Toward Ontology Based Chatbot Endowed with Natural Language Processing and Generation in Commercial Domain

Amine Hallili, PhD student
Catherine Faron Zucker & Fabien Gandon, Advisors
Elena Cabrio, Supervisor
Headlines

- Introduction
  - Motivations
  - Research questions
- Chatbot
  - Definition
  - Categories
  - Our Chatbot?
- Ongoing work
  - Our proposal
  - Knowledge Base
  - Ontology (Schema.org, GoodRelations)
  - Pattern Extraction
  - Property Matching
  - Response Generation
- Perspectives
- References
Introduction
Context & Motivations

Why?

- New means of communication
  - FAQ
  - Social Networks
  - Mobile Applications
  - Search Engines
- Huge amount of underexploited data especially in Commercial Domain
  - Linked Data
  - Log files
  - Raw Text ...
Research questions

- How to construct a Knowledge Base using website APIs?
  - Proposing a platform to extract information
- How to fully understand user’s question?
  - Natural Language Processing
- How to keep users interested in interacting with the system?
  - Natural Language Generation
  - Friendly interface
  - Dialog mode
Scenario

Give me the price of a Nexus 5!

the price of Nexus 5 is 400$!

and who sells it?

several sellers were found. The main one is Google! Do you want to see other sellers?

No, show me the white version, sold by Google and located in France!

here are the images of Nexus 5 white version, sold by Google and located in France...
ChatBot
Chatbot – State of the art

- Chatbot, ChatterBot, CleverBot, Chat-Robot (Allen et al) :
  
  Computer program designed to simulate an intelligent conversation with one or more human users via auditory or textual methods, primarily for engaging in small talk.

- Natural Language Dialog system (NLDs)
- Expert System (Liao 2005)
- Question Answering system (Hirschman & al)
- Multiagent system (Wooldridge 2009)
Chatbot – state of the art
Ongoing work
Our proposal

- Combining the benefits of both QA systems & NLDs to propose:
  - A rich KB for data extraction and reasoning
  - NLP tools to interpret user's question
  - NLG techniques to generate well-formed sentences.
  - Integrating Dialog mode to keep user interacting with the system.
Our starting point

- QAKiS (Cabrio & al 1)
  - Question Answering wiKi framework System
  - Test it at qakis.org
Our contributions

- QAKiS from Open Domain (DBpedia) => Closed Domain (Commercial)
- Natural Language Generation
- Question with constraints (N-Relations)
- Dialog Mode
SynchroBot

Question

Dialog Manager

NLP
Type Recognizer
Property Recognizer
NE Recognizer

Query Generator

N-Relations Handler

KB
Triple store
Subject  Predicat  Value

Off-line Feed
Pattern Finder
Triple Feeder

NLG
Response Formatter
Pattern Picker

Ontology
Knowledge Base creation

[eBay, Amazon, BestBuy] API
Ex: getPrice(Nexus_5) => 400$

Data Transformer

RDF Knowledge Base

<sbo:Product rdf:about='#Nexus_5'>
<sbo:hasPrice>400</sbo:hasPrice>
</sbo:Product>
<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF

xml:sbo="http://synchronext.com/ahallili/synchroBotOntology#"


<sbo:title>HAND CARVED ONE PIECE WOOD CHAIR. UNIQUE 1990 C.</sbo:title>
<sbo:country>US</sbo:country>
<sbo:label>Antiques</sbo:label>
<sbo:price>15.0</sbo:price>
<sbo:locatedIn>Orwigsburg, PA, USA</sbo:locatedIn>
<sbo:itemId>18131935864</sbo:itemId>
<sbo:globalId>EBAY-US</sbo:globalId>
<sbo:soldBy>

<sbo:Seller rdf:ID="AzurMatos">

<sbo:mailBox>contact@azurmatos.fr</sbo:mailBox>
<sbo:address>152 Boulevard Cessele, 06100 Nice</sbo:address>

</sbo:Seller>
</sbo:soldBy>
</sbo:Product>
Ontology reuse

- Why we need an Ontology?
  - Data structuration, Domain representation, Inference.

- Existing ontologies on commercial domain
  - Schema.org Ontology
    - Covers several domains
    - Used by state of the art search engines
    - Partial coverage of the commercial domain
  - GoodRelations Ontology (Hepp 2008)
    - Better coverage of the commercial domain
GoodRelations Ontology
GoodRelations Ontology

Linked Open Commerce

Vertical Ontologies
For specialized e-commerce Domains

Universal e-Commerce Model
http://purl.org/goodrelations/

Product Catalogues
From vendors and aggregators

Product or Service Model
Location
Payment Method
Business Entity
Camera
SLR Camera
Structured Product Sheet
Your Mega Store
Big Shops

The Long Tail
Small shops, SMBs, Blogs, Facebook, etc.

USD$800
"I sell my D90"
Aldo Bucchi
Joe's Photography LLC
Main Office
Open on Saturdays

USD$900
"D90 Just arrived! Photography lessons included"

USD$850
"Nikon D90"

http://linkedopencommerce.com/
Pattern Extraction - Algorithm

API based method
- For each property
  - Parse product pages
    - Get all sentences containing the domain and range values
  - Make generic patterns

- All pages are tested!
- Finds extra patterns
- Easy to implement

Crawler & annotation based method
- For each page => {Subject}
  - Parse annotation
    => Graph representing the page
    - For each property
      - Get all sentences containing the domain and range values
      - Make generic patterns

- Requires annotated pages
- More efficient
- Less time execution
Subject

The Galaxy S5’s matte backing rejects fingerprints like a boss.

Josh Miller/CNET

At the end of the day, the Galaxy flagship feels like it always has: like plastic. That's not necessarily a bad thing, though, because it has all the latest specs at all striving for loftier ambitions, it hasn't reached the heights of HTC's luxe brushed aluminum or even Sony's sleek style.

The GS5 is only a fraction larger than the Galaxy S4: it measures 5.59 inches tall by 2.85 inches wide by 0.32-inch deep, or 142mm by 72.5mm by 8.1mm; and it weighs 5.1 ounces, or 145 grams. Even though the extra hardware makes it a little taller and heavier than the Galaxy S4, I had no problem carting it around. Even outside my reach, usually home to at least two phones and a jumble of weighty items, the GS5 had no problem fitting into the pocket of my skinny jeans, and stayed wedged in my hand during several-mile runs. (I also had a death grip on it.)

This new phone, too, has a 5.1-inch 1080p HD AMOLED display that's 0.1-inch bigger than the Galaxy S4's. That means that the screen's pixel density is just a breath looser, though you'll never notice the difference. Images are still extremely crisp and colorful, with high contrast and true-to-life colors, and videos look especially lush.
Pattern extraction – Crawler Method

Properties

metadata

Sentences expressing properties

Jessica Dolcourt reviews smartphones and cell phones, covers handset news, and tests the latest models. She originally started at CNET in 2006 and spent four years reviewing mobile and desktop software before taking on devices.
Property Matching Module

<sbo:hasPrice>

[Product] price is [Double]

The price of [Product] is [Double]

[Product] costs [Double]

Give me the price of a Nexus 5!
Property Matching (N-Relation)

- 2-relations: Give me the address of Nexus 5 seller!
  - Give me the Nexus 5 seller!
  - Give me his address! => high score
  - NE: Nexus 5 => [Product]

\[\text{Nexus}_5 \xrightleftharpoons{\text{<hasAddress>}} 10 \text{ Jean Medecin, 06000, Nice}\]
Property Matching (N-Relation)

Graph representing the question:

- Property1: Domain: D1, Range: R1
- Property2: Domain: D2, Range: R2
- Property3: Domain: D3, Range: R3
- Property4: Domain: D4, Range: R4
- Property5: Domain: D5, Range: R5

Or / And?

No link???

No domain or no Range?!
Give me the price of a Nexus 5!

Nexus 5 costs 400$!
Give me the price of a Nexus 5!

Dialog Manager

NLP

<sbo:Product>

<sbo:hasPrice>

<sbr:Nexus_5>

Query Generator

Select ?v
where {
  <sbr:Nexus_5> <sbo:hasPrice> ?v
}

KB

Triple store

Subject  Predicat  Value
Nexus5    hasPrice   400$

Off-line Feed

Pattern Finder

Triple Feeder
Perspectives

- **Short term:**
  - NE Recognition improvement
    - KNN, Similarity, N-Gram, TF-IDF algorithms
  - N-Relations Implementation
  - Scale to a bigger KB

- **Middle term:**
  - Dialog Mode
    - Multiagent systems
    - Conversational behavior systems
  - Serendipity
References


