Applying Finite-State Methods to the Swahili Language
Michael Wangia, Isaiah Greene
Advisor: Aravind Joshi

Abstract
The objective is to deliver a recognition device for identifying well-formed Swahili sentences as well as for generating new Swahili strings that have the noun phrases and prepositional phrases enclosed in parenthesis.

High-Level architecture

Why Swahili?
Finite State Methods have already been applied to the English Language in areas such as Text Processing.

Since the Swahili sentence structure is different than the English one, we decided to explore the possibilities of applying finite state methods to it.

In addition, technology companies like Google and Microsoft are creating Swahili products. Thus, we anticipate that Swahili Finite State Methods will be useful software tools in the future.

System Model

Finite State Machines (FSM):

Finite State Transducers (FST):

Simple Noun Phrases FSM

Simple Prepositional Phrase FSM

An Example
In this example, we combine the Noun Phrase FSM and Noun Phrase FST to create a device that outputs parenthesized Noun Phrases in a grammatical Swahili sentence.

Given the grammatical Swahili sentence below:
Watoto werevu sana wanacheza kwenye shule
(N) (ADJ) (ADV) (Ving) (PREP) (N)

Start state (A) (B) (C) (D) (E) (F) (G) (H)

Prepositional Phrase FST
Noun Phrase FST

Empty string

Final correct output: [watoto werevu sana] kwenye [nyumba]