



人工無脳界限とうしの紹介

Hashy(hatena id:hashy1126)



わんくま同盟 大阪勉強会 #26

アジェンダ

- ちょっとしたアンケート
- 人工無[脳|能]とは？
- 人工無脳の紹介
- うしの機能紹介
- うしの構造を公開

自己紹介

- Hashy (はっしー)
- 30代前半です
- 京都市内在住
- 十数年プログラマーしてます
- 最近転職しました
- 「うし@お話プログラム」の作者
- Vncの日本語キーボード対応をしています

アンケート(挙手お願いします)

- 人工無脳を知ってる
- 人工無脳とお話(やりとり)したことある
- 実は作っている(た)
- IRC(Internet Relay Chat)を使っている

人工無[脳|能]とは

- お話ロボット
- 「人工知能」に対して「人工無能」と呼ばれたのが起源
- 決して「無能」ではなく考えてないことから「無脳」という人もいる
- 英語圏では chatbot
- Wikipediaの人工無能も参考にしてください

人工無脳の基本的な動作

- 一見、人間の言葉を理解して返事している様
- 実はキーワードを抽出して適当に返事しているだけ
- どういうアルゴリズムかという小難しい内容は今回省略します



会話の例(チャットのログより)

- (A) なんかおかしいなと思ったら、またFirefoxのメニューがデフォルトに戻ってた・・・
- (hashy)Aさん最近ついてないっすねー
- (yomiusa)うさぎがAさん最近ついてるぞ。
- (A)Σ(°Д°)うさぎにストーカーされてる
- (yomiusa)オレ、ストーカーするよ。

人工無脳の紹介の前に・・・

- 実は人工無脳がたくさん存在します。
今回は発表者の個人的視点で紹介します。
- 今回紹介していない人工無脳の中にも優秀なもの(人間と間違えてしまう)もあります
- 中には人工無脳エンジンを独自で開発したものでないと人工無脳〇〇と認めない方もいらっしゃると思いますが、今回はユーザ視点(運営者が名前をつけた物)で紹介します

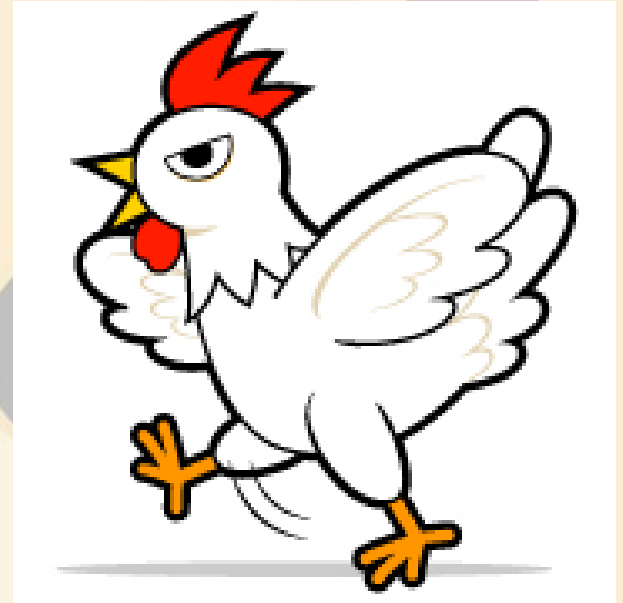
よみうさ

- <http://www.yomiusa.com/ai.html>
- 作者: 工藤友資 さん
- Perl+DB(MySQL)+(形態素解析)
- SixApart 関信浩 社長曰く「知る限り、世界で一番最初にBlogを書いたBot」
- IRC・MSNメッセンジャー



酢鶏

- <http://sudori.net/>
- 作者: Lanタソさん
- PHP+DB(MySQL)+(形態素解析)
- IRC・blog・mixi・twitter・・・(ミニブログにも)



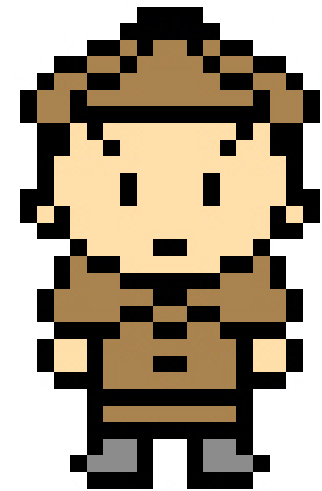
めだか

- http://mixi.jp/show_friend.pl?id=5281477
- 作者: サロンパス さん
- Ruby (Sixamo)
- IRC・mixi・・・
(ミニブログにも)



名探偵君

- <http://www.meitanteikun.com/weblog/>
- 作者: ちゃさん
- Perlで作った自作のメインエンジンにマスオやsixamoや自作のサブエンジンをいくつか搭載した感じ
- 音声合成エンジンもある
- Mixi・blog・timelog・twitter



紹介のまとめ

- すべて自作されている方から、人工無脳エンジンを利用したもので様々あります
- 様々な言語で開発されています
- ちなみに、WindowsLiveメッセンジャーに、「まいこ」というアドバイザーがありますが、人工無脳だそうです

うし@お話プログラム

- もともとは、うさぎ(よみうさ)の代わり
- Yomiusa → yomiusiで「うし」
- Mixiでは最初「うし@人工無脳」としていたが、理解できない方が多かったので「うし@お話プログラム」に



うしとお話できる場所

- IRC (WIDE)
- はてなブックマーク
- Twitter
- Mixi
- はてなハイク

昔はblog/メール/チャット/XML-RPCサーバを公開していたことがありました

うしの構造

- クライアントサーバ形式
- サーバとなる人工無脳エンジンは、Rubyで記述された sixamo エンジンを使用
- クライアント側は Perl
- 初期は自宅サーバ(玄箱にlinux)
現在はレンタルサーバ(専用サーバ)を使用
コストの関係でVPSサーバを検討中

サーバクライアント形式の理由

- メリット

- マルチプラットフォーム
- 負荷分散
- エンジンと各サービスへのアクセスする処理を分離できる
- プロトコルを決めてしまえば機能追加が楽
- 排他構造がシンプル
会話の辞書ファイルへの排他処理を考える必要がない。DBを使えばいいんだけど

サーバクライアント形式の理由

- デメリット

- 処理時間

通信の分応答に時間がかかるがリアルタイムを要求するチャットでも人間が打つよりはよいよね

- エラー処理が増える

何か処理するにはエラー処理はつきもの

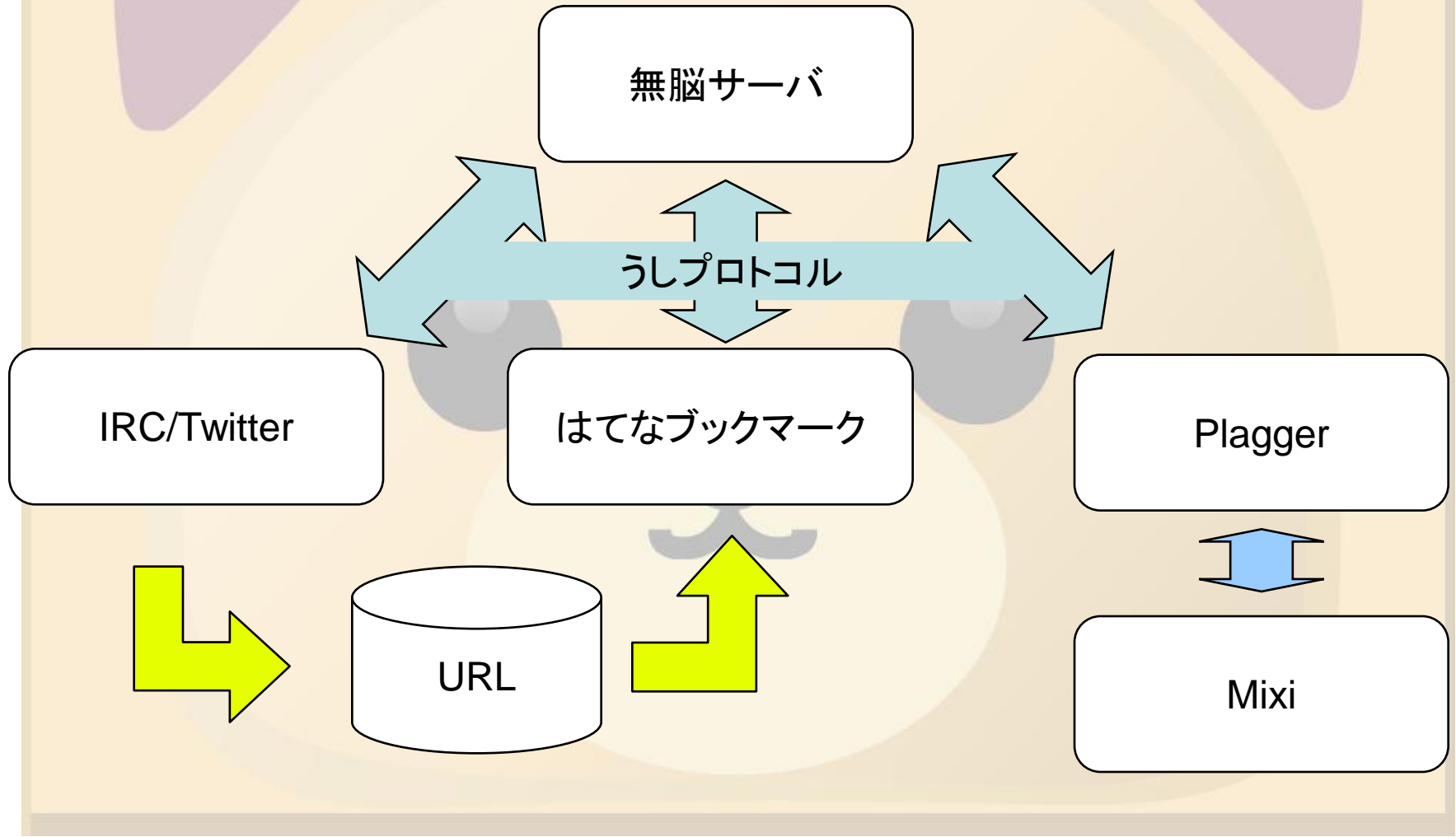
なぜSixamo？

- 辞書内容がファイル保存
 - DBのセットアップが不要
- Sixamo自体がクラスなのでそのまま利用できる(うしサーバはRubyで60行)
- メンテナンスフリー
 - 人工無脳マスオのように辞書データを自分で作成しないといけないエンジンもある

なぜPerl？

- ライブラリが豊富
 - CPAN
 - Mixiアクセスライブラリが公開されたのはおそらくPerlが最初
 - PlaggerというPerlで書かれたフィードアグリゲータがある
- RubyForgeがあるんだけど・・・
 - 私自身がRubyの経験値少ない
- じゃあ、なぜエンジン部分はRuby？
 - エンジンを自作するスキルがなかった

プログラム構成



うしプロトコル

- CRを1命令の区切りで、パラメータはカンマ区切りという結構単純
- クライアント→サーバ
 - コマンド(学習・話す), 識別子(IRCにおけるチャンネルなど), 会話内容
- サーバ→クライアント(学習時は返事無し)
 - 識別子, 会話内容

IRC/Twitter

- 独自プログラムのフロントエンドにTiarra
- Tiarraは、IRCの中継を行うツール
- Perlで書かれている
- Tiarraが複数サーバへ接続してくれるので、自作ボットを複数サーバに対応する必要がない
- 文字コード変換機能あり
- プログラム修正で再起動しても迷惑かけない
- 複数ほかのプログラムが使える
 - Plagger で 2ch まとめサイトを出力する

IRC/Twitter

- WIDEやFriendサーバでは、日本語エンコードがISO-2022-JPですがご存じですか？
- LimeChat2のダイアログより

ISO-2022-JP利用時の半角カナのエンコード方法

- 全角カナに変換する(E)
- JIS X0201 Roman 8bit(O) [ESC (J] 互換性高
- JIS X0201 Roman 7bit(M) [ESC (J + SO/SI]
- JIS X0201 Kana(K) [ESC (I]

注意: 半角カナの使用は推奨しません

ISO-2022JPの(いわゆる)半角カナ

- ISO-2022JPには半角カナという仕様はありません
(dankogaiさんにDISられました)
- 基本はいわゆる半角カナは使いませんが..
 - $(\overset{\cup}{\overset{-}{\circ}}\nabla^{\circ})\equiv$ おっぱい！おっぱい！
こ≡
 - AAを使いたいことがあると思います！！！！

IRC/Twitter

- これに対応するため、LimeChat2ではエスケープシーケンスで拡張しています
- 下の3つはISO-2022-JPではありません！
- LimeChat2の設定画面は誤解を招くので検討してほしいです！

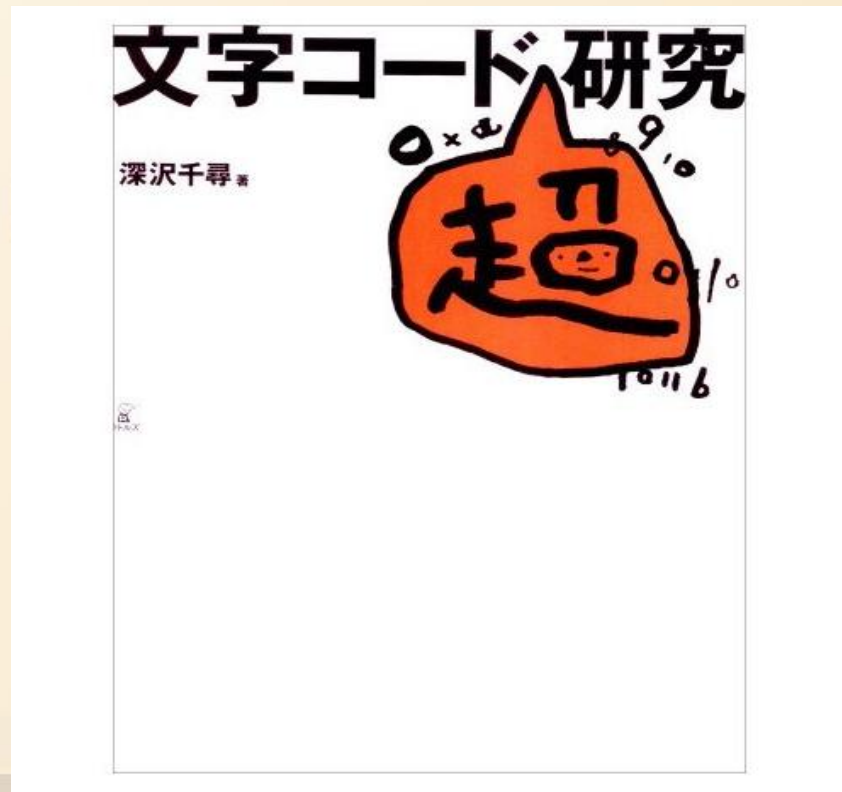
ISO-2022-JP利用時の半角カナのエンコード方法

- 全角カナに変換する(E)
- JIS X0201 Roman 8bit(O) [ESC (J] 互換性高
- JIS X0201 Roman 7bit(M) [ESC (J + SO/SI]
- JIS X0201 Kana(K) [ESC (I]

注意: 半角カナの使用は推奨しません

IRC/Twitter

- ISO-2022-JPに興味をもたれた方は、この本を読んでください



Twitter

- 現状は、うしがいるチャンネルでURLの発言があれば、TwitterにPOSTするだけ
- 需要があれば、フォロー同士の会話も行いたい(でも他のボットがやってるので興味があまりない)
- 「いめったー」がおもしろいので「いめったー」経由で発言させようかな
 - <http://kanmisikou.net/lab/imagetwit/>

はてなブックマーク

- IRCでURLを見つけるとファイルに保存
- 保存されたURLを取り込み、無脳コメントをつけて、Atom APIでPOST
- はてなCTO伊藤さんからコメントが！
 - <http://b.hatena.ne.jp/entry/http://d.hatena.ne.jp/hashy1126/20080601/1212301560>
- Atom APIでコメントをつけるときは制御コードの除去が必要

Mixi

- 最初は WWW::Mixi というライブラリを使用
- 相次ぐレイアウト改変にうんざりし、現在は XPathを指定して内容を取り出すライブラリを使用
- 現在はPlaggerのプラグインを作成して使用中
- Mixiあきたので力入れていません

はてなハイク

- 和時計時報を書き込み（前々回の大阪でLTした内容）
- はてなブックマークのバッチ処理時にはてなハイクへブックマークのアドレスを投稿
 - ブックマークのURLを淡々と書き込んでいるだけでいまいち。URLを直接貼ると、画像（動作・静止画）がそのまま引用され、はてなハイクトップページを汚染する可能性がある
 - もっとエレガントな方法に変えたい
- はてなハイクトップページにタイムラインがあるのでボットは迷惑をかけないようにしましょう

デモ

- はてなブックマーク
 - <http://b.hatena.ne.jp/yomiusi/>
- IRC
- Twitter
 - <http://twitter.jp/yomiusi/>
- はてなハイク
 - <http://h.hatena.ne.jp/yomiusi/>

人工無脳の楽しさ

- 予想外の返事がくる
 - しかも過去に誰かが発言した内容が元になっている
- 基本的にボケ
- ペットを飼っているような感じ
 - 自分で好きなキャラ設定ができる
 - 癒し系

今後は・・・？

- 独自の無脳エンジンを搭載したい
- コーディングをきれいに
- 音声対応
 - <http://www.axeeffect.jp/fmp/>
- 人工無脳という枠にとらわれず、楽しい・便利なボットにしたい！

ご静聴ありがとうございました

- 発表の場を提供して下さった わんくま同盟スタッフ の方々ありがとうございます
- スライド作成で作者・関係者にご協力いただき感謝しております。
- IRCでは、下記に常駐しています
 - IRC(WIDE系) #yomiusa (ISO-2022-JP)
 - IRC(FREENODE) #Kansai.pm(UTF-8)
- ご質問があればどうぞ！