



# IMPROVING EFFICIENCY AND ACCURACY IN STRING TRANSFORMATION ON LARGE DATA SETS

Jeyalakshmi.S<sup>1</sup>, Rathika.T<sup>2</sup>

<sup>1</sup>M.Tech final year, Dept. of CSE, SRM University, Ramapuram, jeyalakshmibetch@gmail.com

<sup>2</sup>A.P (O.G), Dept. of CSE, SRM University, Ramapuram, rathika.t@rmp.srmuniv.ac.in

---

## Abstract

This paper discusses the problems in information processing on data mining, information retrieval, and bioinformatics can be put forwarded to string transformation. The  $k$  most likely output strings are generated corresponding to the given input string for string transformation. It proposes a probabilistic approach such as log linear model-a training method and algorithm for generating top  $k$  candidates to string transformation. The log linear model is defined as a conditional probability distribution of an output string and a rule set for the transformation conditioned on an input string. The maximum likelihood parameter estimation is employed for learning method. The optimal top  $k$  candidates are generated using this string generation algorithm and commentz walter algorithm. Correction of spelling errors in queries as well as reformulation of queries in web search is made using our proposed method. Experimental results on large scale data show that the proposed approach is very accurate and efficient improving upon existing methods in terms of accuracy and efficiency in different settings.

**Keywords:** Log Linear Model, Parameter Estimation, Query Reformulation, Spelling Error Correction, String Transformation, commentz walter algorithm

---

Full Text: [www.ijcsma.com/publications/march2014/V2I314.pdf](http://www.ijcsma.com/publications/march2014/V2I314.pdf)