

Tri-gram model for language identification and comparison

Jean Lucien Randrianantenaina
Applied Mathematics
Machine Learning and Artificial Intelligence
Stellenbosch University
161 Decanting Building, Hammanshand Road.
Email: lucien.randrianantenaina@aims-cameroon.org

Abstract—Abstract

I. INTRODUCTION

II. METHODOLOGY

A. Text normalization

B. Language modelling

C. Language identification

D. Byte-pair encoding for language similarity

E. Perplexity

III. RESULTS AND DISCUSSION

IV. CONCLUSION

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.