

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!

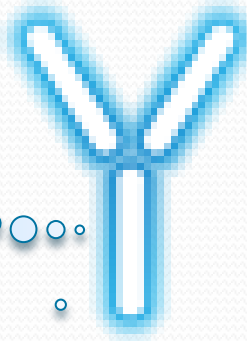
the price of Nexus 5 is 400\$!

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...



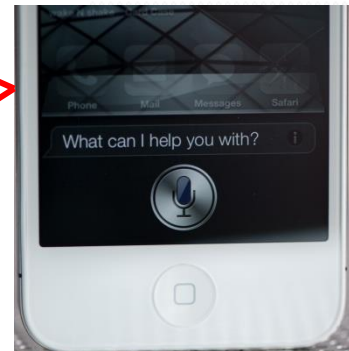
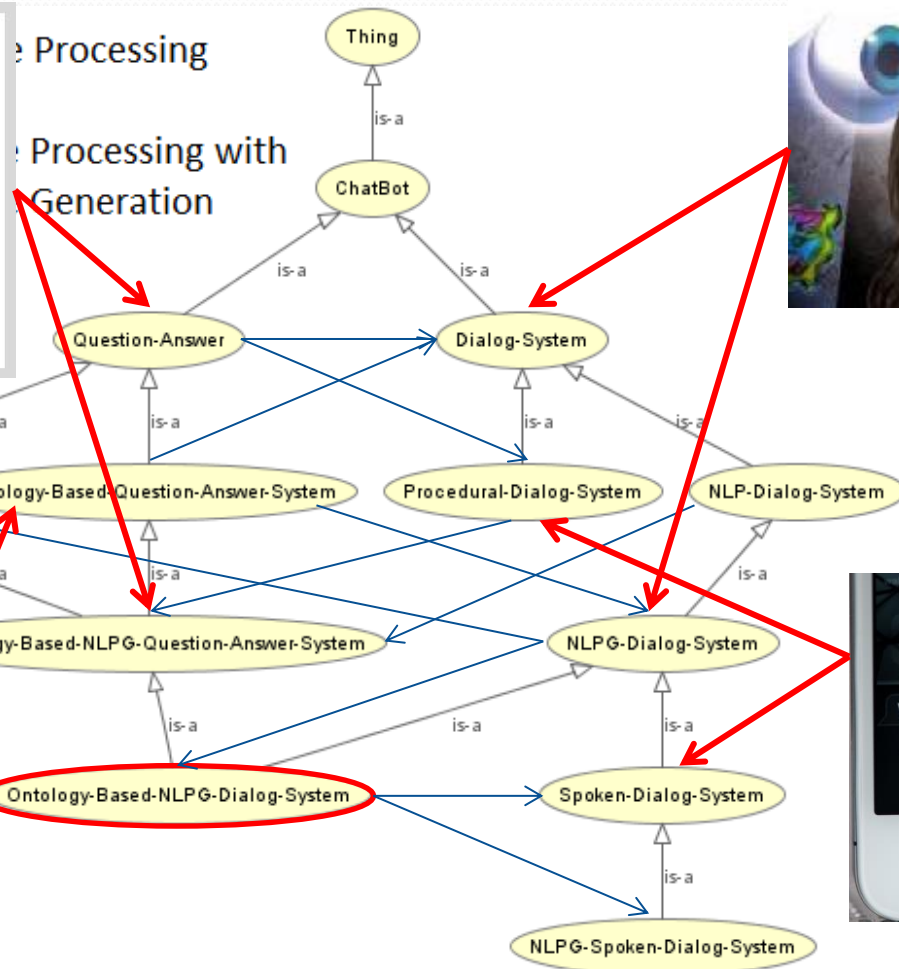


ChatBot

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





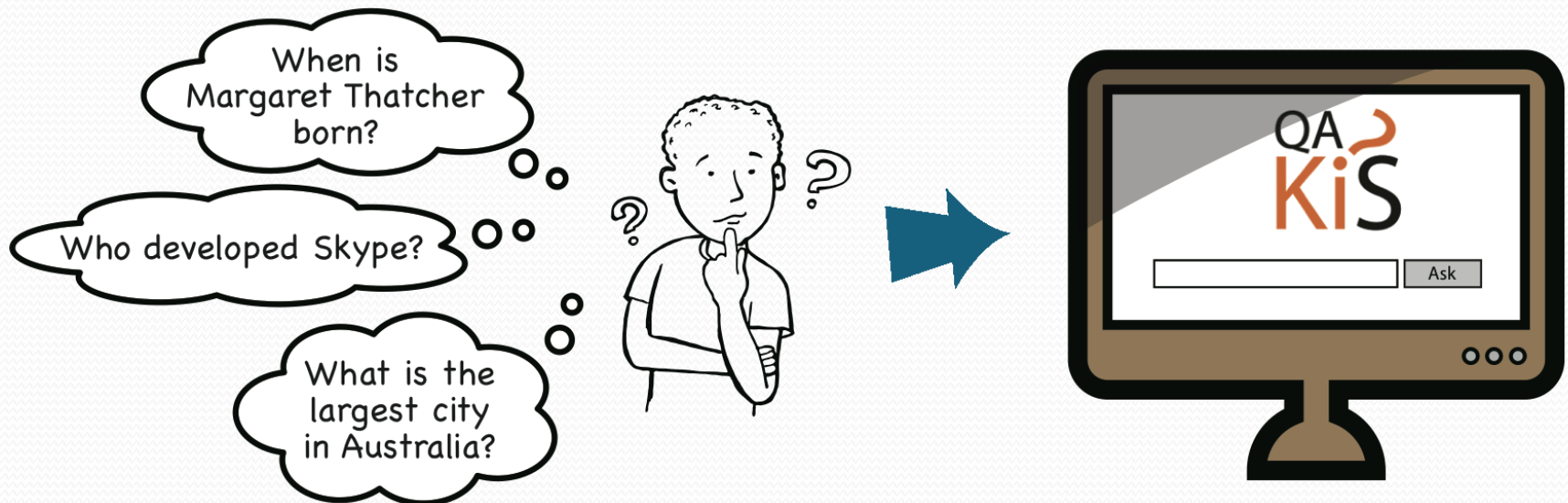
Ongoing work

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

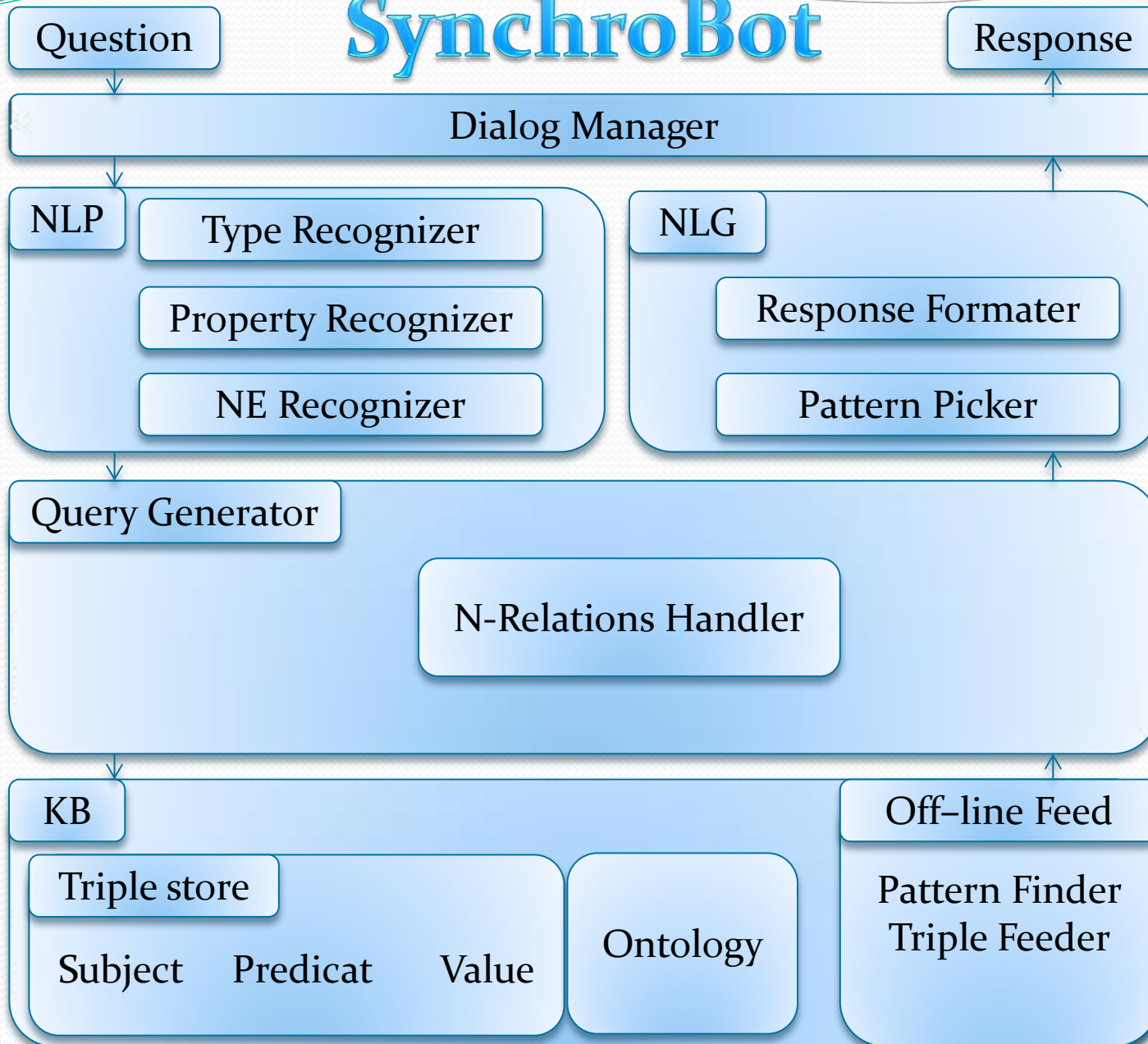
- QAKiS (Cabrio & al 1)
 - Question Answering wiKiframework System
 - Test it at qakis.org

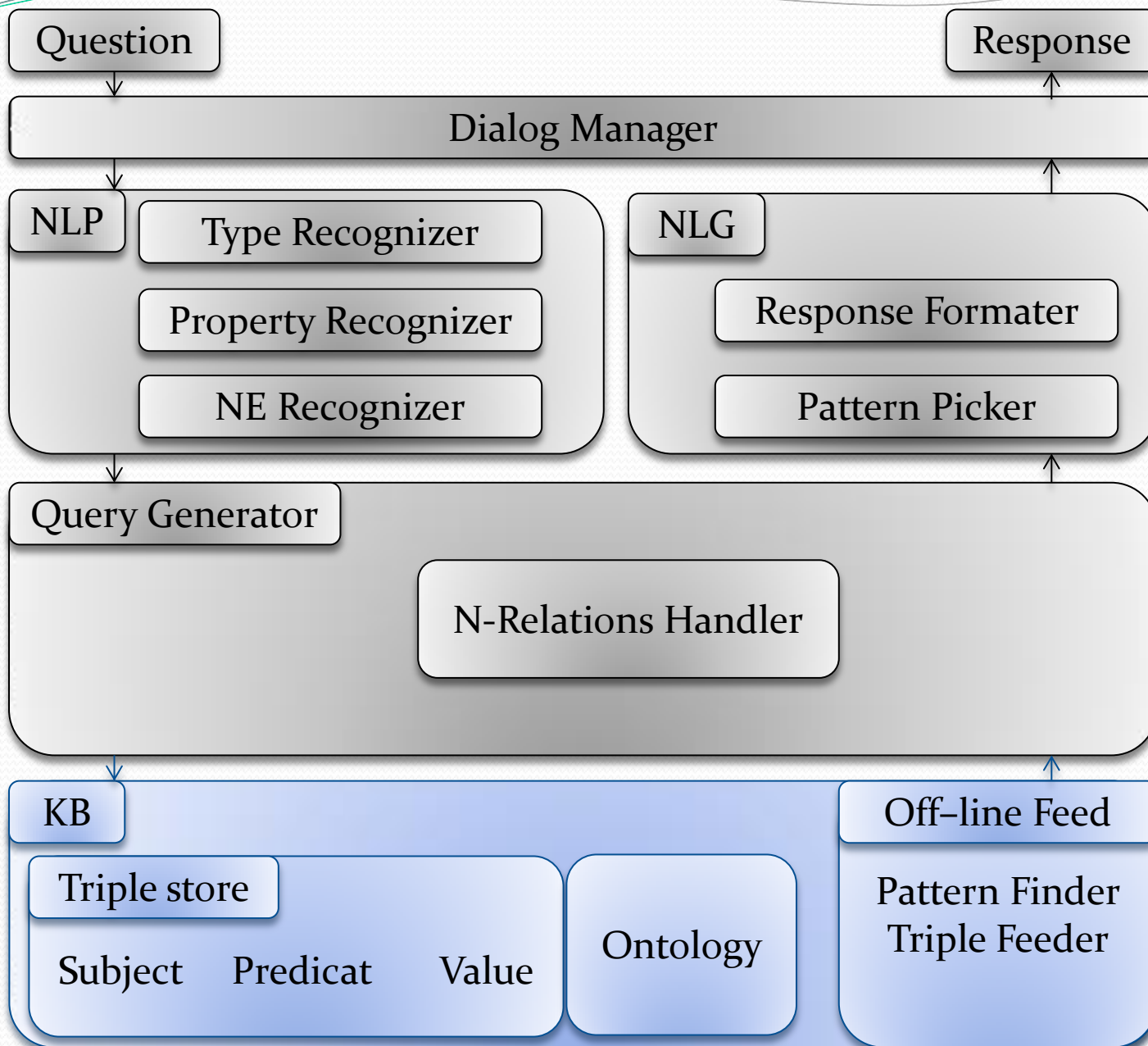


Our contributions

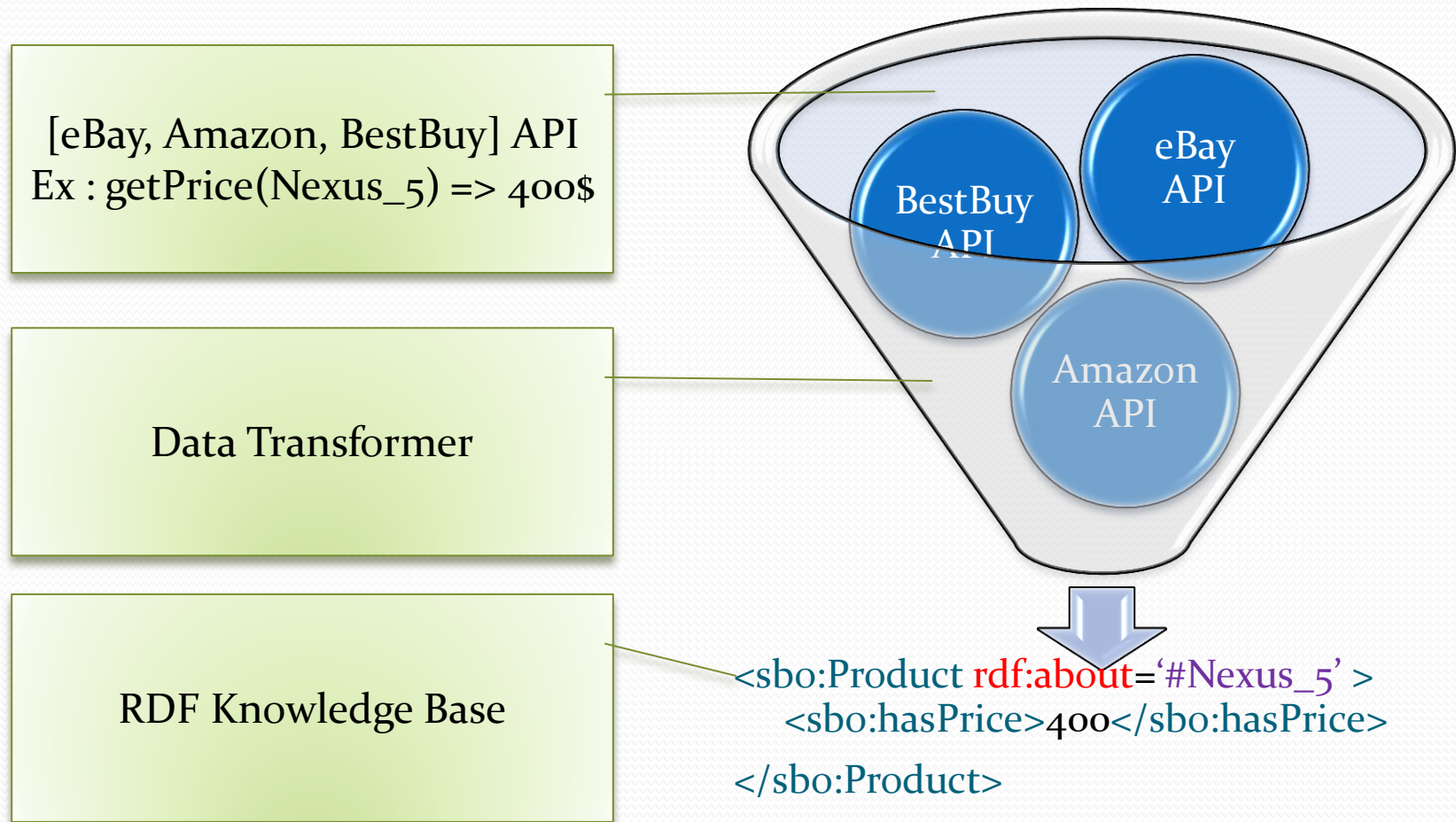
- QAKiS from Open Domain (DBpedia)
=> Closed Domain (Commercial)
- Natural Language Generation
- Question with constraints (N-Relations)
- Dialog Mode

SynchroBot



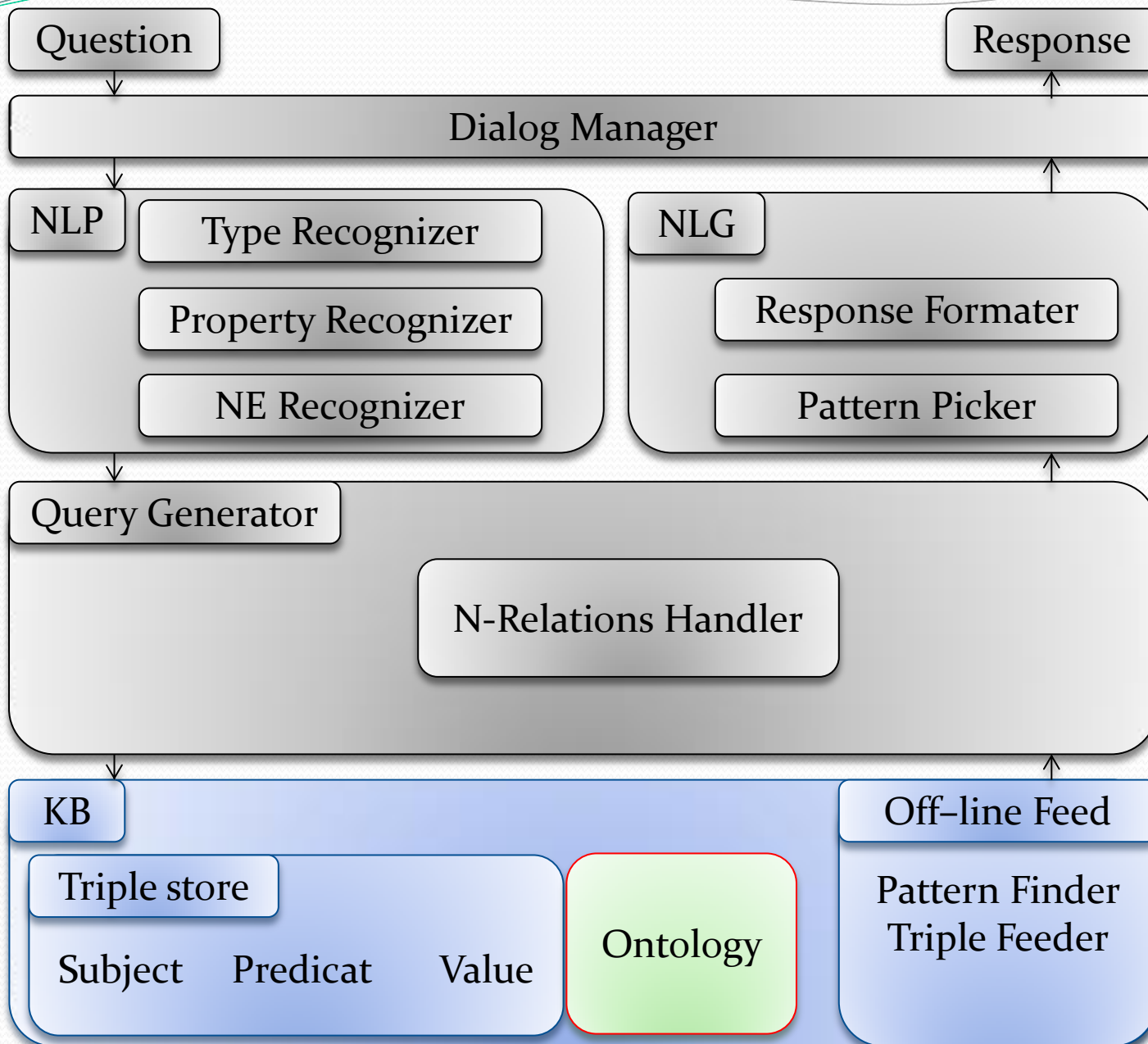


Knowledge Base creation



Knowledge Base - Example

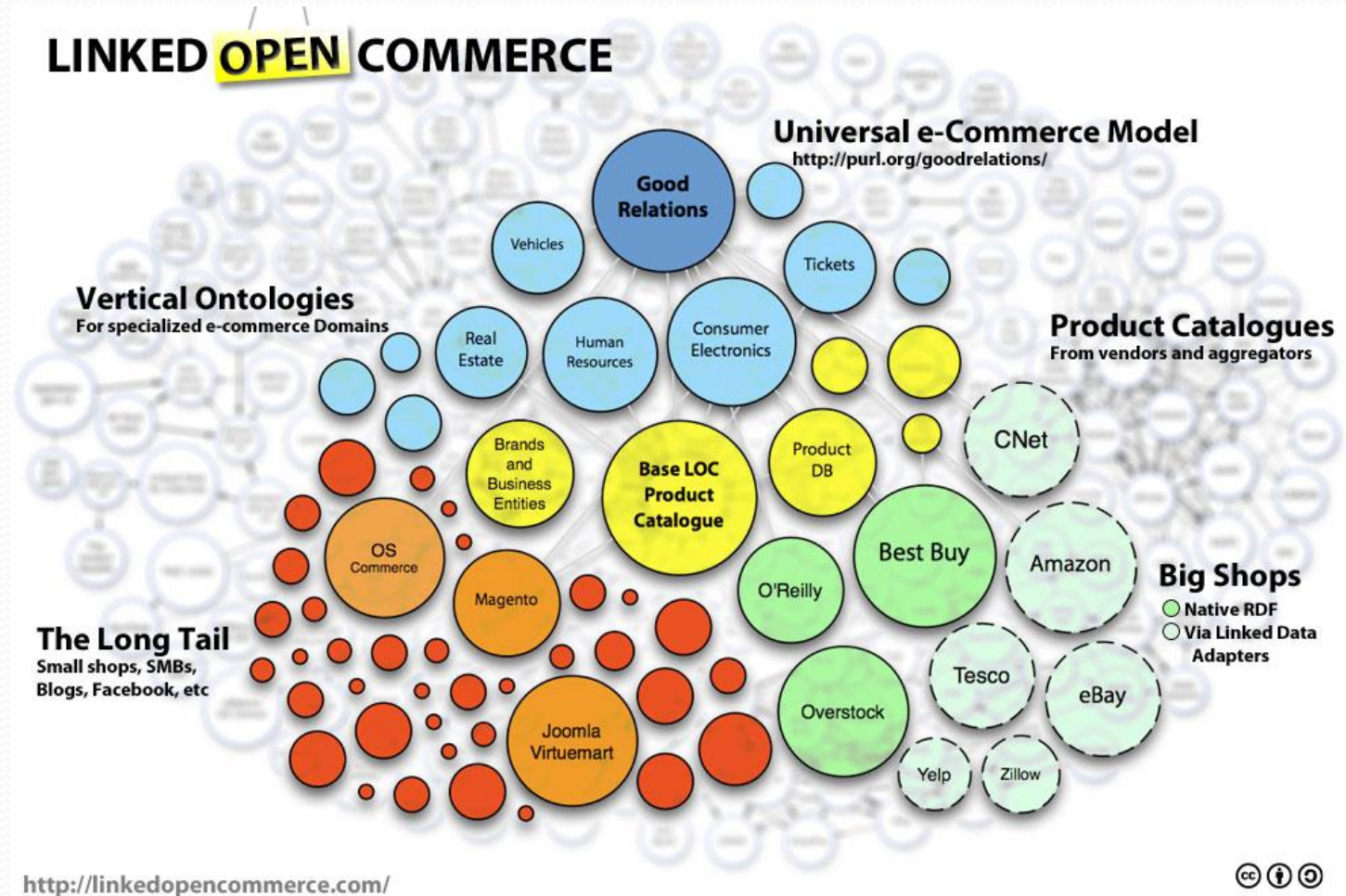
```
1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
5   xmlns="http://www.w3.org/2000/01/rdf-schema#"
6   xmlns:owl="http://www.w3.org/2002/07/owl#"
7   xml:base="http://synchronext.com/ahallili/synchroBotResource"
8   xmlns:sbo="http://synchronext.com/ahallili/synchroBotOntology#"
9
10  <sbo:Product rdf:about="http%3A%2F%2Fwww.ebay.com%2Fitm%2FHAND-CARVED-ONE-PIECE-WOOD-CHAIR-UNIQUE-1990-c-%2F181319358864%3Fpt%3DFolk_Art">
11    <sbo:title>HAND CARVED ONE PIECE WOOD CHAIR . UNIQUE 1990 c.</sbo:title>
12    <sbo:galleryURL>http%3A%2F%2Fthumbs1.ebaystatic.com%2Fm%2Fmf2ojYXqqiPL4NhCRcOMw8g%2F140.jpg</sbo:galleryURL>
13    <sbo:country>US</sbo:country>
14    <rdfs:label>Antiques</rdfs:label>
15    <sbo:price>15.0</sbo:price>
16    <sbo:locatedIn>Orwigsburg, PA, USA</sbo:locatedIn>
17    <sbo:itemId>181319358864</sbo:itemId>
18    <sbo:globalId>EBAY-US</sbo:globalId>
19    <sbo:soldBy >
20      <sbo:Seller rdf:ID="AzurMatos">
21        <sbo:mailBox>contact@azurmatos.fr</sbo:mailBox>
22        <sbo:address>152 Boulevard Cessole, 06100 Nice</sbo:address>
23      </sbo:Seller>
24    </sbo:soldBy>
25  </sbo:Product>
```



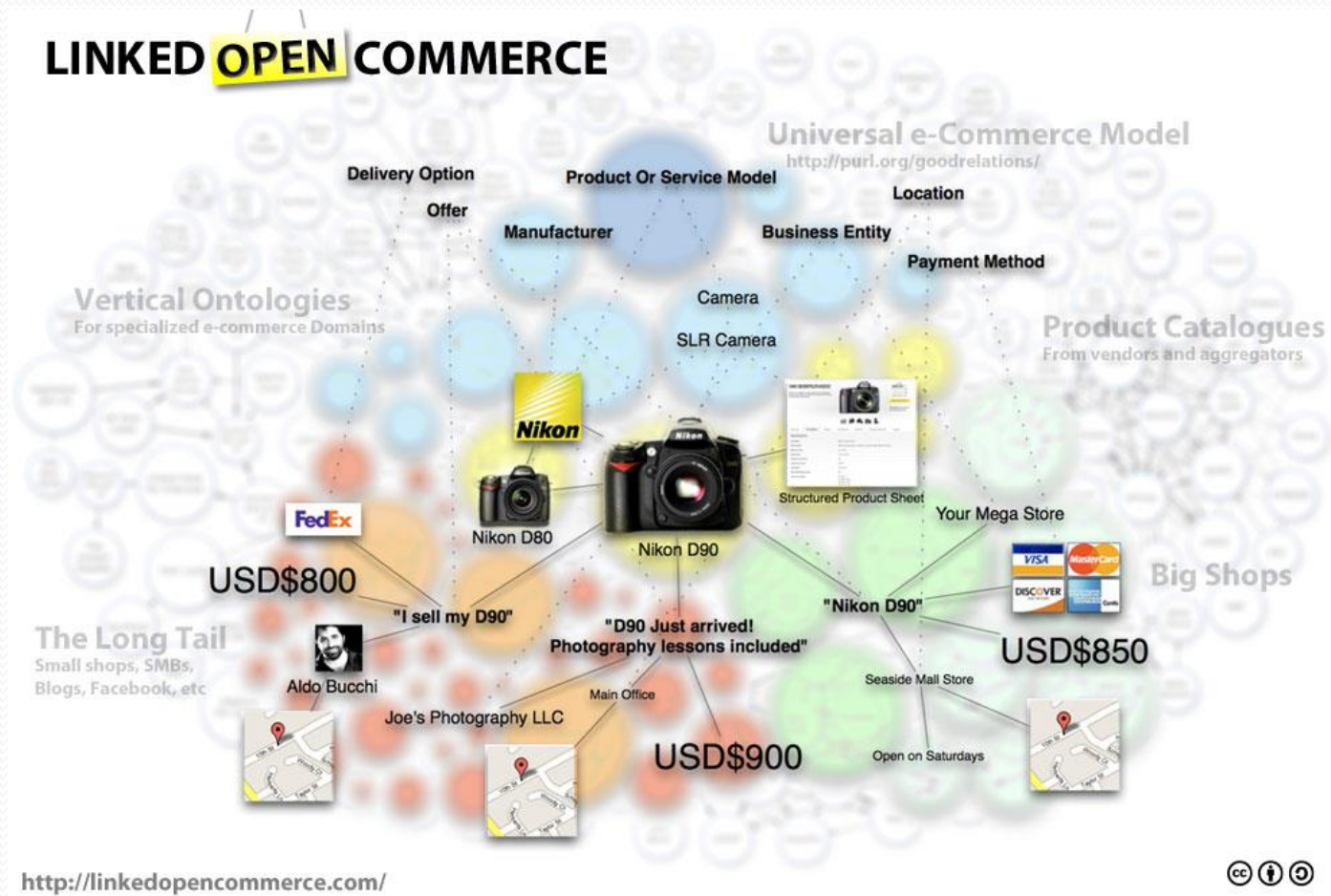
Ontology reuse

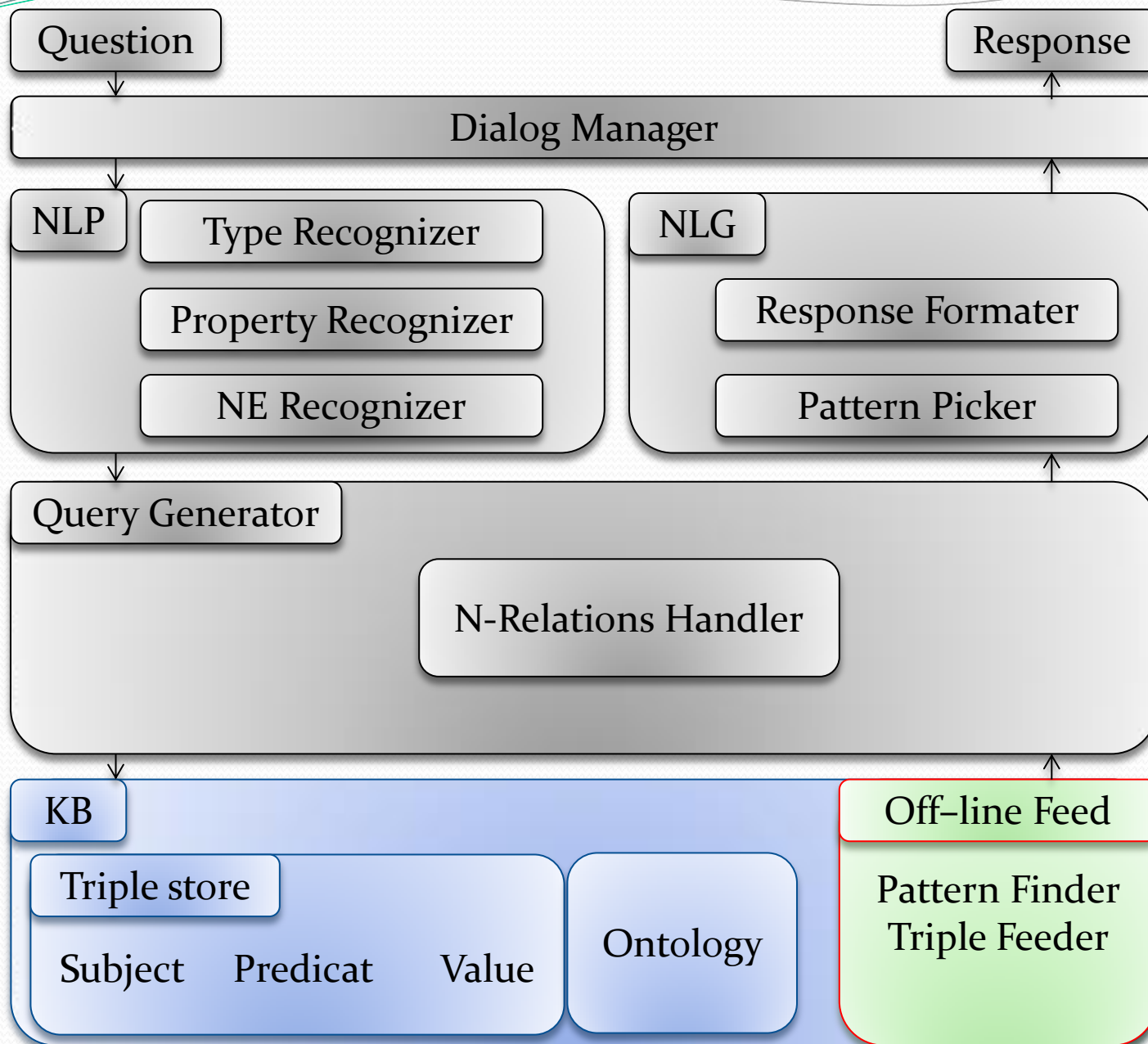
- Why we need an Ontology ?
Data structuration, Domain representation, Inference.
- Existing ontologies on commercial domain
 - Schema.org Ontology
 - Covers several domains
 - Used by state of the art search engines
 - Partial coverage of the commercial domain
 - GoodRelations Ontology (Hepp 2008)
 - Better coverage of the commercial domain

GoodRelations Ontology



GoodRelations Ontology





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

Pattern extraction – API method

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

Josh Miller/CNET

Subject

At the end of the day, the Galaxy flagship feels like it always has: like plastic. That's not necessarily a bad thing. As 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 room, I usually home to at least two phones and a jumble of weighty items, the phone stayed wedged in my hand during several-mile walks. I also had a death grip on it.)

<sch:hasDimension>

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 videos look especially lush.


<sch:hasDisplay>

Pattern extraction — Crawler Method

Properties

metadata

Sentences expressing properties



Search CNET

ReviewsNewsVideo

Starting at: **\$149.99**
★★★★★

Starting at: **\$129.99**
★★★★★

Compact
★★★★★

(Verizon Wireless)
Starting at: **\$149.99**
★★★★★

LOW PRICE:
\$189.99
Best Buy
\$749.99
Amazon.com
\$189.99
See all prices >

Here's why the Samsung Galaxy S5 should grab your attention: it looks good, it performs very well, and it has everything you need to become a fixture in nearly every aspect of your life. But, like a candidate running for reelection, the GS5 gets where it is today based on experience and wisdom, not on flashy features or massive innovation.

With the exception of a few nonessential hardware and software additions -- like the fingerprint scanner and novel heart-rate monitor -- and a few design tweaks, you're pretty much looking at the same phone Samsung released in 2013. The S5 is more of a Galaxy S4 Plus than it is a slam-the-brakes, next-generation device; it makes everything just a little smoother and faster.


The 5.1-inch, quad-core **Android 4.4 KitKat** machine with a terrific 16-megapixel camera is well worth snapping up, both on-contract for about \$200, or off-contract for about \$650 (and about £450 in the UK and a ~700). However, it isn't the only phone worth your time. The gorgeous, all-metal **HTC One M8** has a more sophisticated design, better speakers, and greater internal storage for about the same price (32GB versus 16GB). Should you buy the GS5? If you want to go to sleep at night certain that you own the most capable, robust phone, yes.

Jessica Dolcourt

Jessica Dolcourt reviews smartphones and cell phones, covers handset news, and pens the monthly column Smartphones Unlocked. A senior editor, she started at CNET in 2006 and spent four years reviewing mobile and desktop software before taking on devices.

EXTREME CLOSEUP: SAMSUNG'S GALAXY S5 (PICTURES)

SEE FULL GALLERY



1 - 4 of 19

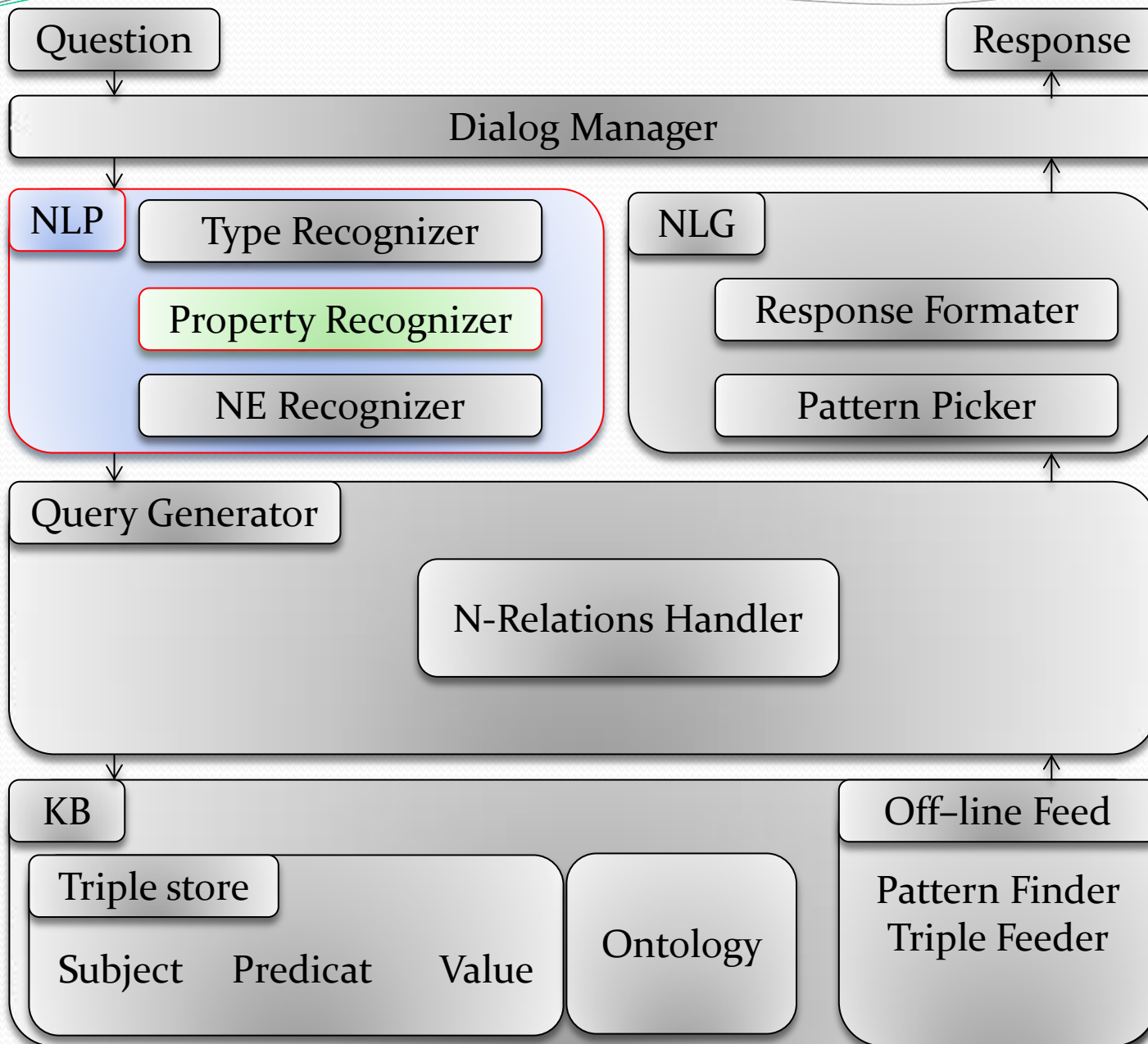
< PREV | NEXT >

Dimpled design an S4 redux

ElementsNetworkSourcesTimelineProfilesResourcesAuditsConsole

```
<meta property="og:site_name" content="CNET">
<meta property="og:title" content="Samsung Galaxy S5 review - CNET">
<meta property="og:description" content="Subtly improved and smartly refined, the Samsung Galaxy S5 is a superior superphone that hits every mark but the sharpest design.">
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/774ce9ab-7c96-40f1-a312-d35a1c1b0780/thumbna1/770x433/f5c687c7a139a5f4edc9520405399a3e/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/9e26712d-7997-450c-9a65-76927deb4f5b/thumbna1/770x433/0debceff672217eb043fbcacae2a34/samsung-galaxy
<meta property="og:image" content="http://cnet3.cbsistatic.com/hub/i/r/2014/04/07/1149dbc7-b3eb-4b70-ac9d-9c0c8b9c08f7/thumbna1/770x433/52a6cd7f5f487d1b4a0dbd2515b9c9ed/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/b1147740-aalb-42f8-8860-9a876bbe25b3/thumbna1/770x433/01d747c790e02e25b5f1e57471a55ebd/samsung-galaxy
<meta property="og:image" content="http://cnet3.cbsistatic.com/hub/i/r/2014/04/07/81252d4e-a9ba-4a27-ae51-f49775200e43/thumbna1/770x433/9e6125752111885af704f57fc8496713/samsung-galaxy
<meta property="og:image" content="http://cnet3.cbsistatic.com/hub/i/r/2014/04/07/7b8162b3-f060-47e2-8927-16098fcd54f0/thumbna1/770x433/b314d611aa06e96aea64e8e069bee0e1/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/e5911e1f-b8c9-458c-ac8b-993f5210497a/thumbna1/770x433/c195fd9e6454cdc23fd4fcca03db68/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/f0e77114-aa04-426f-a7c6-4968baf2ca68/thumbna1/770x433/13f4120f24f70c5e074a01208236856b/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/95c13f3a-b819-487c-92b7-874ba5e5b79e/thumbna1/770x433/a47f42677cc0f8c452a0599c34181c7d/samsung-galaxy
<meta property="og:image" content="http://cnet3.cbsistatic.com/hub/i/r/2014/04/07/292cac84-ebad-446f-a8d2-51f12ae770d6/thumbna1/770x433/81c34f88fe6b7672398297f9c199d659/samsung-galaxy
<meta property="og:image" content="http://cnet3.cbsistatic.com/hub/i/r/2014/04/07/0abab7d85-9c20-4961-93ed-eddd0747e38b/thumbna1/770x433/f8c7624d89e6b9fde0247e4059b8edae/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/08/ba79e002-e852-4415-8c9a-c854788b1b4a/thumbna1/770x433/ed778272c7ccea9a3330816cb6f4e78/samsung-galaxy
<meta property="og:image" content="http://cnet4.cbsistatic.com/hub/i/r/2014/04/07/1a6c6c0b-28b6-46a1-b9f1-16a79cb9facb/thumbna1/770x433/eb4d7e94d0a470b8808c511fb93f0829/samsung-galaxy
<meta property="og:image" content="http://cnet3.cbsistatic.com/hub/i/r/2014/04/07/8561f838-047c-4d7a-85de-87d053f80b7d/thumbna1/770x433/21a95114d0a0673166272ad3ad38ba0/samsung-galaxy"
```

25



Property Matching Module

<sbo:hasPrice>

[Product] price is [Double]

The price of [Product] is [Double]

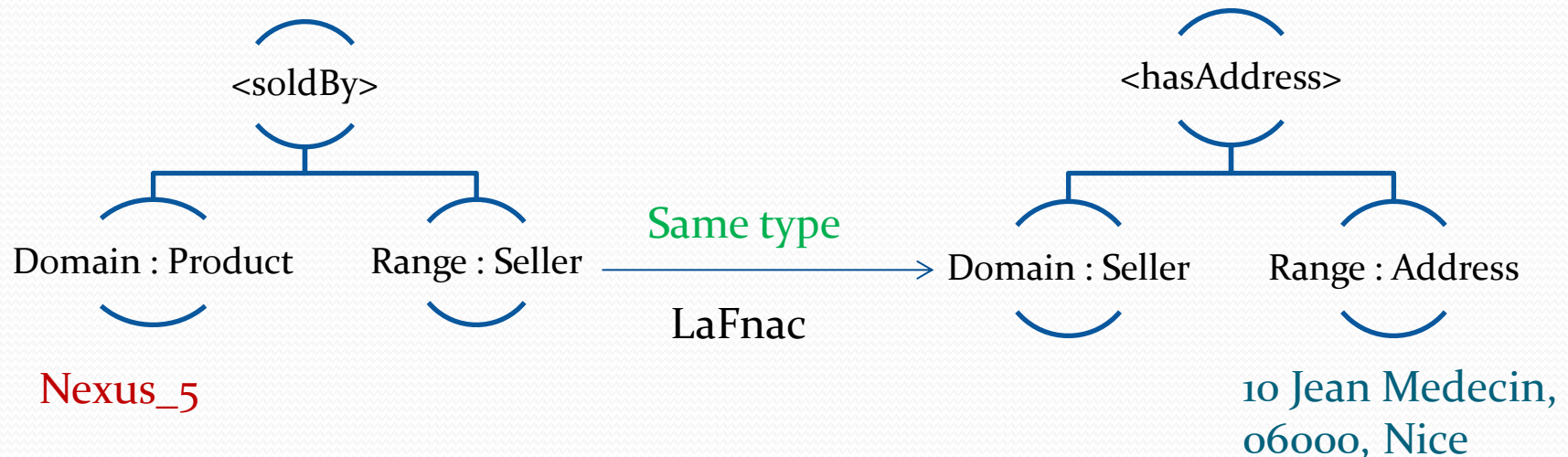
[Product] costs [Double]

High score

Give me the price of
a Nexus 5!

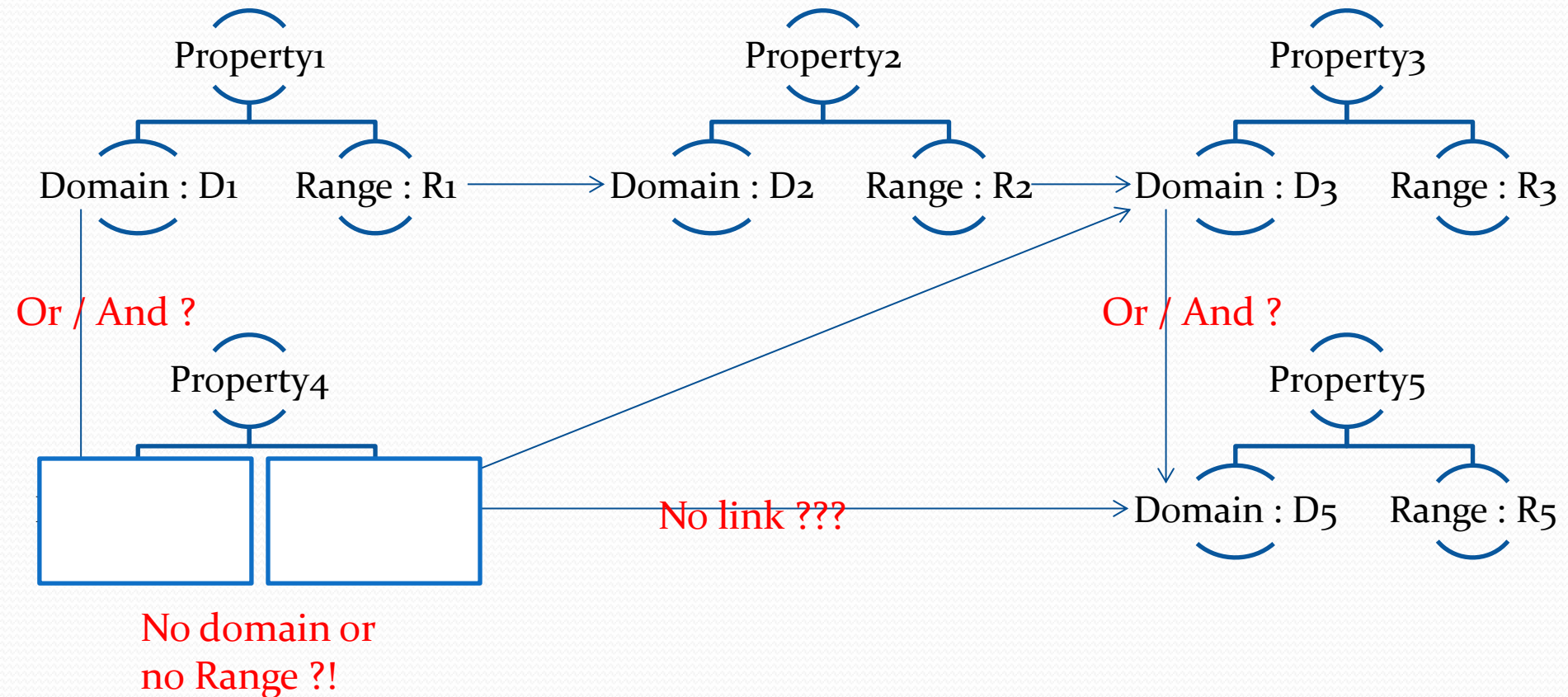
Property Matching (N-Relation)

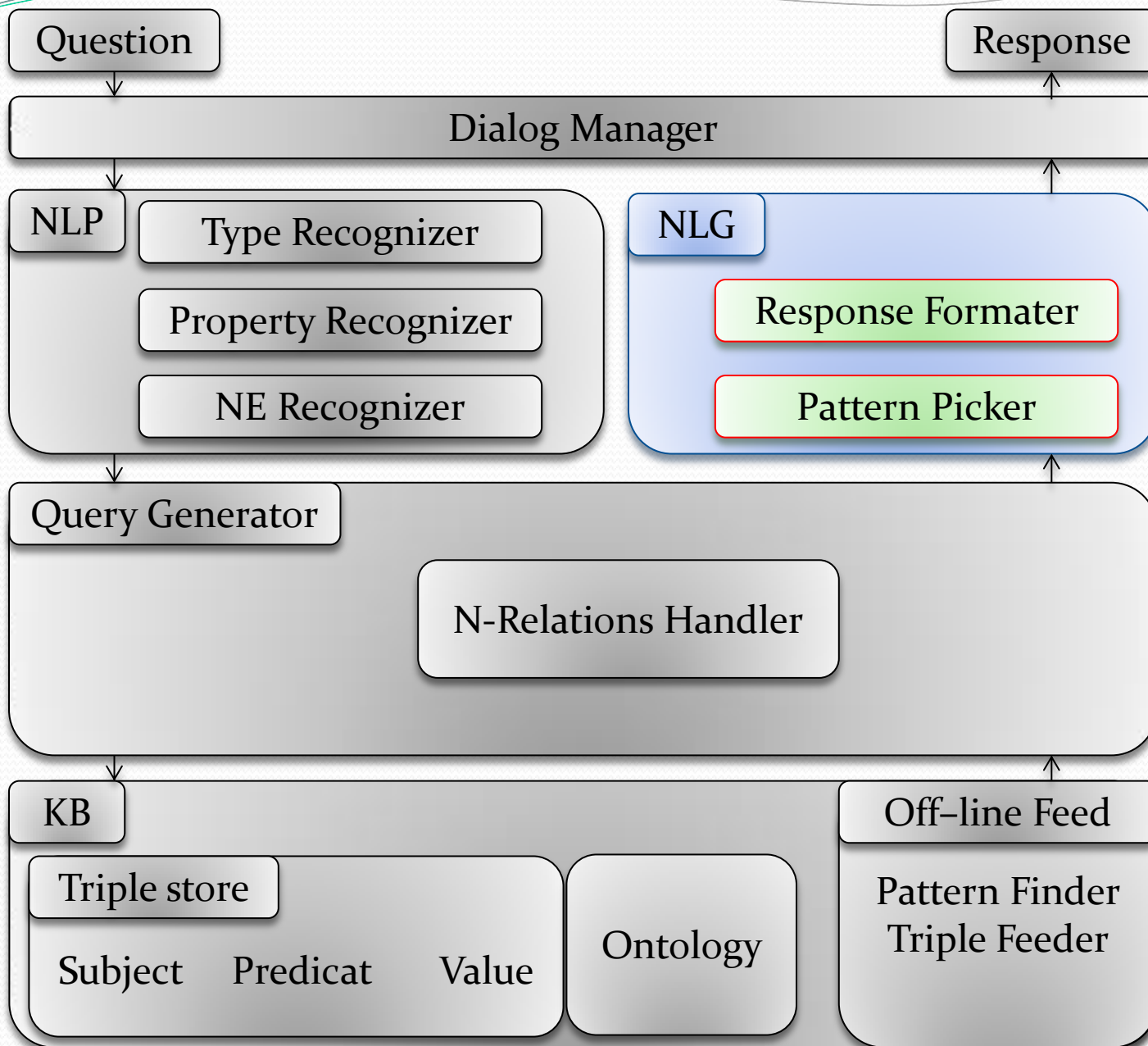
- 2-relations : Give me the address of **Nexus 5** seller !
 - Give me the Nexus 5 seller !
 - Give me his address ! => high score
 - NE : Nexus 5 => [Product]



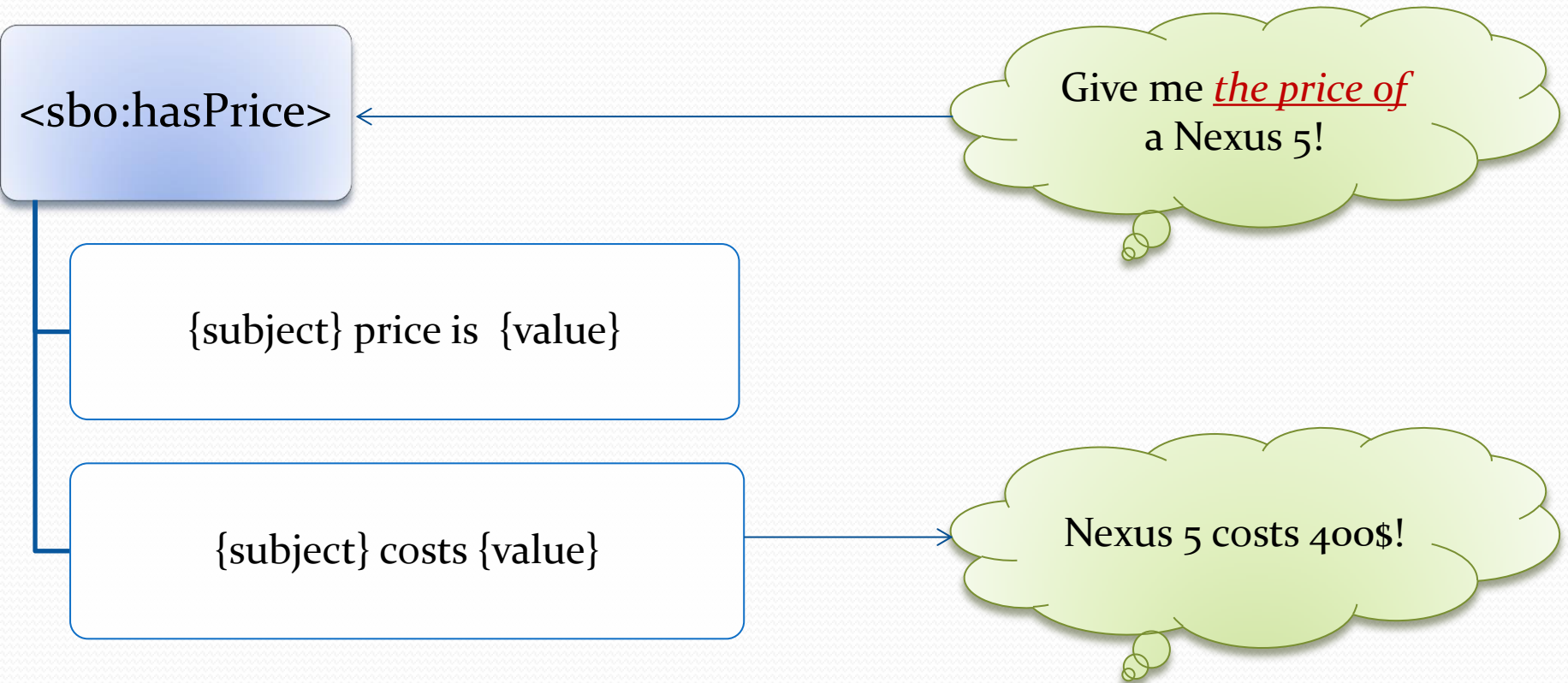
Property Matching (N-Relation)

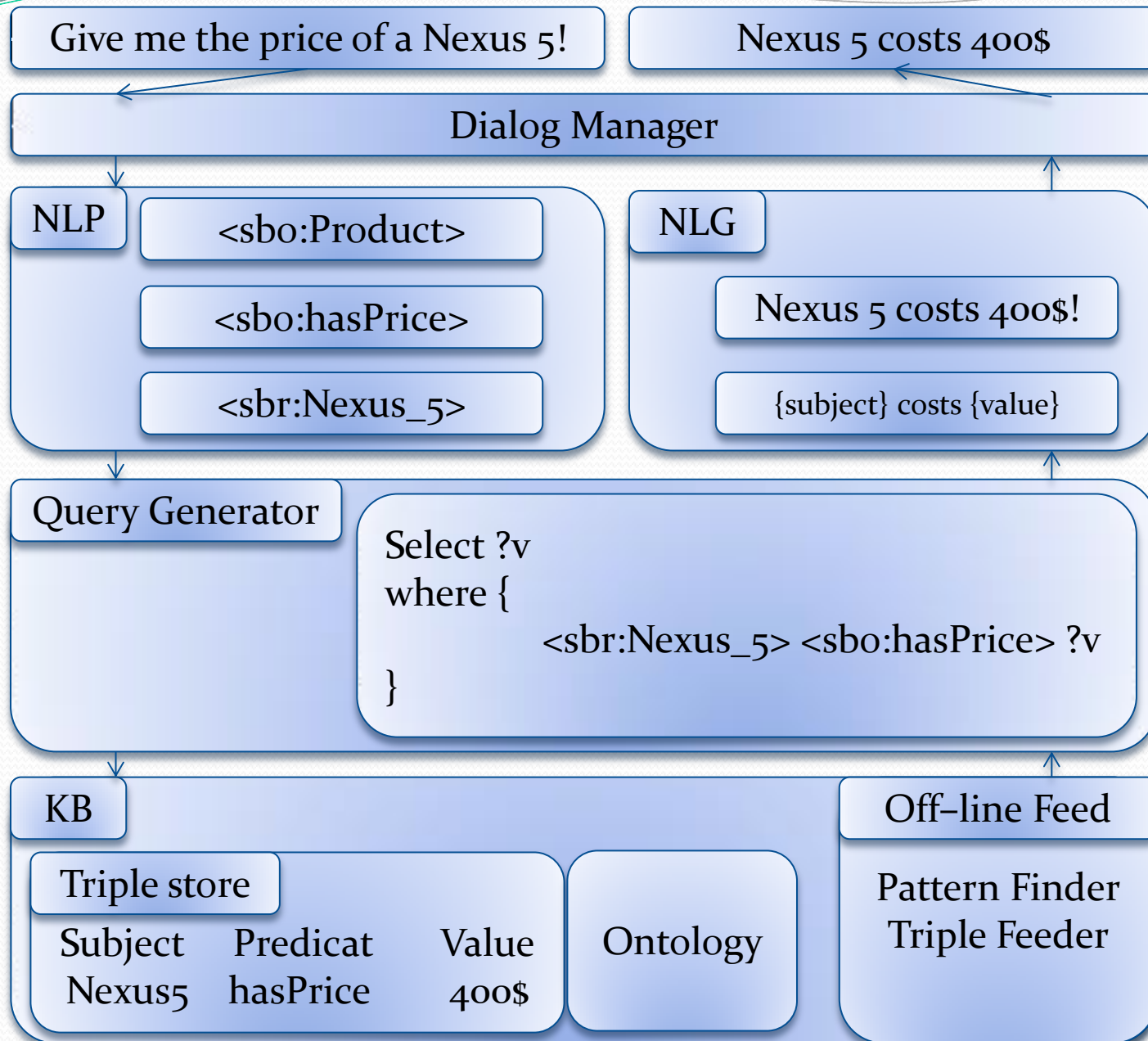
Graph representing the question :





NL Generation





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

- (Allen et al) J. F. Allen, D. K. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. Toward conversational human-computer interaction. *AI Magazine*, 22(4):2738, 2001.
- (Liao 2005) S.-H. Liao. Expert system methodologies and applications - a decade review from 1995 to 2004. *Expert Syst. Appl.*, 28(1):93-103, 2005.
- (Hirschman & al) L. Hirschman and R. J. Gaizauskas. Natural language question answering: the view from here. *Natural Language Engineering*, 7(4):275300, 2001.
- (Wooldridge 2009) M. J. Wooldridge. *An Introduction to MultiAgent Systems* (2. ed.). Wiley, 2009.
- (Cabrio & al 1) E. Cabrio, J. Cojan, A. P. Aproso, B. Magnini, A. Lavelli, and F. Gandon. Qakis: an open domain qa system based on relational patterns. In *International Semantic Web Conference (Posters & Demos)*, 2012.
- (Cabrio & al 2) E. Cabrio, J. Cojan, A. Palmero Aproso, and F. Gandon. Natural language interaction with the web of data by mining its textual side. *Intelligenza Artificiale*, 6(2):121-133, 2012.
- (Augello & al 1) A. Augello, G. Pilato, G. Vassallo, and S. Gaglio. A semantic layer on semi-structured data sources for intuitive chatbots. In *CISIS*, pages 760-765, 2009.
- (Augello & al 2) A. Augello, G. Pilato, A. Mach, and S. Gaglio. An approach to enhance chatbot semantic power and maintainability: Experiences within the frasi project. In *ICSC*, pages 186-193. IEEE Computer Society, 2012.
- (Hepp 2008) M. Hepp. Goodrelations: An ontology for describing products and services offers on the web. In *EKAU*, pages 329-346, 2008.