Abstract - The objective in extreme multi-label learning is to learn a classifier that can automatically tag a data point with the most relevant subset of labels from an extremely large label set. Extreme multi-label classification is an important research problem as it not only lets us tackle web-scale classification problems but it has also opened a new paradigm for solving ranking and recommendation problems. This thesis focusses on developing scalable algorithms and appropriate loss functions for extreme classification problems which can lead to state-of-the-art performance on large-scale ranking and recommendation applications. In particular, it makes the following contributions –

1) It proposes loss functions suitable for extreme multi-label learning that do not erroneously treat missing labels as irrelevant but instead provide unbiased estimates of the true loss function even when ground truth labels go missing under arbitrary probabilistic label noise models. These loss functions also naturally promote the accurate prediction of infrequently occurring, hard to predict, but rewarding tail labels.

2) It develops the SLEEC algorithm which is an embedding based extreme multi-label learning approach. SLEEC addresses some of the major limitations of embedding based methods such as high training and prediction cost and low prediction accuracy.

3) It develops the Slice algorithm for extreme multi-label learning with low-dimensional dense features that scales to 100 million labels and 240 million training points.

4) It reformulates the problem of recommending related queries on a search engine as an extreme classification task and demonstrates that Slice could significantly improve recommending related searches on Bing.