チャットボットのススメ

Kazuhiro NISHIYAMA

株式会社Ruby開発

Osaka RubyKaigi 02 #osrk02

2019/09/15

自己紹介

- 西山 和広
- Ruby のコミッター
- twitter, github など: @znz
- 株式会社Ruby開発 www.ruby-dev.jp

Agenda

- チャットとは
- チャットボットとは
- なぜおすすめなのか
- 色々なボット例
- 人工無脳の話
- ボット作成時の注意

チャットとは?

- 主にテキストで
- リアルタイムに
- 会話(chat)するシステム

使ってきたチャットシステム

- NIFTY-Serve の RT
  - → RT を含むサービス終了 (@nifty 自体は存続)
- IRC → 今も利用
- Slack → 今のメイン
- Idobata, Gitter, Discord, ... → たまに利用

チャットボットとは?

- 誰かの発言に対して反応して発言
- 時間や通知など他のトリガーで自動発言
- システムによって使えるトリガーは違う
  - RT は入退室で挨拶するボットが可能だった
  - IRC は入室の挨拶は可能だが退室前は不可能
  - Slack は使われ方が違うので挨拶ボットは困難

なぜおすすめなのか? (1/3)

- プログラムの基本は入力を受けて出力を返す
- 出力がないと
  - 計算だけしてもできているのかどうかわからない
- 入力がないと
  - 同じことしかできない (そういう用途も多い)
    - テキスト以外に時刻なども入力

なぜおすすめなのか? (2/3)

- チャットボットはテキスト入出力で単純
- 昔ながらのチャットシステム (IRC など) だと行ごとの入出力でさらに単純
- テストもしやすい
- 標準入出力で試せるようにするのも簡単

なぜおすすめなのか? (3/3)

- 複数対応するといろんな経験を積める
  - 本質的な部分とシステム依存の分離
  - システムによる機能差
  - システム依存だが便利な機能の実装
- 長期運用でシステム移行も経験可能
  - 例: RT から IRC への移行を経験

小ネタ系ボット例

- `ping` → `pong` と返すだけ
  - ボットの生存確認やユーザー側の接続確認

便利ボット例

- URL → タイトル
  - 文字コード, JavaScript, HTML 以外の対応など、簡単そうに見えて実際やってみると大変
- g> 検索キーワード → Web 検索のトップ
- tenki> 大阪 → 天気予報
- amedas> 大阪 → アメダスの最近の値

反応例

```
g> OsakaRubyKaigi02
osrk02 - https://twitter.com/hashtag/osrk02 (and 3 hits)
tenki> 大阪
tenki bot: 大阪府 大阪 の天気: 今日:晴時々曇(max:31),
 明日:晴れ(min:24,max:35), 明後日:晴時々曇 - 近畿地方は、
 北部や中部では高気圧に覆われておおむね晴れていますが、南部では
 湿った空気の影響でおおむね曇り、雨の降っている所があります。
 (2019-09-14T10:31:00+0900)
 - http://weather.livedoor.com/area/forecast/270000
amedas> 大阪
amedas: 2019年09月14日　大阪(オオサカ)　時刻:15時,
 気温:31.6℃, 降水量:0.0mm, 風向:北東, 風速:3.8m/s,
 日照時間:0.6h, 湿度:49%, 気圧:1008.9hPa
 http://www.jma.go.jp/jp/amedas_h/today-62078.html?groupCode=45&areaCode=000
amedas: 2019年09月14日　大阪(オオサカ)
 最低気温(℃):22.7 at 05:53, 最高気温(℃):32.0 at 13:43,
 最大瞬間風速(m/s)(風向(16方位)):9.1(東) at 01:14
```

時報ボット

- 「西暦2019(平成31/昭和94/大正108/明治152/皇紀2679)年(己亥)09月(長月)14日(土)08時00分28秒(インターネットタイム@000.3)です。」
  - 過去の元号や干支や旧暦の月なども対応
  - Swatch のインターネットタイムも対応 (24時間=1000ビート、タイムゾーンなしの世界共通の時刻)

時報ボット

- 「西暦2019年(令和元年/平成31年/昭和94年/大正108年/明治152年/皇紀2679年)(己亥)09月(長月) **13日(金)** 18時00分00秒(インターネットタイム@416.7)です。」
  - 13日の金曜日は装飾

挨拶ボット

- 挨拶ボット
  - IRC のように接続通知がないと難しい
  - 「prefix+時間の挨拶+suffix > nick+敬称」(秘蔵のランダムデータで生成)

挨拶ボット例

- 「次はおはようございますみゅ > NICK御中」
- 「なんでもいいおはようございますするか？ > NICK代表」
- 「積極的におはようがなんともうれしい > NICKタン」

運用系ボット

- uptime : サーバーの uptime 確認
  - 「| uptime」の出力を発言するだけ
- upgradable : apt で upgradable になっているパッケージ一覧
  - 「apt list -qq -o APT::Cmd::use-format=true -o APT::Cmd::format=${Package}(${installed:Version}->${candidate:Version}) --upgradable」

人工無脳とは?

- 「人工無能」とも
- AI (人工知能) ほど高度なものではないということから
- ここでは人間の発言に自動発言を返すボット

人工無脳の種類 (1/3)

- 単純な部分文字列マッチ
  - Slackbot のようにチャットシステム側に存在することも
- 正規表現マッチ
  - 個人運用しているものはこれ

人工無脳の種類 (2/3)

- 形態素解析を利用
  - 人工無脳ししゃも (Sixamo) など
  - 書籍「恋するプログラム―Rubyでつくる人工無脳」の後半はこれ

人工無脳の種類 (3/3)

- ELIZA
  - 人工無脳の起源, 英語のパターンマッチング
  - 内容によってはかなり自然な会話ができるらしい
- 複数手法の組み合わせ
  - 人工無能うずら (ソース非公開)
    - 日本語の IRC (IRCnet) では多分一番有名

人工無脳への攻撃対策 (1/2)

- 問題例: ReDoS (regular expression denial of service)
  - 正規表現をユーザー登録可能な場合
- 不適切な語彙の学習
  - 任意の言葉を登録可能な場合

人工無脳への攻撃対策 (2/2)

- 対応例 (クローズドなグループ向けを想定)
  - 技術的な対応はあまりしない
  - アカウントを ban するなどのソーシャルな対応が楽 (ボットが無視するアカウントにするなど)

不適切な学習の問題

- Tay というマイクロソフトの Twitter ボット
- ユーザーによる不適切な学習によりヘイトスピーチなどをするようになったらしい
- (その後停止)

対処は難しそう

ボット作成時の注意

- ボットとボットのループに注意
  - ボットの発言にボットが発言するとループの可能性があって危険
  - IRC ではボットは NOTICE で発言して NOTICE には反応しないのが原則
- 大量発言に注意
  - アカウントが ban される可能性も

脆弱性にも注意

- 任意コード実行ボットはかなり難易度が高い
- タイトル取得ボットは情報漏洩に注意
  - 例えば localhost に制限しているもの
  - “169.254.169.254”？

発言量の問題

- ほとんどの発言に反応 → 迷惑?
- 受け入れられていれば問題なし
  - 以前からそういうものとして存在
  - 新規ならボットとの会話用チャンネルに隔離?

単調にみえる反応

大量のパターンを用意しても…

- キーワード反応型
  - ユーザーの発言のバリエーションが少ない
    → 反応も単調
- 時間に依存するもの
  - ユーザーが使う時間がほぼ同じ
    → 反応もほぼ同じ

最後に

- ping pong のような単純なものから始めるのがオススメ
- Twitter のような公開の場所は攻撃される可能性もあがるので Slack などのクローズドな場所で始めるのがオススメ
- アイデア次第で簡単なものでも便利だったり面白くなったり