

# 日本語入力の危機を乗り越える

インプットメソッド・フレームワークと  
かな漢字変換に訪れている課題とその対策

Open Developers Conference 2019 Tokyo  
2019-08-24

Cross Distribution Developers Camp

# アジェンダ

- Cross Distribution Developers Campとは
  - 解決したい問題
  - 活動内容
- 日本語対応の技術紹介
  - 日本語対応を行う仕組み
  - 日本語入力を行う仕組み
- 日本語入力環境の現状
  - インputメソッド・フレームワークのメンテナンス状況
  - かな漢字変換エンジンのメンテナンス状況
- まとめ

# Cross Distribution Developers Camp

# Cross Distribution Developers Campとは

- Linux/BSD等のOSSディストリビューションで横断して抱える諸問題を解決していこうとする団体
  - **現在の主なモチベーションは「日本語対応」**
- 主なメンバー
  - 日本openSUSEユーザ会メンバー
  - 東海道らぐメンバー
  - Debian JPメンバー
- このセミナーの登壇者
  - 日本openSUSEユーザ会 武山さん
  - 東海道らぐ はしもとまささん
  - Debian JP 吉野さん、杉本

# Cross Distribution Developers Campが解決したい問題

- 問題意識
  - Linux/BSD等のOSは数多くのディストリビューションがある
  - それぞれのディストリビューションが抱える問題は似ている
  - ディストリビューション間の共通課題を共同で解決できれば互いにメリットがある
- 改善したいと考えている問題
  - GNOMEやKDEなどが採用する新しいGUI toolkit (GTK/Qt) への対応
  - Waylandといった新しい方式のデスクトップ環境への対応
  - コンテナ環境への対応 (Flatpakなど)
  - スクリーンキーボード、音声によるキー入力
  - かな漢字変換エンジンの開発停滞と辞書のライセンス (機械学習と関連)

# Cross Distribution Developers Campの活動内容

- 個人がそれぞれのディストリビューションで活動
- 主にslackやMLでコミュニケーションをとり、困った話題を共有
  - <https://floss-os-jp.slack.com/>
  - floss-os-jp@googlegroups.com
- 定期的に「もくもく会」を実施し共同で問題の解決を図る
  - 次回は8月末～9月に開催予定
- 合宿もやっています
  - 2019年6月に開催
  - 2019年内にもう一度開催を計画中



# 日本語対応の 技術紹介

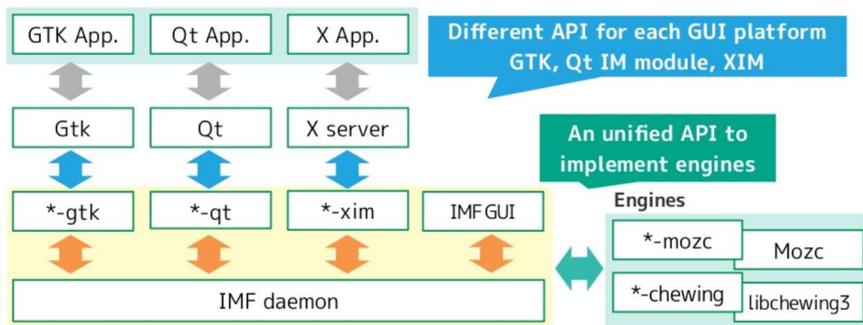
# (技術紹介) 日本語対応を行う仕組み

- 一般的に「国際化(I18N)」「地域化(L10N)」と呼ばれる
- 構成要素
  - フォント
  - 日本語入力
    - インputメソッド・フレームワーク(例:ibus、uim、fcitx)
    - かな漢字変換エンジン(例:mozc、anthy、libkkc、freewnn、skk)
    - 辞書
  - アプリケーションの対応
    - gettext(翻訳した各言語用の文字列リソースを差し込む仕組み)
      - \*-i18n、\*-l10nという名前と言語別に文字列リソースを提供
    - マルチバイト文字の処理への対応
    - 日本語文章の特有表現への対応(ルビ、縦書きなど)

# (技術紹介) 日本語入力を行う仕組み

## What is an input method framework (IMF)?

- An interface between applications and input method engines (conversion software)
  - Also allows switching multiple input method engines
  - e.g., IBus, Fcitx, GCIN, SCIM, UIM, IIIMF, Kinput2, ...



JapaniUG

2018/08/11 4/23

出典:

<https://events.opensuse.org/conferences/summitasia18/program/proposals/2056> P.4

# 日本語入力環境の 現状

# インプットメソッド・フレームワークのメンテナンス状況

- ディストリビューションの選定状況
  - ibus (openSUSEにおける日本語環境のデフォルトに採用)
  - uim (Debianにおける日本語環境のデフォルトに採用)
- 現在のメンテナンスと開発の状況
  - 「What is necessary for the next input method framework」(武山さん)
    - <https://events.opensuse.org/conferences/summitasia18/program/proposals/2056>
  - 「Busterデスクトップの日本語入力」(吉野さん)
    - <https://tokyodebian-team.pages.debian.net/pdf2019/debianmeetingresume201907-presentation-yyoshino.pdf>

# かな漢字変換エンジンのメンテナンス状況

- ディストリビューションの選定状況
  - mozc(openSUSE、Debianにおいて日本語環境のデフォルトに採用)
  - libkkc(Fedoraにおいて日本語環境のデフォルトに採用)
  - anthy(Debianにおいてmozc非対応CPUアーキテクチャで利用可能)
- 現在のメンテナンスと開発の状況
  - mozc(upstream: <https://github.com/google/mozc>) **最終リリース日時: 2016-12-08**
    - 現在はpull requestを受け入れていない状況
  - libkkc(upstream: <https://github.com/ueno/libkkc>) **メンテナンスは継続中**
  - anthy(upstream: <https://ja.osdn.net/projects/anthy/releases/>) **最終リリース日時: 2009-02-09**
    - <https://salsa.debian.org/debian/anthy> へforkしてメンテナンスを継続中
- 新しいかな漢字変換エンジンの開発
  - 「自作かな漢字変換「Genji」をつくったよ」(はしもとまささん)
  - <https://www.slideshare.net/hashimom/genji-97697705>

まとめ

# まとめ

- openSUSEとDebianの実例を交えて日本語入力環境のおかれた状況を整理しました
- ディストリビューション界隈には、日本語環境に必要な知識と技術を持っている人たちがいます
- コンピュータにおける日本語変換処理と日本語を使った機械学習の基礎技術は同じ
  - 日本語処理の理論と実践を学ぶことで機械学習をより深く理解できる
- 日本語環境の発展とより使いやすいOSをつくりたいと考えている人たちが力を合わせて成果を出す場として**Cross Distribution Developers Camp**があります

ディストリビューションの垣根を  
超えた開発にご興味がある方は  
ぜひ一緒に活動しませんか！

活動に参加してみたい方はQRコードの  
slackチームへの参加リクエストをお願いいたします！

