the Joke Ontology and Web Application

January 22, 2015

Falitokiniaina Rebearison
Marcello Benedetti

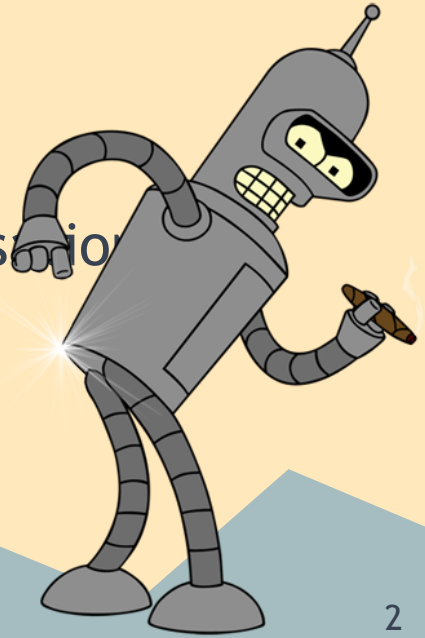
Introduction

So many ontologies have been written about movies... we propose something new, a Joke ontology!

We imagine a time when strong Artificial

Intelligence comes and we'll be able to have conversations with machines…

they need to have sense of humor!



Jokes aside...

...we followed *Ontology 101*

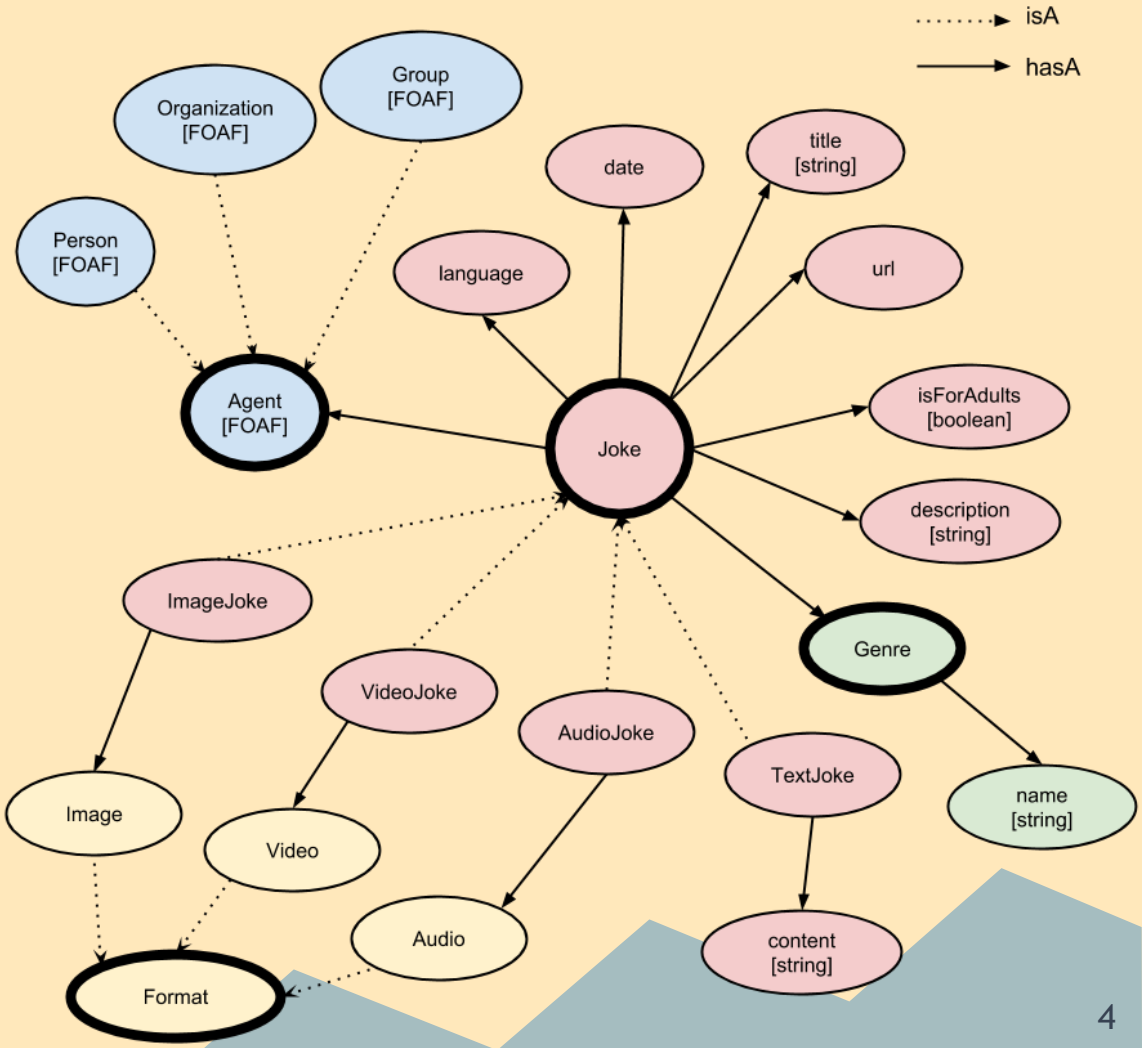
1. determine the scope: *Jokes*
2. consider reuse: *FOAF used for authors*
3. enumerate terms: *joke, genre, language, media…*
4. define classes
5. define properties
6. define constraints

} *see next slide*

7. create instances: *from Comedy Central's web site*

The ontology

- Color of the bubbles identifies different reusable domains
- Bold bubbles are the main classes
- Arrow type identify the type of relation



Architecture

- RDF/OWL design with *Protégé*
- web crawling in Java *crawler4j*
- dynamic pages:
  - server side with Python *NLTK*, *cherrypy*, *rdflib*
  - client side with Javascript *jQuery*
  - cross-domain compatibility with PHP
- user interaction via web browser

How does it work

1. The user inputs a topic in the web interface
2. An appropriate genre of joke is selected
3. A random joke instance is returned from that genre
4. BONUS: if you're using *Chrome* a nice voice will read the joke

Interpretation of user's request

- user's input and genre names are assumed to be nouns
- shortest paths in the WordNet taxonomy are computed
- most similar genre is selected

Example:

```
input = 'spaghetti'
genres = ['doctor', 'food', 'school']
sim('spaghetti', 'doctor') = 0.06
sim('spaghetti', 'food') = 0.2
sim('spaghetti', school) = 0.0625
```

Return a food joke!

http://www.kawaiikitchen.com/jokes.htm

Result

Not even close to an Artificial Intelligence with sense of humor but…

...let's see a demonstration!

http://127.0.0.1:8080/generate

Current limitations and TODO 1

Requests assumed to be nouns → needs Part-of-Speech tagging

Content of the jokes irrelevant → but including an inverted index would move everything towards Information Retrieval and away from Semantic Web

Inference mechanism not exploited → needs a good reason to invest time here

Current limitations and TODO 2

Some jokes are not suitable for the youngest. The ontology already provides the *isForAdult* property, but we lack an automatic detection of explicit content.

Comedy Central is a great source of English textual jokes. The ontology provides classes for multimedia types and languages as well. Need to extend with video and image instances from *9gag*.