemes, but with the use of same set of measures as it is proposed for evaluating IR domains only. Though, few evaluation initiatives are developed for multilingual scenarios, they have not used all the performance evaluation metrics which are available to evaluate any of these MLIR systems. This dimension motivates this thesis to identify different schemes of IR measures thereby, enhancing them to measure the performance of the MLIR systems.

This work reported in this thesis aimed at to offer a set of application-independent fine grained quantification schemes, which will act as a generic
model for evaluating the performance attributes of MLIR systems. In this respect, the goals of this research are derived in order to extend the support to the developers to evaluate the MLIR systems as follows:

I. Reckon the performance measures of MLIR systems to get the fine grained interpretations
II. Qualitative analysis is accomplished to epitomize the different characteristics of MLIR systems
III. Guide the designer to design and develop better MLIR systems at different approaches.

The above specified goals are much attentive and more quantifiable. That is, the language-specific granularities at several schemes of MLIR systems with respect to the performance evaluation is delineated by the First goal and consequently the various characteristics of MLIR systems are delivered in the Second goal. The upshots of these two goals yield the Third goal i.e. the implementer to implement better MLIR systems; this exhibits the prominence of the proposed research.

The work reported in this thesis describes the greater efforts that are taken to develop a set of Metrics under four different schemes for Evaluating the Performance of MLIR Systems: *Ranked relevance and Ranked retrieval; Binary relevance and Ranked retrieval; Binary relevance and Continuous retrieval and Binary relevance and Binary Retrieval*. In case of Ranked Relevance and Ranked Retrieval scheme, the distance properties are covered. As per the retrieval systems concerns, relevance is subjective in nature and system relevance is different from user relevance. In case of Binary Relevance and Ranked retrieval scheme, usability issues are discussed and the corresponding properties are identified. Usability is an important
characteristic of every MLIR system. There are different approaches to measure the usability. In case of Binary Relevance and Binary Retrieval scheme, internal characteristics are studied. Internal characteristics deal with qualitative attributes that should be generalized to every MLIR system. Finally, in case of Binary relevance and Binary retrieval scheme, design issues are taken into account; design issues facilitate some peculiar clues to the designer to design qualitative model in the near future. These proposed measurement schemes can play the decisive roles for developing qualitative designs to maintain the optimized performance in any MLIR system. Experiments of the proposed research are chosen such that, to assess the proposed metrics both empirically and statistically in order to validate their usefulness.

It is also outlined that further research work can be carried out such that to consider other available important traditional IR Metrics. In addition to the schemes and metrics, this work can also be extended in terms of delineating new measures for specific issues in the MLIR systems.