

# Group Charlie

## Multi-Lingual SMS

Andrew, Chihang, Daria, Swaraj and Tanvi

University of Cambridge

March 4, 2015

# Background

- ▶ Africa's Voices: Opinion polling done by broadcasting a variety of questions from several radio stations, where answers were received as SMS messages.

# Background

- ▶ Africa's Voices: Opinion polling done by broadcasting a variety of questions from several radio stations, where answers were received as SMS messages.
- ▶ Main problem: The SMS messages were in different languages and sometimes included irrelevant information, such as details of the sender.

# Background

- ▶ Africa's Voices: Opinion polling done by broadcasting a variety of questions from several radio stations, where answers were received as SMS messages.
- ▶ Main problem: The SMS messages were in different languages and sometimes included irrelevant information, such as details of the sender.
- ▶ Sub-problem: Machine translation is not feasible for minority languages with little economic value. As a result of this, translation was taken to be an expensive task.

# Background

- ▶ Africa's Voices: Opinion polling done by broadcasting a variety of questions from several radio stations, where answers were received as SMS messages.
- ▶ Main problem: The SMS messages were in different languages and sometimes included irrelevant information, such as details of the sender.
- ▶ Sub-problem: Machine translation is not feasible for minority languages with little economic value. As a result of this, translation was taken to be an expensive task.
- ▶ Problem we're trying to solve: allow for easy detection of opinion trends in such data, despite it being in several languages.

# Our goals

- ▶ This is a tool for scientists — it should have a friendly and interactive UI while not getting in the way.

# Our goals

- ▶ This is a tool for scientists — it should have a friendly and interactive UI while not getting in the way.
- ▶ Due to translation being expensive, we need to allow for the researchers to determine more important words/topics for which they can prioritise translation.

# Our goals

- ▶ This is a tool for scientists — it should have a friendly and interactive UI while not getting in the way.
- ▶ Due to translation being expensive, we need to allow for the researchers to determine more important words/topics for which they can prioritise translation.

Additional requirements set along the way:



# Our goals

- ▶ This is a tool for scientists — it should have a friendly and interactive UI while not getting in the way.
- ▶ Due to translation being expensive, we need to allow for the researchers to determine more important words/topics for which they can prioritise translation.

Additional requirements set along the way:

- ▶ Allow for identification of the language that a message was written in.

# Our goals

- ▶ This is a tool for scientists — it should have a friendly and interactive UI while not getting in the way.
- ▶ Due to translation being expensive, we need to allow for the researchers to determine more important words/topics for which they can prioritise translation.

Additional requirements set along the way:

- ▶ Allow for identification of the language that a message was written in.
- ▶ Allow for 'cleaning up' of the data by removing names, txt-speak and other slang.

# Features/what we've accomplished

# Nitty-gritty details of our code

# Lessons learned