



Machine Learning Research in Odia : Where We Stand

— Dr. Shantipriya Parida —
Idiap Research Institute



Current Scenario

Below are some of the major challenges in Odia machine learning research:

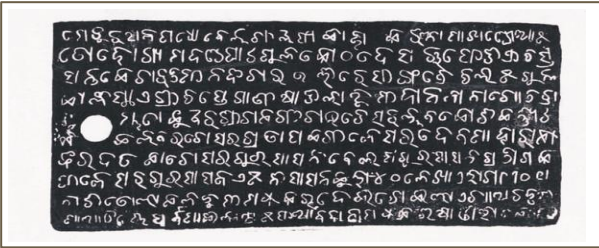
- Low content digitization.
- Limited availability of NLP resources for research.
- Limited/unavailability of better Machine Translation System.
- Low/unavailability of services in Odia language.

Ongoing Research and Contribution

- OdiaWiki is getting richer day by day
 - Thanks to volunteers for their dedication
- Odia NLP Resource Catalog
 - <https://github.com/shantipriyap/Odia-NLP-Resource-Catalog>
- Odia Corpus for Machine Translation
 - <https://ufal.mff.cuni.cz/odiencorp>
 - Odia↔English already added in Workshop on Asian Translation 2020 challenge
- Machine Translation system (open-source)
 - Will be available soon...
- OCR system
- App development

Going Forward (Possible Research Direction)

Digitized ancient contents/
applying NLP techniques.



Source: <https://www.quora.com/How-did-Odia-get-its-script>

Build NLP resource/tools/app
for Odia dialect

ଲିଖିତ ଓଡ଼ିଆ : ମୋର ଆଦି ବହୁତ କାମ ଅଛି
ଦେଖିଆ: ମର କେତେ କାମ୍
ଗଞ୍ଜନାମୀ: ଆଦି କାମ ଟିକେ ଅଧିକା କା
ସମ୍ବଲପୁରୀ: ମୋର ଆଦି ଟିକେ ବେଶି କାମ ଅଛି
ବାଲେଶଠରୀ: ମର ଆଦି ଭେର କାମ ଅଛି

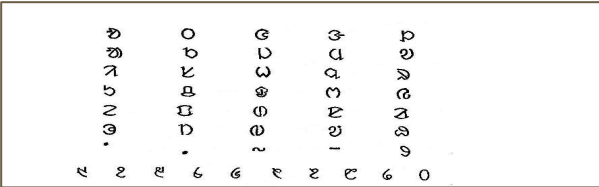
Build NLP resource for Odia
Fakenews

Angul youth arrested for fake coronavirus
news on social media

Fake news hits Odisha VIPs

<https://ommcomnews.com/odisha-news/angul-youth-arrested-for-sreading-fake-news-on-coronavirus>
<https://www.telegraphindia.com/india/fake-news-hits-odisha-vips/ci/d/1789442>

Build NLP resource/tools for
the Tribal languages



Source: [https://commons.wikimedia.org/wiki/File:Ol_chiki_script_\(writing_of_santa_li_language\)_2014-05-30_15-21.jpg](https://commons.wikimedia.org/wiki/File:Ol_chiki_script_(writing_of_santa_li_language)_2014-05-30_15-21.jpg)

Shared task/Promotion in event
/international conference

WAT 2020

The 7th Workshop on Asian Translation

- Indic tasks:
 - UFAL (EnOdia) corpus: Odia <--> English (NEW!)
 - Bengali/Hindi/Malayalam/Tamil/Telugu/Marathi/Gujarati <--> English

Product localization/service
accessible in Odia language

Microsoft Translator Blog

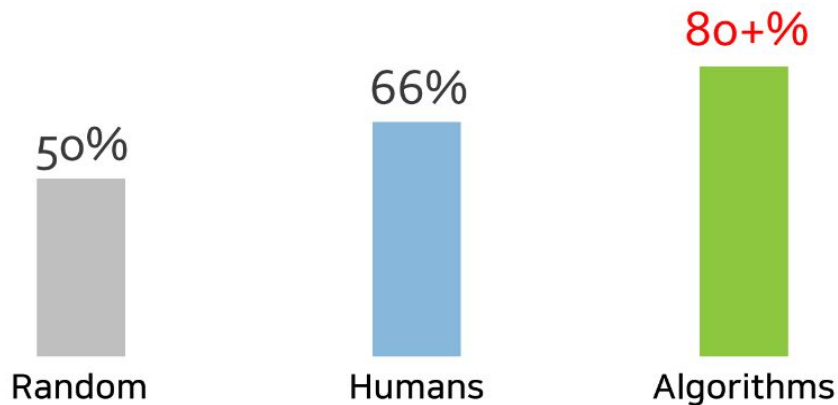
Odia Language Text Translation Is Now Available in Microsoft Translator

Source: <https://www.microsoft.com/en-us/translator/blog/2020/08/13/odia-language-text-translation-is-now-available-in-microsoft-translator/>

Ethics, Bias and Rights in AI/Machine Learning

- Developing AI leads to new questions, including moral and ethical ones.
- How we can set boundaries and set regulations.

Who can Detect Fake News Better ? Human vs. Algorithm



Kumar-West-Leskovec, WWW 2016

Can AI generate Original Expressions ?

A novel written by Japanese AI program passed the first round of screening for a national literary prize

A novel called <The Day A Computer Writes A Novel> was created by the AI program developed by researchers at the Future University Hakodate, Japan.



2016 <https://www.digitaltrends.com/cool-tech/japanese-ai-writes-novel-passes-first-round-national-literary-prize>

GPT-3 CREATIVE FICTION

Creative writing by OpenAI's GPT-3 model, demonstrating poetry, dialogue, puns, literary parodies, and storytelling. Plus advice on effective GPT-3 prompt programming & avoiding common errors.

<https://www.gwern.net/GPT-3>

Take-Away Message

- To avoid bias in machine learning:
 - Select appropriate learning model.
 - Verification of data being used to train the model.
 - Monitor performance (tool, real data)