

**SINSENSE - WORD SENSE  
DISAMBIGUATION TOOL FOR  
SINHALA LANGUAGE**

**Timeth Subasinghe**

A dissertation submitted in partial fulfilment of the requirement for Bachelor of  
Engineering (Hons) degree in Software Engineering

**Department of Computing**

**Informatics Institute of Technology, Sri Lanka**

**in collaboration with University of Westminster, UK**

**2020**

# Abstract

Semantic ambiguity in human language has been a challenge for Natural Language Processing from the inception of it. Solving this would give an advantage to many NLP tasks such as Machine translation, Information retrieval, Question answering, speech recognition. Few years around, it was recognized as an AI-Complete problem which means it is as hard as central problems or Artificial intelligence.

Sinhala is the mother tongue in Sri Lanka with around 16 million people using the language. There have been many researches carried out for Sinhala NLP in areas such as Text to speech, speech text, and so on. There have been few researches done on Sinhala word sense disambiguation as well, but lack of resources has made it hard to improve further.

Sin-Sense tries to solve Sinhala word sense disambiguation Using a non-conventional approach, cross-lingual sense disambiguation where another language is used to aid the sense disambiguation process in the target language. This approach is possible because English word sense disambiguation has been greatly improved and there are services that people can use for sense disambiguation.

This system is first of its kind for the Sinhala Language with better accuracy than existing systems. Sin-Sense is available as a web application for everyone.

**Keywords:** Natural Language Processing, Word Sense Disambiguation, Cross Lingual WSD

Sinhala WSD