

Nginx で日本語入力を遊んでみよう！

はしもとまさひこ

オープンソースカンファレンス 2015 Hamanako
2015/2/11

おーぶん万葉

自己紹介

- 東海道らぐ(Tokaido Linux Usergroup)とかやってる人。
- 某大学文学部日本文学文化学科卒
 - ケータイメールの日本語研究もしてました
- 去年(?)から、おーぶん万葉を始めてます。
 - 少しでも大学時代に学んだことを生かせればね!

おーぶん万葉

おーぱん万葉プロジェクトとは？

- 目的「自由な日本語入力環境を手に入れよう！」
 - 現状の問題点：
 - ビックデータと叫ばれる時代に、開発がアクティブで自由にコミットメントできる日本語入力システムがない？
 - 例: mozc, Anthy... (但し、SKKを除く!)
- そんな現状を打破したい!てのが目的です。

おーぱん万葉

直近の活動

- Cannaをフォークして、Izumoを開発中!
 - Github: <https://github.com/hashimom/Izumo>
- 変換システムに依存しない辞書を作成したい!
 - 昔は「Pubdic」ってありましたよね?
 - オープンデータを活用した辞書を作りたい!
(例: Wikipediaとか)

おーぶん万葉

本日のトピック

- Cannaは新しい時代のかな漢字変換システム？
 - サーバークライアント型のシステム
- Node.jsと連携させてみよう！
 - Cannaライブラリ呼び出し部分をC++で実装
 - JavaScriptからライブラリを呼んでみよう
- さらにNginxと連携させてみよう！
 - 設定方法と実演

おーぶん万葉

ここでお詫び:

(既に気づいたかも知れませんが)本日のタイトルは、、、

「Nginxで日本語入力を遊んでみよう」
ではなく
「Node.jsで日本語入力を遊んでみよう」

みたいな感じになってしまいましたm(_ _)m

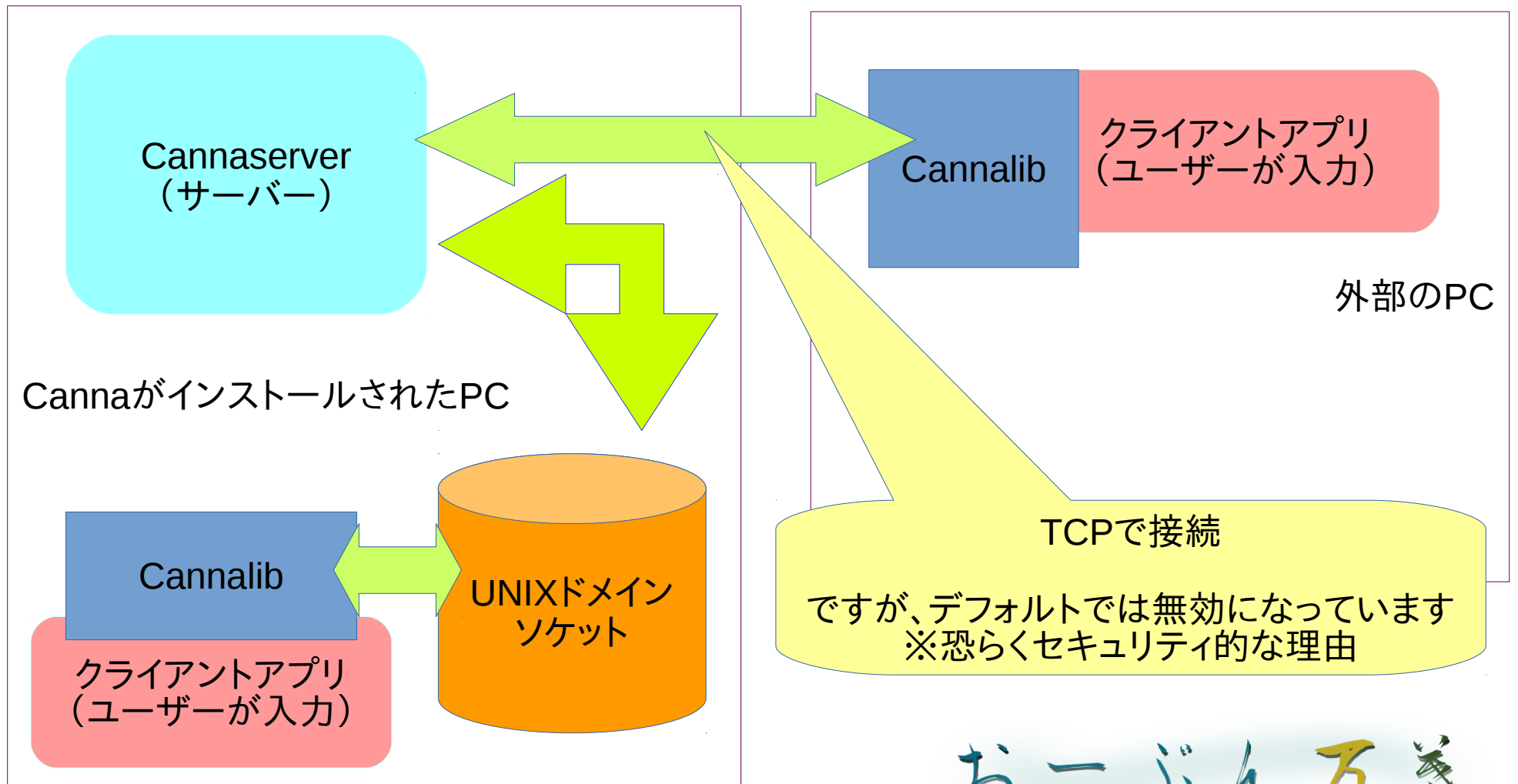
おーぶん万葉

Cannaとは？

- 1989年にNECが開発したかな漢字変換システム
- 現在はオープンソースとなっている
 - MITライセンス → Izumoのフォーク元
- が、残念ながら開発停止状態orz
 - 最後のリリースは2004年。
(MLで呼びかけても返事がない???)

おーぶん万葉

Cannaはサーバークライアント型！



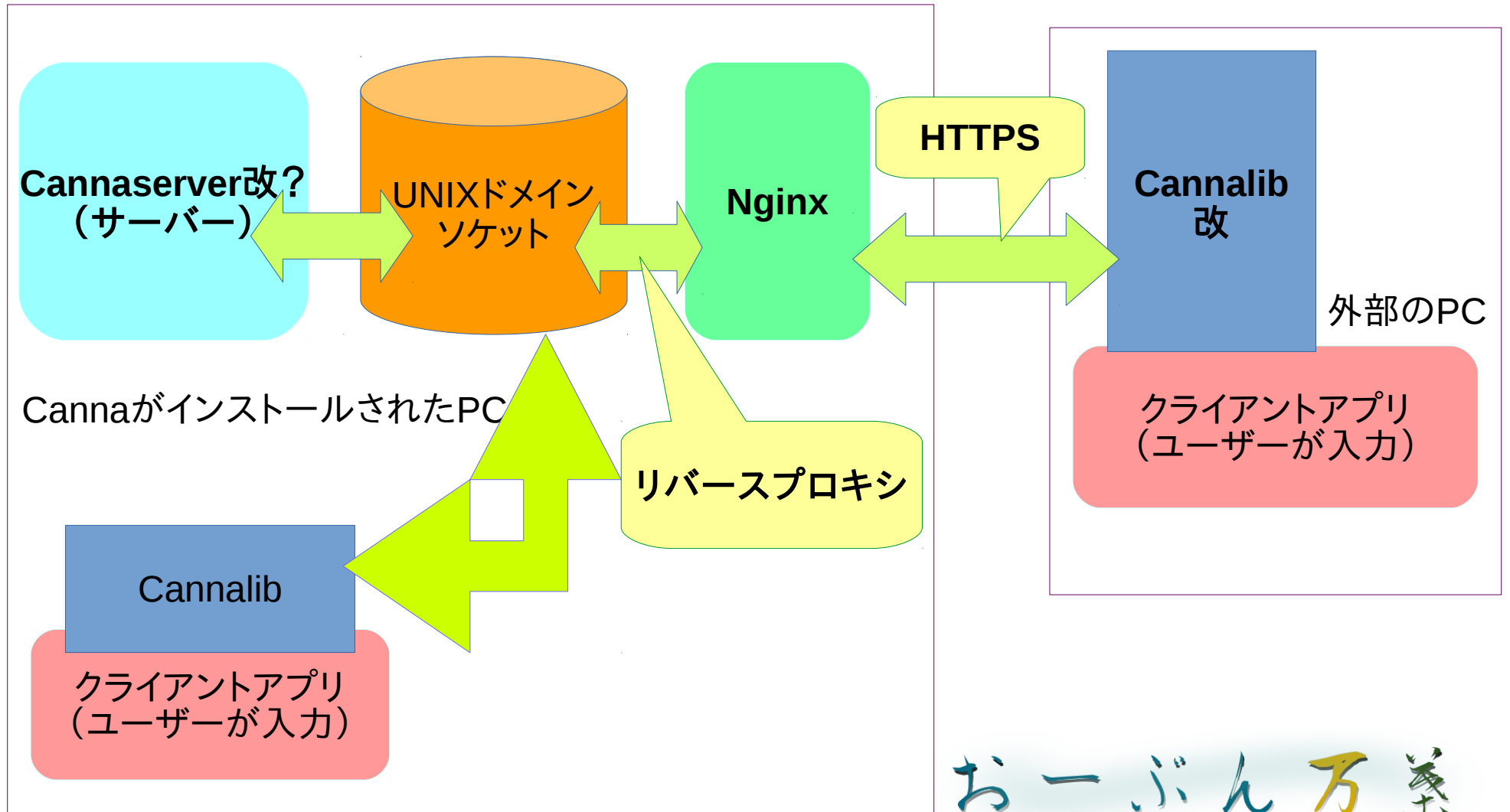
おーぶん万葉

Cannaがサーバーってことは…

- 常にどこでも同じ環境の日本語入力ができる!
 - これって今時の「クラウド」だ!!! (え?)
- がしかし、そもそも「通信暗号化」が考慮されていない
 - 通信が丸見え。→誰が何を入力したかバレバレ
 - 結局使えるのはUNIXドメインソケットのみ
 - ローカルPCでのみとか、サーバーの意味なくね?
- ようは、通信を暗号化すればいいんでしょ?

おーぶん万葉

そしたらこんなのはどうだ!



おーぶん万葉

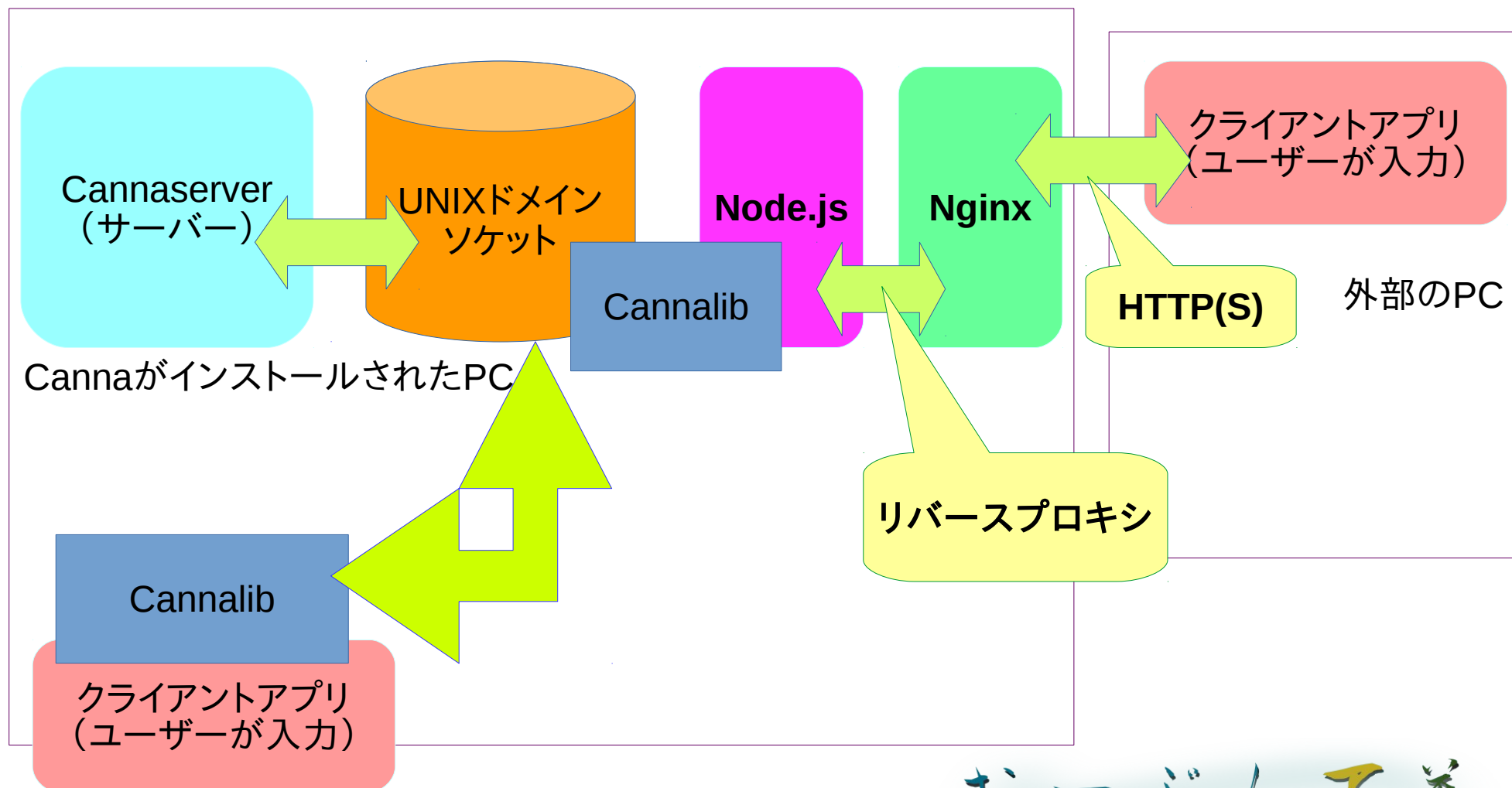
さてどうだろうか？

- おそらく、「可能」だろう。
 - 試してないです。。

却下した理由：

- 環境構築がちょっと複雑
- クライアント・サーバー側にもSSL対応が必要
 - 実装も大変ですね…

そして、(今回は)こうなった!



おーぶん万葉

利点

- 実装する箇所はNode.jsの箇所のみ!
 - 既存のクライアントライブラリもそのまま使える!
 - コード行数も少ない!
- SSL通信に関してはNode.jsに任せられる
 - cannaサーバーには手を入れずに済む

※今回の実験はHTTP(暗号化前)までですが。。。

おーぶん万葉

んじゃNode.jsのところを実装してみよう

- Node.jsのC++アドオン作成して、そこからCannaのクライアントライブラリを呼んでみます。
- サンプルコードも本家日本語サイトにあるので、まずはそこからえいっと持ってこよう!
- あっさり書けた! 動かしてみよう!!!

…あれれ～、動かないぞ～!??
(それ以前にビルドが通らん!!!(^.^;

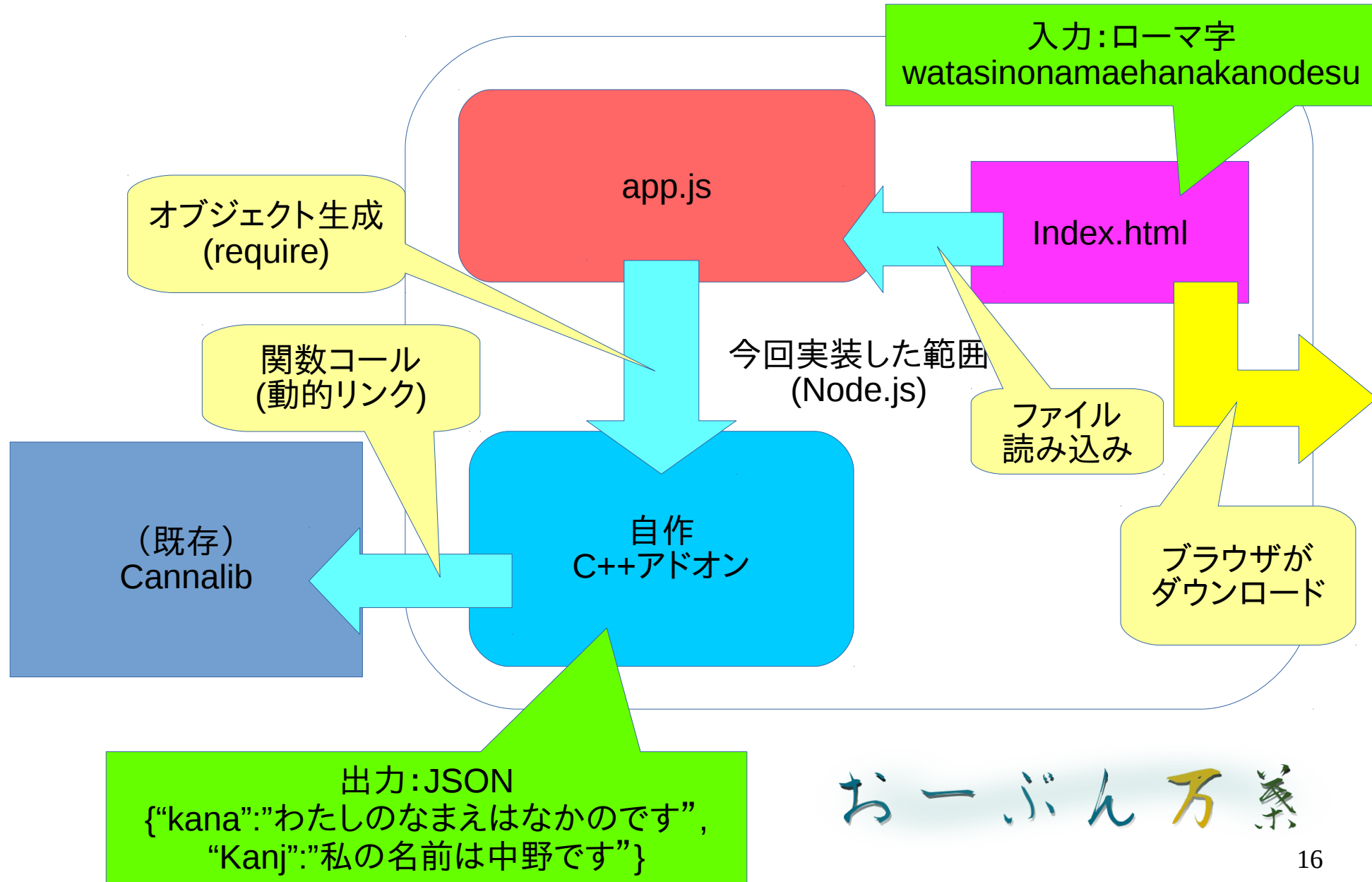
おーぶん万葉

ビルドが通らない…ってことは

- 理由は単純
「本家日本語サイトは全然翻訳されてません」orz
URL: <http://nodejs.jp/> ←このこと。
 - Ver. 0.10とか書いてながら、そのバージョンではビルドも通らないものが載ってたりします。
 - 今翻訳に関わってる暇ないのだorz（申し訳ね～×2）
- てわけで、githubにあるサンプルを利用しましょう！
 - URL: <https://github.com/rvagg/node-addon-examples>

おーぶん万葉

今回はこんなサンプルを作ります



どんな実装？

ソースを読んでね！（え？

<https://github.com/hashimom/node-canna>

※気になる点がございましたら
ブースに来ていただけたら返答致しますm(_ _)m

おーぶん万葉

C++の実装

- JavaScriptについては、私よりもぐーぐるさんに聞いていただけたらよろしいかと思えます^^;
- C++の肝となるのは「**binding.gyp**」というファイル

```
{ "targets": [  
  {  
    "target_name": "node_canna",  
    "sources": [  
      "src/mod_node_canna.cc",  
      "src/nodecanna.cc",  
      "src/cannasession.cc"  
    ],  
    "link_settings": {  
      "libraries": [ "-lcanna"  
    ]  
  }  
]  
}
```

JSONで書かれたMakefileみたいですね
下記のコマンドでビルドします。

node-gyp configure
node-gyp build

こんな感じで
共有ライブラリ(libcanna)をリンクします

おーぶん万葉

実はちょっとした仕掛けが組みられています

- 単純に、Requestが来たらResponseを返すという実装だと、どこか面白くないかな〜と。
 - 通信速度的にもボトルネックになるかも???
- というわけで、少し仕掛けを作ってみました
 - 正確には今回Node.jsにした理由はこれだったりします

おーぶん万葉

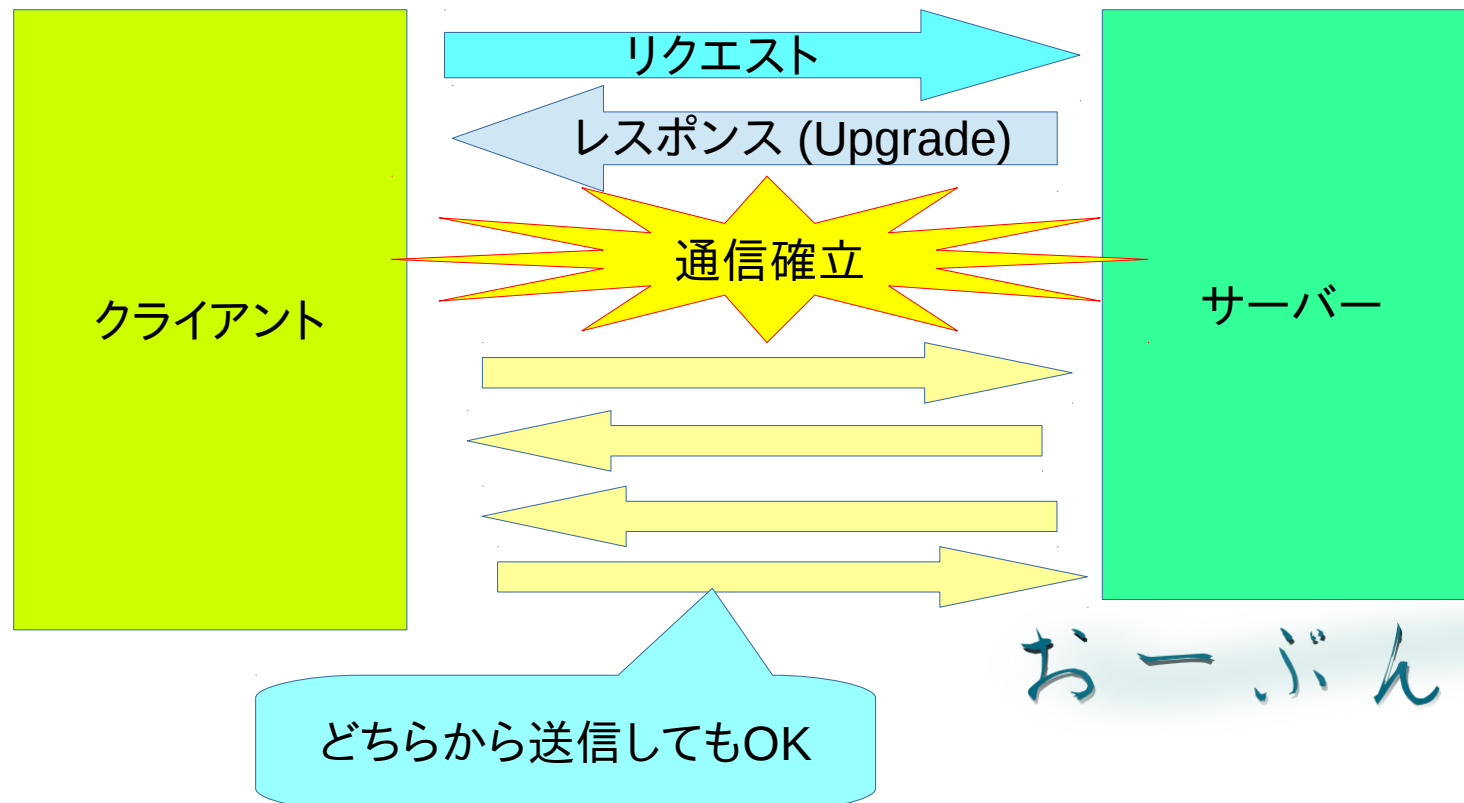
新しい日本語入力として提案

WebSocketを使ってみよう!

おーぶん万葉

WebSocketとは？

- RFC 6455で定義されている新しい通信規格
- 一度コネクションをはってしまえばつなぎっぱなしという双方向通信の規格です。



これを使えば日本語変換が速くなる？

- もともとCannaには「逐次変換機能」というものが実装されています。
- 逐次変換機能と、WebSocketを使えば、入力と非同期で変換結果が返ってくるはず。

…が、今回はそこまで辿り着けませんでしたorz
Cannaライブラリについて少し研究が必要です。

→近いことはできたかも？

おーぶん万葉

Nginxについて

- Nginxをリバースプロキシとして利用します。
 - 「http://(IPアドレス):(Node.jsのポート番号)」ではなく「http://(IPアドレス)」のみでアクセスできるようにします
 - Nginx 1.3からWebsocketに対応してます

```
server {  
    listen    80;  
    server_name localhost;  
  
    location / {  
        proxy_pass http://localhost:3000/;  
        proxy_http_version 1.1;  
        proxy_set_header Upgrade $http_upgrade;  
        proxy_set_header Connection "upgrade";  
    }  
}
```

こんな感じの設定が必要です。

おーばん万葉

実演…の前に

うまくいけば

- 今回の実演には、BeagleBoneBlackを使います。
 - クライアント側ではなく、サーバー側ですw
- microSDにはDebianがインストールされています
その場合の注意点は以下のとおり。
 - Nginxはセルフビルド(バージョンが古かったため)
 - Node.jsもセルフビルド(パッケージがない)

Node.jsのビルド前のconfigureは
「./configure **-without-snapshot**」で実行する!
(ARMの場合は必須なの???)

おーぶん万葉

それではお待ちかね!

実演です。

おーぶん万葉

今後の課題 (Izumoの取り組み)

- **WebSocket**と**逐次変換**はぜひやってみたい!
 - ただし、Node.jsを使う必要は・・・(ちょっとね)
 - WebSocket使いたい場合は、Javaの方がいいかも?
 - WebSocketだとサーバー側はマルチスレッド必須?
- JSONもいいが、速度を出すならバイナリ転送?
 - 元々Cannaプロトコルはバイナリプロトコル
 - 但し、アプリ連携を考えたらJSONか?

おーぶん万葉

おーぷん万葉からお願い

日頃、日本語入力を利用して
普段思っていることを是非ご連絡ください

- フォーラム:

<https://groups.google.com/d/forum/openmanyo>

どんな話題でもOKです!

コミュニティの力で新しい形の
日本語変換システムが作れたらと考えています。

よろしくお願いいたします!!!

おーぷん万葉

ご清聴、ありがとうございました!

次回予告: 2015/2/28 OSC東京 LT大会
「アヒルヤキを変換してみよう」
(通ればね…)

おーぶん万葉