

What exactly is NLP (Natural language processing), and how is Anusaaraka helping?

Jasveen S, Anusaaraka lab,
LTRC, IIIT-Hyderabad (Professor
Vineet Chaitanya)

オープンソースカンファレンス2020

Online/Kyoto セミナープログラム

Computers are incredibly fast, accurate and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination. (*Albert Einstein*)

アルバートアインシュタインが
指摘したように、コンピュー
ターは基本的に完全に愚かで
す。

TALKS & ME (私について?)



**Open Source
Conference
2020 Online
/北海道**



**PyDelhi
2020**

**PyCon アフリカ
2020**

01

ANUSAARAKA

Anusaarakaは、可能な限り、ソース言語の構文をターゲット言語の対応する構文にマップします。

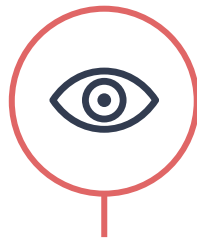
Some standard components (MTシステム の)



**Word
analyzer**



**ローカル単語グ
ルーパー**



**Mapper using
bilingual
dictionaries**



**ワードシンセサ
イザー**



Proposed system (およびデザイン)

CONDENSING NLP



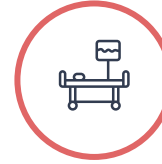
Text Preprocessing

ノイズ除去
レキシコン正規化
オブジェクトの標準化



Syntactical Parsing

構文解析



Entity Parsing

フレーズ検出
名前付きエンティティの認識
トピックモデリング
N-グラム



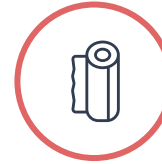
Statistical features

TF - IDF
周波数/密度機能



Word Embeddings

Saturn has several rings



Text Matching

レーベンシュタイン距離
音声マッチング
柔軟な文字列マッチング



INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY

HYDERABAD

Important Libraries for NLP (python)

- Scikit-learn: Pythonでの機械学習
- Natural Language Toolkit (NLTK): すべてのNLPテクニックのための完全なツールキット。
- Pattern – NLPおよび機械学習用のツールを備えたのWebマイニングモジュール

THANKS!

何か質問がありますか？

userjs76@gmail.com

<https://africa.pycon.org/speakers/jasveen-sandra>