Almond: Keeping the Internet Open with An Open-Source Virtual Assistant

Monica Lam
Computer Science Department
Stanford University
lam@cs.stanford.edu

With Giovanni Campagna, Michael Fischer, Mehrad Moradshahi, Rakesh Ramesh, Richard Socher, Silei Xu, Richard Yang
Sponsors: AVG, Google, HTC, Hitachi, ING Direct, Nokia, Samsung, Sony Ericsson, UST Global
Consumer Privacy at Stake

- Facebook owns and sells 2-billion people’s personal data
- Cambridge Analytica incident
- EU GDPR (General Data Protection Regulation)
- There is no meaningful alternative
Unhealthy Commercial Ecosystem

- Platform Monopolies/Duopolies
- Google and Apple app stores: 30% revenues
- Google and Facebook: 60% digital marketing revenues

Monopoly — Open Competition — Innovation?
Virtual Assistants

50 millions in 2 years
Internet: 50 millions in 4 years

https://techcrunch.com/2018/03/07/47-3-million-u-s-adults-have-access-to-a-smart-speaker-report-says/
Virtual Assistants

Personalized: sees all personal info
Linguistic User Interface (LUI)
Intermediates all digital services
Controls choice of vendors
Human behavior analytics

Amazon, Facebook, Google Combined!
## Mobile & Ubiquitous: Graphical->Linguistic

<table>
<thead>
<tr>
<th>Graphical Web</th>
<th>Linguistic Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphical user interface (GUI)</td>
<td>Linguistic user interface (LUI)</td>
</tr>
<tr>
<td>Browser</td>
<td>Virtual Assistant</td>
</tr>
<tr>
<td>Web page addresses</td>
<td>Intents</td>
</tr>
<tr>
<td>Hosted by owners</td>
<td>Hosted by virtual assistants</td>
</tr>
<tr>
<td>Open Platform</td>
<td>Proprietary Platform? (like AOL?)</td>
</tr>
</tbody>
</table>

We are witnessing the start of proprietary linguistic webs.
Almond Vision:

- Open-source best virtual assistant technology
- Privacy: open federated virtual assistants -> choice
- New capability: Put users back in the driver seat!
  - Connect disparate resources
  - Share them with “who, what, when, where, how”

“Program” our virtual assistant in natural language!
Example: Asthma Patient

**people**
Dr. Smith:
“if Bob’s peak flow-meter drops below 180L/min let me know”

**environment**
Dr. Smith:
“when the air quality index is above 500 and Bob is running, warn him”

**location**
“Let my Dad know if I am at the hospital”

**devices**
“when I use my inhaler, record my GPS location in logfile on Box”

Bob
Impact of Natural Language Programming

- Power of language
- Consumers: today’s software cannot satisfy the long tail of user needs
- Professionals: automate their own repetitive tasks
- Big data analytics + privacy
Technology of Natural Language Programming
Core Concepts

Natural Language → LUINet Neural Network → ThingTalk → Thingpedia

- Formal Personal Web Programming Language
- Simple construct + Thingpedia Functions
- Signatures of all APIs

Open Interoperable Web
Natural Language Programming

“When I use my inhaler, get my GPS location, if it is not home, write it to logfile in Box.”

- Event-driven program
- Multiple function calls
- Parameter passing
- Filters on values
Almond: 1st Programmable Virtual Assistant

Natural Language Commands

```
When I use my inhaler, get my GPS location, if it is not home, write it to logfile in Box.
```

Semantic parsing

```
monitor @Inhaler-use(), => @GPS(), location <> “home”
=> @Box-write(file=“logfile”, data=location)
```

LUInet

Linguistic User Interface Neural Network

Formal target language

ThingTalk

Thingpedia

LAM Giovanni, Ramesh, Xu, Fischer, Lam, WWW 2017
**Thingpedia: Encyclopedia of Things**

- **Interoperability**
  - API signatures + corresponding NL
  - Not just intents
- **Open repository**
  - Available to Alexa, Google Assistant, ...

> 60 devices / 200 functions

<table>
<thead>
<tr>
<th>Action</th>
<th>Natural Language</th>
<th>API Signatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHEN</td>
<td>@Stanford tweets</td>
<td>Monitor (@home_timeline(), ...) author==“Stanford”)</td>
</tr>
<tr>
<td>GET</td>
<td>tweets matching “#Cardinal”</td>
<td>search(...), contains (hashtag, ...)</td>
</tr>
<tr>
<td>DO</td>
<td>tweet “Stanford won!”</td>
<td>post (status)</td>
</tr>
</tbody>
</table>
THINGTALK Compound Statement

**WHEN** [FILTERS] → **GET** [FILTERS] → **DO**

FILTERS: =, <, >, <=, >=, <>, contains, starts with, ends with

- When I use my inhaler, get my location, save them to Dropbox
- If I get taken to a hospital, let my dad know.
- When the air quality index is above 500, and I am running, send me an SMS.

- When the Bitcoin price reaches $10,000, search for a “bitcoin” picture, and tweet it with caption “I’m rich!”
Expressiveness of ThingTalk

- Inspired by IFTTT
- ThingTalk is a superset of IFTTT recipes (2 clauses)
- IFTTT has 250,000 unique recipes
- IFTTT provides a GUI: no formal or natural language
- IFTTT is proprietary: user must share credentials
Real Natural Language Input

When my car is at home, and it is not plugged in, send me a reminder email.

Remind me if my car is not plugged in at home. If I am not charging my car when it is home, let me know. Remind me to plug in my car whenever I’m home.
Technical Challenges

• Natural language training-data acquisition (Liang)
  • Formal language
    → Natural language templates
    → synthetic sentences
    → paraphrased sentences (and repeat)

• Co-design of Thingpedia, ThingTalk and LUInet
  • Combine when + get functions into 1 set

• Compositionality to handle scale
LUInet Results

- Dataset (60+ devices, 200+ functions)
  - Synthetic: 515K programs, 2.9M sentences
  - Paraphrased: 175K programs, 400K sentences
- Model:
  - Seq2seq bi-LSTM with attention, pointer network
- Accuracy: 89%
- Future work: real user input
Sharing with Privacy
Sharing is Broken Today

- Services have limited options
- With the cost of data ownership
- Sharing credentials

Let your virtual assistant help you share
General+Fine-Grain: ThingTalk Extension

Requester: GET-PREDICATE [FILTERS]

WHEN [FILTERS] → GET [FILTERS] → DO

FILTERS: =, <, >, <=, >=, contains, starts with, ends with

Let Dr. Smith monitor my peak-flow-meter, if it drops below 180L/min

Let my father monitor my security camera for motion,

Let my secretary, whenever I am out of town, read email messages whose subject is marked urgent.

Let my daughter, from 6-8pm, watch Netflix

Let my boyfriend get pictures from my dropbox, taken on Feb 14, and post them on Facebook
Almond: 1st Federated Virtual Assistant

Expressiveness:
Any ThingTalk command

Privacy:
Remote execution model
Owner executes requests
Returns need-to-know

Giovanni, Xu, Ramesh, Fischer, Lam, Ubicomp 2018
Distributed ThingTalk Protocol

(a) Request
“Ask @alice to notify me when her security camera detects motion.”

(b) Send program
σ=SELF, c=@alice: monitor @security_camera.event(), has_motion=true
⇒ return

(c) Check
σ=SELF, c=@alice: monitor @security_camera.event(), has_motion=true
⇒ return

(d) Ask for permission
σ=@dad, c=SELF: monitor @security_camera.event(), has_motion=true && @phone.get_gps() {location!=home}
⇒ return

(e) Respond
“Only if I’m not home.”

(f) Save

(g) Return detected events

Dad’s Assistant

Policy Database

Alice’s Assistant

“Notification from monitor security camera: motion detected ...”

Alice
Conformance of Access Control

- Natural Language
- 2nd-Party Program Access Control
- Satisfiability Modulo Theories (SMT)
Conformance Algorithm

• SMT: Generalization of boolean satisfiability (SAT) with theories of strings, arrays, …

• Provably correct programs for conformance, and synthesis of conforming code

• NP-hard, but fast enough in practice
Needs and Acceptance?
Do Consumers Need Access Control?

Role-Based Permission

- **Teenage daughter** to use credit card
  - % People comfortable in giving permission (200 person survey)
  - 100%

- **Amazon courier** to unlock door
  - If the package is over $1000
  - If your security camera is on

- **Friends** to access cloud drive
  - Photos with their faces in them
  - Photos in a specific folder

- **Parent/kid** to see security cameras
  - If you are not at home
  - Cameras facing the front yard/garage

- **10-year-old kid** to use Netflix
  - Between 7 PM to 9 PM
  - Free G or PG rated movies

Attribute-Based Permission

- With a $20 budget limit
- For restaurants only
More Examples

Willingness to share doubles with attribute-based access control
Expressiveness of ThingTalk?

Solicit use cases by showing AMT workers 3 examples, without describing ThingTalk or TACL

**Enforceable:**
Mom: “You need to follow this guy on Twitter, give me your Twitter account”.
Me: “OK, add him but don’t follow any other twitter user”.

**Unenforceable:**
Friend: “Can I use your library card?”
Me: “OK, only if you return the book on time”.

ThingTalk is Expressive

60 workers; 220 suggestions; 85 unique assets

Diverse use cases

- Business Accounts: 21%
- IoT: 27%
- Social Media: 16%
- Services: 1%
- Personal Data: 34%

85% in the scope of TACL

- Existing API: 70%
- New API: 15%
- Out of Scope: 9%
- Unenforceable: 6%

LAM
User Study: “Sharing Without Passwords”

Like the concept?

Like the app?

Use the app?

#users

![Bar graphs showing ratings for concept, app, and use.]
Almond: an Open Virtual Assistant

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperable Personal Web</td>
<td>Open, crowdsourced skill repository</td>
<td>Thingpedia</td>
</tr>
<tr>
<td></td>
<td>Personal web language</td>
<td>ThingTalk</td>
</tr>
<tr>
<td>Autonomy</td>
<td>Natural language programming</td>
<td>LUInet</td>
</tr>
<tr>
<td>Privacy, Openness</td>
<td>Open-source federated virtual assistants</td>
<td>Almond</td>
</tr>
</tbody>
</table>

Put users back in the driver seat in 5 years!
Last Chance for Privacy and Open Competition

http://almond.stanford.edu
Almond virtual assistant in Android

Starting: Massively Open Online Project

Stanford Profs. Bernstein, Boneh, Lam, Landay, Manning, Mazieres, Re
Closing Remarks