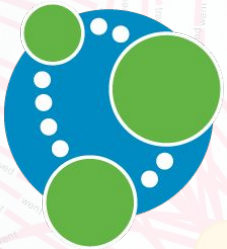
Graph NLU



Natural Language Understanding with Python and Neo4j

Dan Kondratyuk

July 2017

Summary

- About me
- Research Question
- Experiences with Neo4j
- Natural Language Implementation

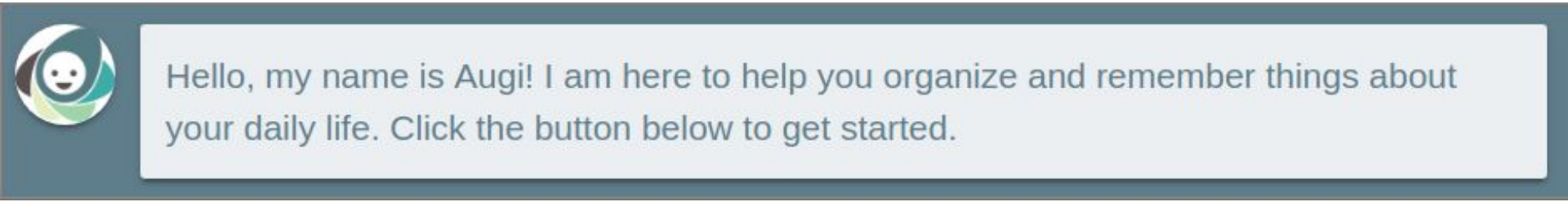
About Me

About me

- Researcher & Developer
- Recently graduated with a Bachelor's in Computer Science
  - Boise State University
- Member of Speech, Language, and Interactive Machines (SLIM)
  - Interaction ∩ Computation

About me

- Full Stack Developer
- Hello Augi
  - A personal assistant that remembers when you forget
  - helloaugi.com



Research Question

Intelligent Personal Assistants

Intelligent Personal Assistants of today are rapidly transforming how humans interact with machines.

Cortana

Siri

Alexa

Assistant

Memory Problems

Personal assistants behave like amnesiacs

- Their memory is inflexible, or nonexistent
- Interactivity is limited to predefined knowledge and known tasks

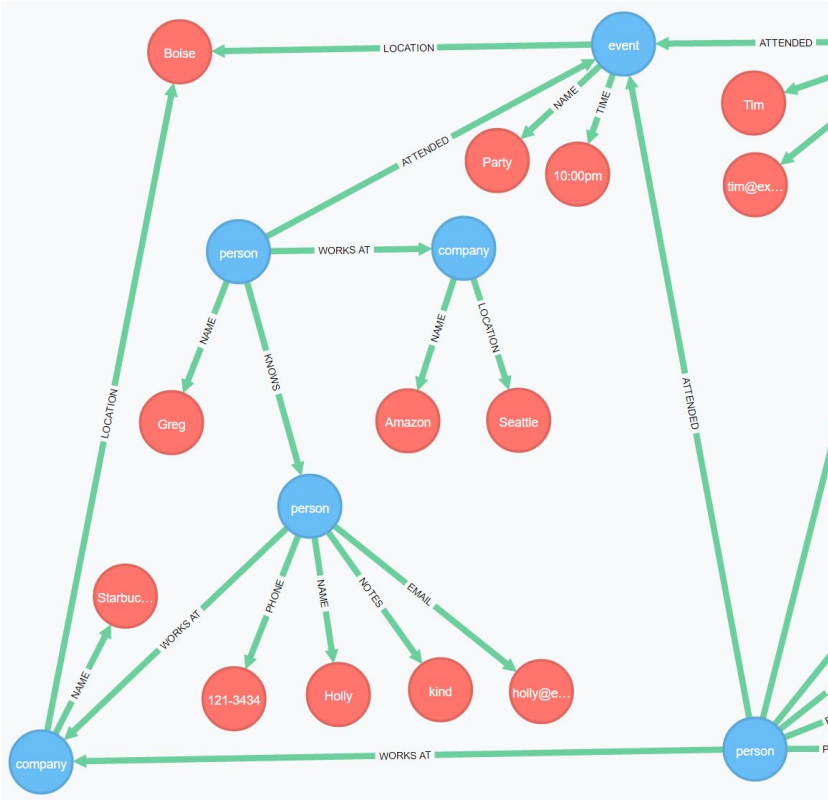
A Graph-Based Approach

How could interaction be stored in memory? How can that information be leveraged for further use?

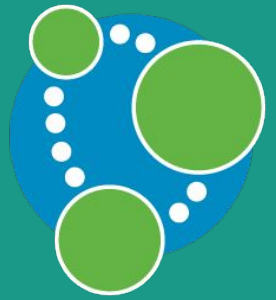
Real world information is highly relational

Goals

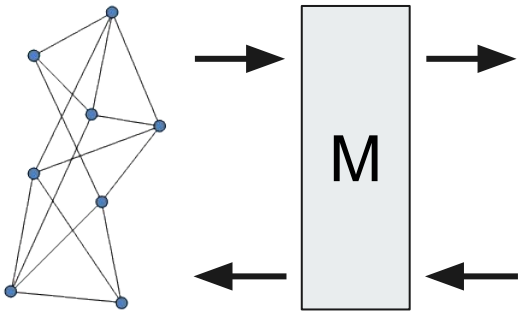
- Dynamic
- Interpretable



Experiences with Neo4j

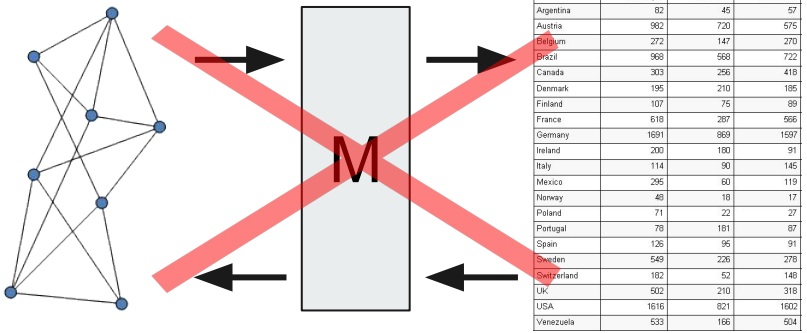


Neo4j + Natural Language Understanding

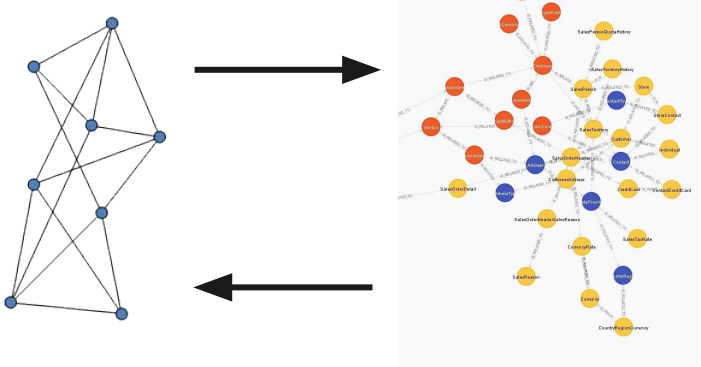


	Beverages	Condiments	Confections
Argentina	82	45	57
Austria	982	720	575
Belgium	272	147	270
Brazil	968	568	722
Canada	303	256	418
Denmark	195	210	185
Finland	107	75	89
France	618	287	566
Germany	1691	869	1597
Ireland	200	180	91
Italy	114	90	145
Mexico	295	60	119
Norway	48	18	17
Poland	71	22	27
Portugal	78	181	87
Spain	126	95	91
Sweden	549	226	278
Switzerland	182	52	148
UK	502	210	318
USA	1616	821	1602
Venezuela	533	166	504

Neo4j + Natural Language Understanding



	Beverages	Condiments	Confections
Argentina	82	45	57
Austria	982	720	575
Belgium	272	147	270
Brazil	968	568	722
Canada	303	256	418
Denmark	195	210	185
Finland	107	75	89
France	618	287	566
Germany	1691	869	1597
Ireland	200	180	91
Italy	114	90	145
Mexico	295	60	119
Norway	48	18	17
Poland	71	22	27
Portugal	78	181	87
Spain	126	95	91
Sweden	549	226	278
Switzerland	182	52	148
UK	502	210	318
USA	1616	821	1602
Venezuela	533	166	504



Neo4j + Natural Language Understanding

- Persistent storage that represents the problem domain
  - Excellent for ad-hoc relationships
  - Very flexible
- Intuitive syntax

Neo4j + Natural Language Understanding

- Efficient querying capabilities
- Future
  - Real-time synchronization
  - Event triggers
  - Prefetch

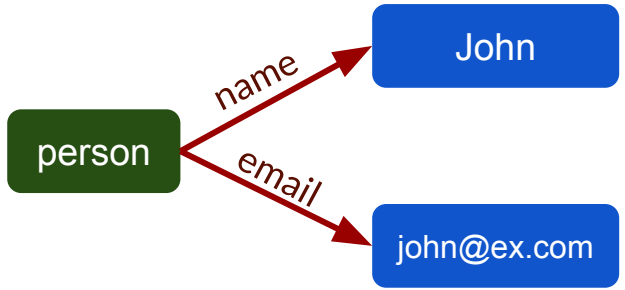
Natural Language Implementation

Use Case: Natural Language Database

1. Statement → update knowledge

"John's email is john@ex.com"

```
type: statement
name: John
email: john@ex.com
```

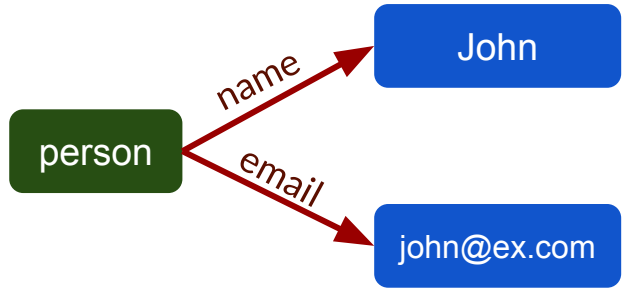


Use Case: Natural Language Database

1. Statement → update knowledge

"John's email is john@ex.com"

```
type: statement
name: John
email: john@ex.com
```



2. Question → retrieve results

"What is John's email again?"

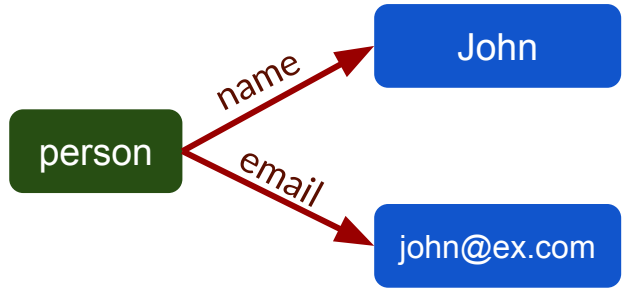
```
type: question
name: John
ask: email
```

Use Case: Natural Language Database

1. Statement → update knowledge

"John's email is john@ex.com"

```
type: statement
name: John
email: john@ex.com
```



2. Question → retrieve results

"What is John's email again?"

```
type: question
name: John
ask: email
```

"john@ex.com" ← `answer: john@ex.com`

Corpus: Facebook bAbI Tasks QA (1-20)

Steps

1. Tokenize and POS tag to produce *(subject, relation, object)* triple

2. Merge statement triples into a single graph

3. Answer each question using the graph, compute accuracy

Task 1

```
1 John travelled to the hallway.
2 Mary journeyed to the bathroom.
3 Where is John?  hallway 1
4 Daniel went back to the bathroom.
5 John moved to the bedroom.
6 Where is Mary?  bathroom 2
```

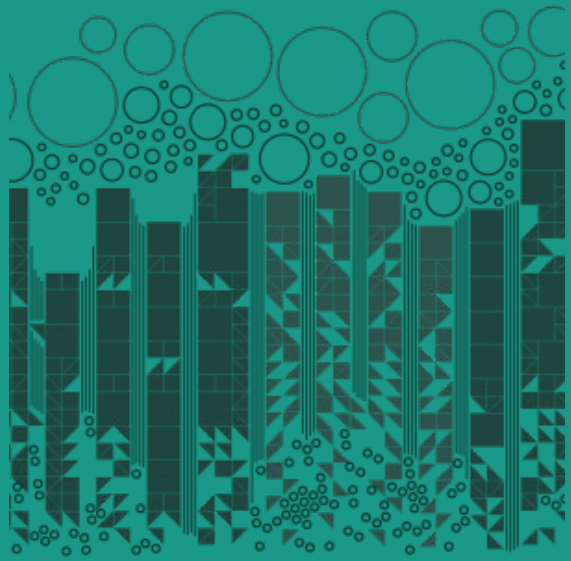
Results

- 100% accuracy for finding where a person is (simple data, no surprise)
- Allows for more complicated questions
  - What rooms has a person been in?
  - How many times has a person changed rooms?
  - How many times has a room been visited?

Future: use entity extraction to fill a frame, then merge with the database

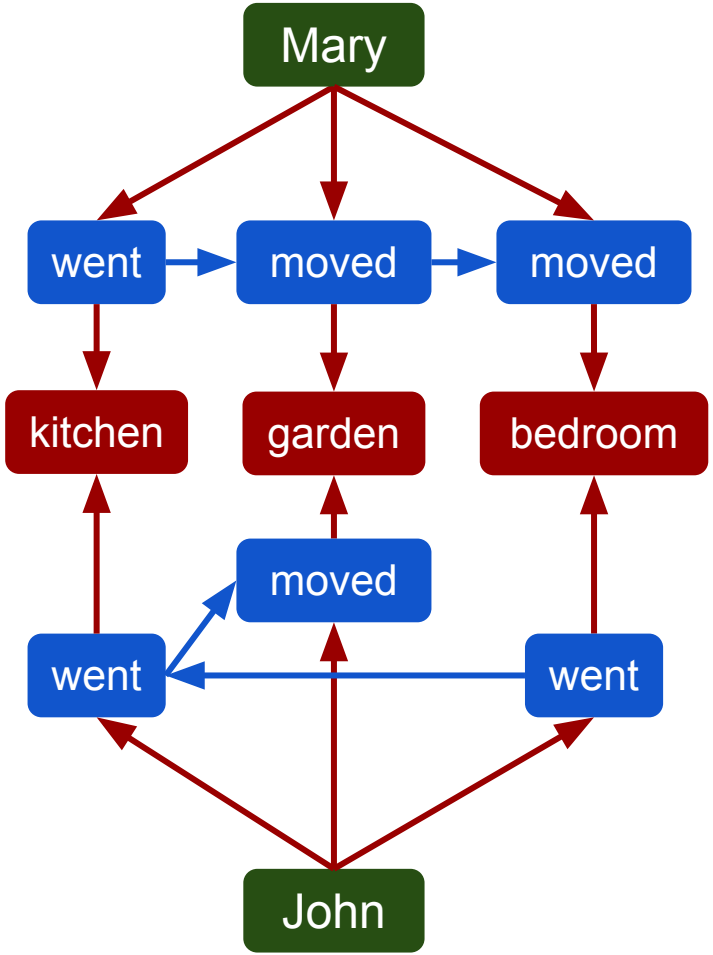
NLU + Neo4j = Promising

Extra Slides

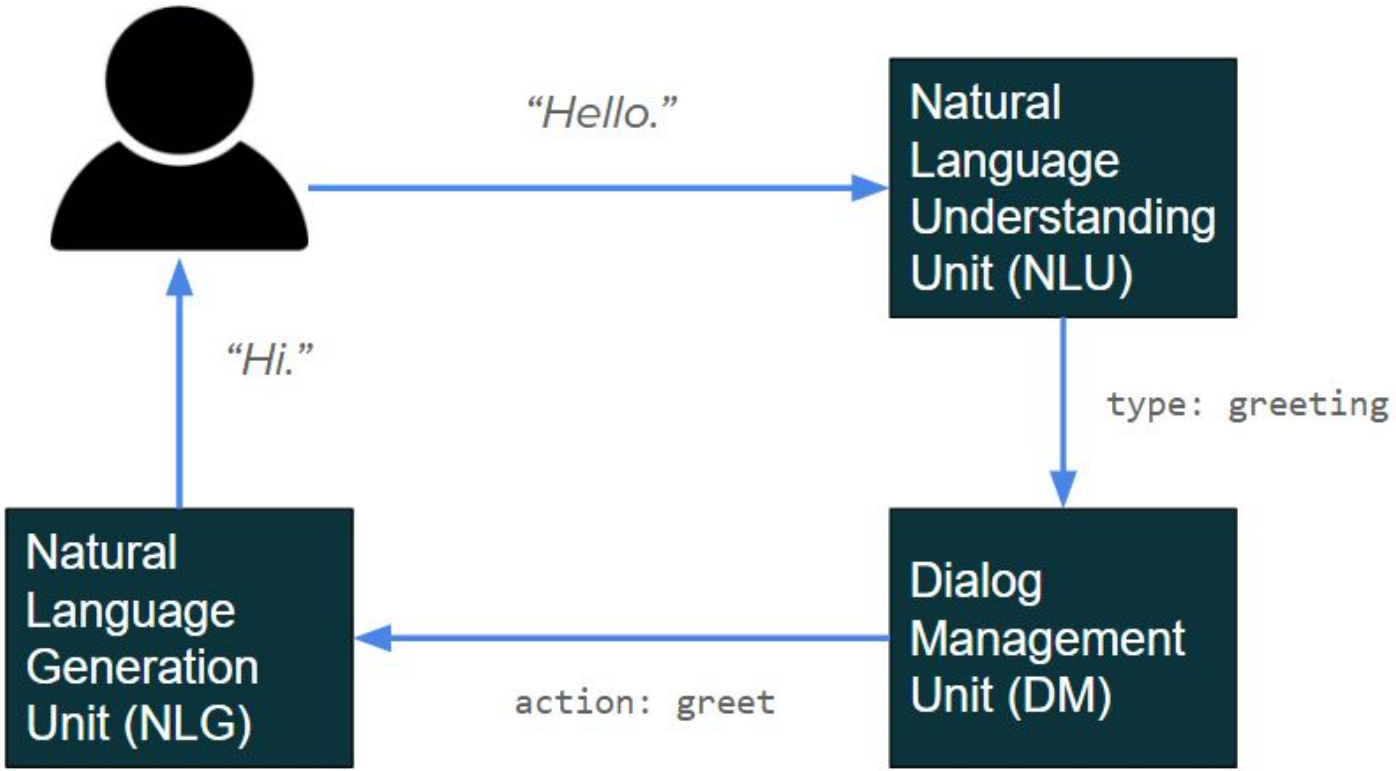


Approach: bAbI Tasks

- A graph of ***syntax*** over a sequence of ***statements***
- ***Semantics*** are derived from ***traversals*** over the graph



Dialogue Systems



Further Reading

- Siri shortcomings. https://www.reddit.com/r/SiriFail/
- Why Graph Databases? https://neo4j.com/why-graph-databases/
- Robinson, Ian, Jim Webber, and Emil Eifrem. Graph databases: new opportunities for connected data. O'Reilly Media, Inc., 2015.
- Lison, Pierre. Structured probabilistic modelling for dialogue management. Diss. University of Oslo, 2013. Chapter 2.
- Quepy natural language database http://quepy.machinalis.com/