

How Can I Help?:

Zero-Shot Multi-Modal Automation with QA



Michael Du, Sam Masling
Nancy Xu

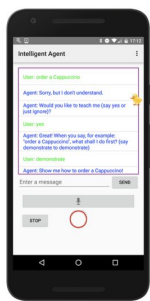
The **Average American** Spends **6hrs/day** on the Internet

- **Imagine an agent automated some of those tasks.** And we spent less time!
- **Virtual Personal Assistants (VPA)** ex. Alexa, Google Assistant, Siri, Cortana, and Bixby unable to cover **long tail of user requests**.
- **Programming by Demonstration** systems allow us to demonstrate new skills to agents.
 - 1. Prompting the user to provide a natural language utterance to refer to the skill
 - 2. Asking users to demonstrate the skill in the browser
 - 3. Capturing and name relevant variables and the sequences of clicks.
 - 4. Saving the demonstration to be called by name in the future.

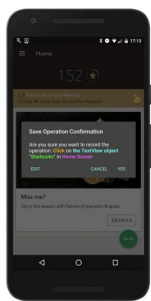


Programming **Dialogue Agents** on the Web is **Hard**

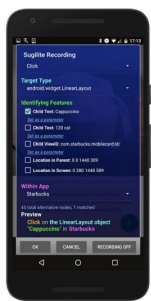
1. Require end-user to **demonstrate full space of possible browser actions** => time-consuming + incomplete.
2. **CSS selectors** are **brittle**.
3. Skills are **not generalizable** to new domains or sites.
4. Training **dialogue systems** is **non-trivial**.



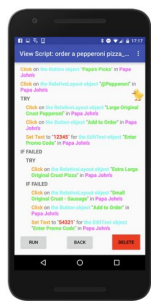
(a)



(b)

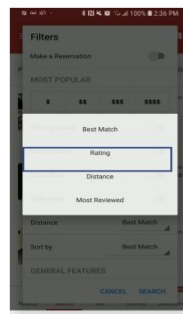


(c)



(d)

VASTA



(a)



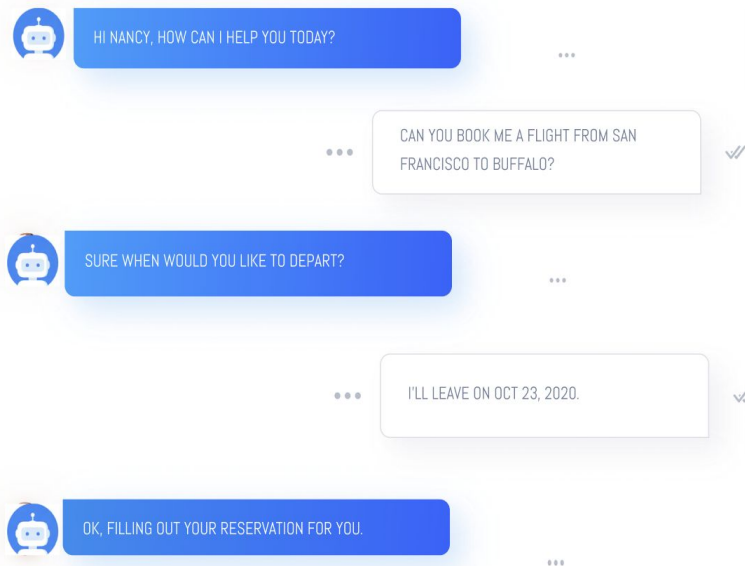
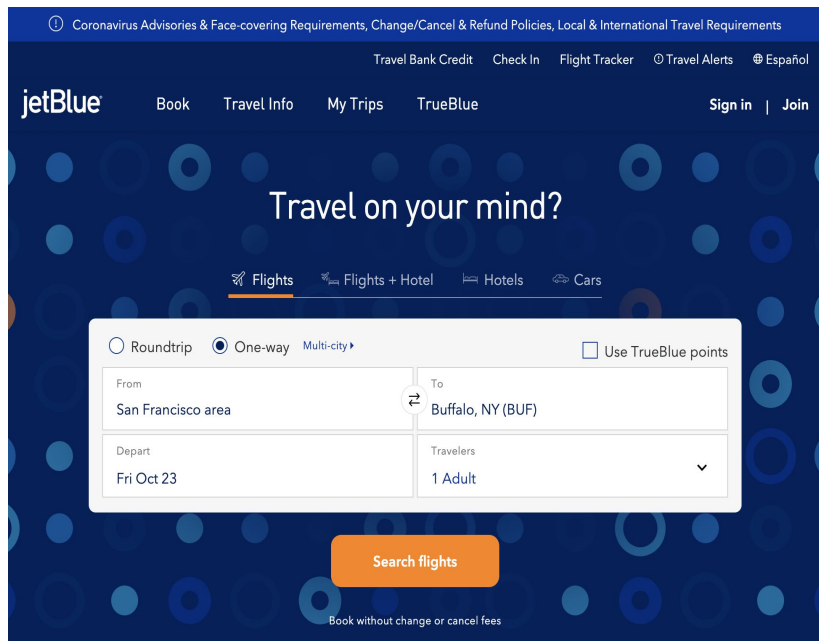
(b)



(c)

SkillBot

What if you could generate **an agent from any website?**



- Like a human reading a website -- no extensive demonstration needed.

Web Elements Perform 3 Main Purposes: **Inform** / **Request** / **Act**

CONTENT

Travel Bank Credit Check In Flight Tracker Travel Alerts Español

jetBlue Book Travel Info My Trips TrueBlue Sign in | Join

Travel on your mind?

Flights Flights + Hotel Hotels Cars

Roundtrip One-way Multi-city Use TrueBlue points

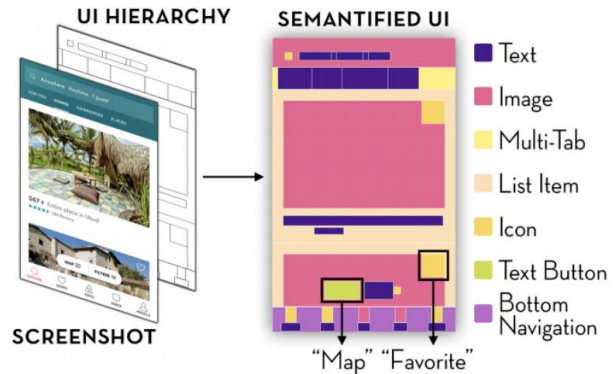
From **SLOT** To **SLOT**

Depart **SLOT** Travelers **SLOT**

ACTION

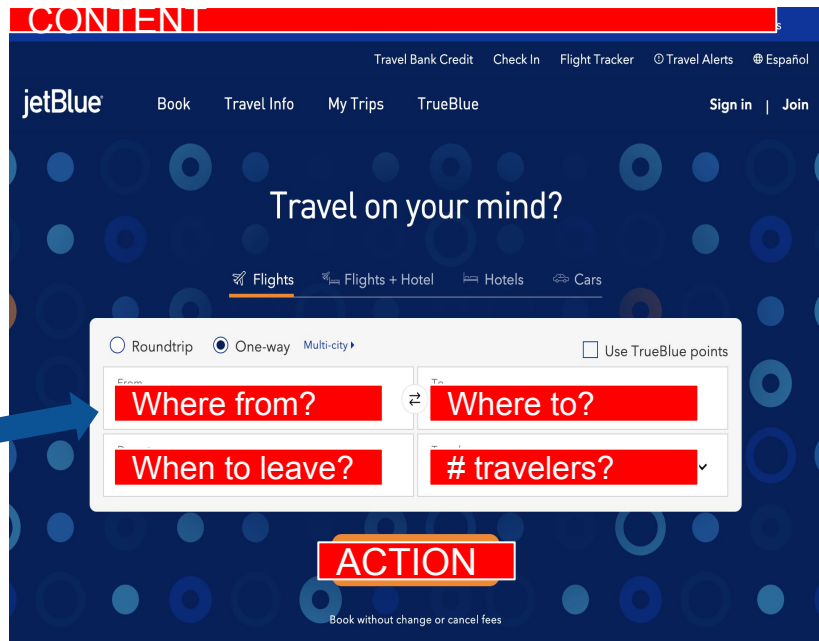
Book without change or cancel fees

```
<span _ngcontent-yqp-c59 class="clip ng-star-inserted" id="city-selector_ee61wkq7dlabel"> To Buffalo, NY (BUF) use arrow keys to select drop menu options </span> == $0  
<input _ngcontent-yqp-c59 autocorrect="off" class="avenir f5 fl ph3 pv2 pt0 bn core-blue w-100 h-100 border-box no-flying-focus ng-pristine ng-valid ng-touched" spellcheck="false" type="text" id="city-selector_ee61wkq7d" autocomplete="city-selector_ee61wkq7dac" aria-activedescendant="city-selector_ee61wkq7d" placeholder="Where to?" aria-labelledby="city-selector_ee61wkq7d">
```



HTML induced questions (with language models?) + UI Grammar Templates

```
▼<div _ngcontent-yeni-c39 class="w-100 center jb-max-width br3 mh1 mb5 bg-off-white pa3 flex flex-column mt1 flex-row-ns flex-wrap-ns justify-between-ns">
  ▶<div _ngcontent-yeni-c39 class="flex pa0 mb2">...</div>
  ▶<div _ngcontent-yeni-c39 class="ml-auto self-center dib-ns dn">...</div>
  ▼<div _ngcontent-yeni-c39 class="flex-row flex-ns w-100 justify-center" data-qa-id="cityPairSelector">
    ▼<div _ngcontent-yeni-c39 class="w-two-thirds-ns w-100-s">
      ▼<div _ngcontent-yeni-c39 class="w-100 order-2-ns ng-tns-c46-19" _ngghost-yeni-c46">
        ▼<div _ngcontent-yeni-c46 class="city-selector-wrapper flex justify-around flex-row-ns flex-column relative">
          ▶<jb-city-selector _ngcontent-yeni-c46 class="self-start w-100 w-50-ns mb1 mb0-ns mr1-ns" data-qa-id="citySelector" _ngghost-yeni-c48">...</jb-city-selector>
          ▶<jb-dialog-temp-wrapper _ngcontent-yeni-c46 class="ng-tns-c46-19" _ngghost-yeni-c49">...</jb-dialog-temp-wrapper>
          <!-->
          ▶<button _ngcontent-yeni-c46 class="t-fast relative-ns absolute cb z-2 b--light-gray bg-white icon-swap flex items-center justify-center h2 w2 ba br-100 pointer mr1 n13-ns nr3-ns ng-tns-c46-19 ng-star-inserted" data-qa-id="reverse" type="button">...</button>
          <!-->
          ▼<jb-city-selector _ngcontent-yeni-c46 class="self-start w-100 w-50-ns ng-tns-c46-19 ng-star-inserted" data-qa-id="citySelectorFrom" _ngghost-yeni-c48">
            ▼<div _ngcontent-yeni-c48 class="flex flex-column relative">
              ▼<div _ngcontent-yeni-c48 class="ipv2 pbl-ns bg-white contains-city-input b--medium-gray ba">
                <label _ngcontent-yeni-c48 class="slate-gray flex copy-xs ph3 dialog-dn clip-s">To/</label>
                ▼<div _ngcontent-yeni-c48 aria-haspopup="listbox" aria-owns="listbox" class="overflow-hidden" aria-expanded="false">
                  <!-->
                  <span _ngcontent-yeni-c48 class="clip ng-star-inserted" id="city-selector_pkuj4cem9label"> To Buffalo, NY (BUF) use arrow keys to select drop options </span>
                  <input _ngcontent-yeni-c48 autocorrect="off" class="avenir w-100 h-100 p2 pt0 bn core-blue w-100 h-100 border-box no-flying-focus ng-untouched ng-pristine ng-valid" spellcheck="false" type="text" id="city-selector_pkuj4cem9" autocomplete="city-selector_pkuj4cem9ac" aria-activedescendant="city-selector_pkuj4cem9" placeholder="Where to?" aria-labelledby="city-selector_pkuj4cem9label">
                </div>
                <!-->
                ▶<jb-region-swiper _ngcontent-yeni-c48 class="dn-ns ng-star-inserted" _ngghost-yeni-c52 hidden">...</jb-region-swiper>
              ...
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>
```



Where from?

Where are you flying from?

Where are you departing from?

What is the departure city?

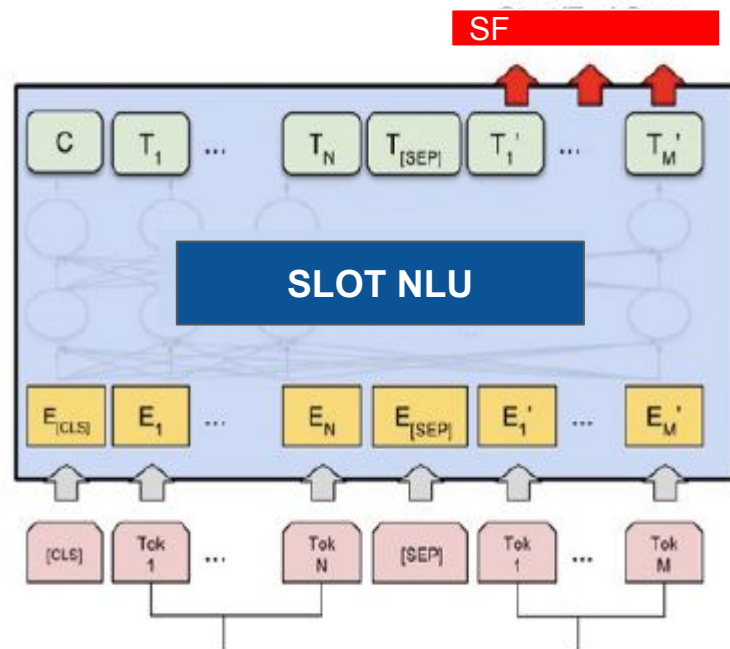
Zero-Shot Slot Filling + Navigation as Question-Answering

Please help me book a flight from **SF** to **JFK** departing on **Oct 30, 2020**.

The screenshot shows the JetBlue website interface. At the top, a red bar contains the word "CONTENT". Below it, the navigation menu includes "Book", "Travel Info", "My Trips", and "TrueBlue". The main heading is "Travel on your mind?". Underneath, there are tabs for "Flights", "Flights + Hotel", "Hotels", and "Cars". A booking form is displayed with the following elements:

- Radio buttons for "Roundtrip", "One-way" (selected), and "Multi-city".
- A checkbox for "Use TrueBlue points".
- Input fields for "Where from?" and "Where to?".
- Input fields for "When to leave?" and "# travelers?".

At the bottom of the form, a red box highlights the word "ACTION". Below the form, there is a small text: "Book without change or cancel fees".

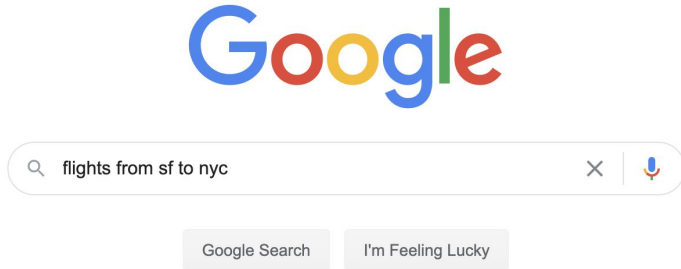


Where from?

Please help me book a flight from **SF** to **JFK** departing on **Oct 30, 2020**.

Demo: **SiteBot**, a multi-model conversational interface.

Book a flight by **navigating through Google** -> **OneBox** via Chrome extension chatbot.
Powered by QA NLU + Induced Questions



flight sf to nyc

Flights from San Francisco, CA (SFO) to New York, NY (all airports)

San Francisco, CA (SFO) New York, NY (all airports)

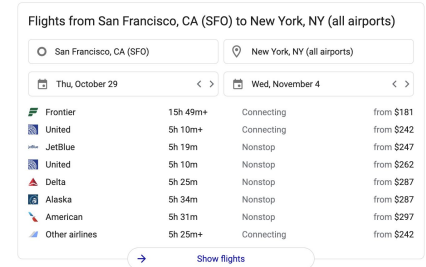
Thu, October 29 Wed, November 4

| | | | |
|----------------|----------|------------|------------|
| Frontier | 15h 49m+ | Connecting | from \$181 |
| United | 5h 10m+ | Connecting | from \$242 |
| JetBlue | 5h 19m | Nonstop | from \$247 |
| United | 5h 10m | Nonstop | from \$262 |
| Delta | 5h 25m | Nonstop | from \$287 |
| Alaska | 5h 34m | Nonstop | from \$287 |
| American | 5h 31m | Nonstop | from \$297 |
| Other airlines | 5h 25m+ | Connecting | from \$242 |

Show flights

Project Timeline:

- **Week 4:** Build a simple puppeteer agent that comprehends user utterance -> executes multi-modal automation for Google.
- **Week 5-6:** Study web structure + classify element types. Create question templates w/ ARIA etc. Also experiment with learning questions automatically from HTML with GPT 3 / language models. BoolQA models for actions (or CoQA) + ExQA on content.
- **Week 7:** Finetune Q&A models on synthetic training data generated by UI grammars + paraphrasing. Collect test data (user utterance + slots) on 10 websites using Mechanical Turk.
- **Week 8:** Build chrome extension interface within puppeteer browser for chatting with the agent.
- **Week 9:** Validate results on test data. Compare zero-shot QA technique against known benchmarks for slot-filling etc.
- **Week 10:** Leeway. Presentation. Paper. Etc.
- **Week 10 + Reach:**
 - Identify necessary slots for actions the seed multi-turn dialogue.



Flights from San Francisco, CA (SFO) to New York, NY (all airports)

San Francisco, CA (SFO) New York, NY (all airports)

Thu, October 29 Wed, November 4

| | | | |
|----------------|----------|------------|------------|
| Frontier | 15h 49m+ | Connecting | from \$181 |
| United | 5h 10m+ | Connecting | from \$242 |
| JetBlue | 5h 19m | Nonstop | from \$247 |
| United | 5h 10m | Nonstop | from \$262 |
| Delta | 5h 25m | Nonstop | from \$287 |
| Alaska | 5h 34m | Nonstop | from \$287 |
| American | 5h 31m | Nonstop | from \$297 |
| Other airlines | 5h 25m+ | Connecting | from \$242 |

Show flights