International Journal on Recent Researches In Science, Engineering & Technology

A Journal Established in early 2000 and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy.
It is an absolutely free (No processing charge No publishing charge etc) Journal Indexed in DIIF and SJIF.

Research Paper

Available online at: www.jrrset.com

Chief Editor : 1. Dr. M.Narayana Rao, Rtd. Professor, NIT, Trichy. (Engg.&Technology division)
2. Dr. N.Sandyarani, Ph.D., Professor, Chennai based Engg.College, (Science division)

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

Volume 1, Issue 12, Dec. 2013

DIIF IF :1.46
SJIF IF: 1.329

Developing Corpora for Statistical Graphical Language Models

Andrew Sullivan

Abstract - In this work Statistical Graphical Language Models (SGLMS), a technique adapted from Statistical Language Models (SLMs), are applied to the task of graphical object recognition. SLMs are used in Natural Language Processing for task such as Speech Recognition and Information Retrieval . SGLMs view graphical objects as belonging to graphical languages and use this view to compute probabilistic distributions of graphical objects within graphical documents . SGLMs such as N-grams require large Corpora of training data, which consists of graphical objects in contextual use (real world graphical documents). Constructing Corpora is an important stage in developing the models and many issues need to be addressed . This paper discusses the development of graphical corpora and presents approaches to some of the problems encountered .