



Give your Ontology The Cognitive Diet It Deserves

Author: Ramona Pauna, PhD Linguistics, Director Customer Success

Presenter: Ian Cameron. VP Digital Media

Session Topics

- Building taxonomies and ontology has been for years a current practice for **data modelling** to power **information applications**
- For an ontology, the biggest focus can be on creating its **data schema** (classes, attributes of classes and relationships between the classes) and then on **manually populating** it with **instances** of classes, **properties** and **relationships**
- Using a cognitive technology to **automate** the feeding of the ontology with **facts** derived from the **content** can help address this challenge

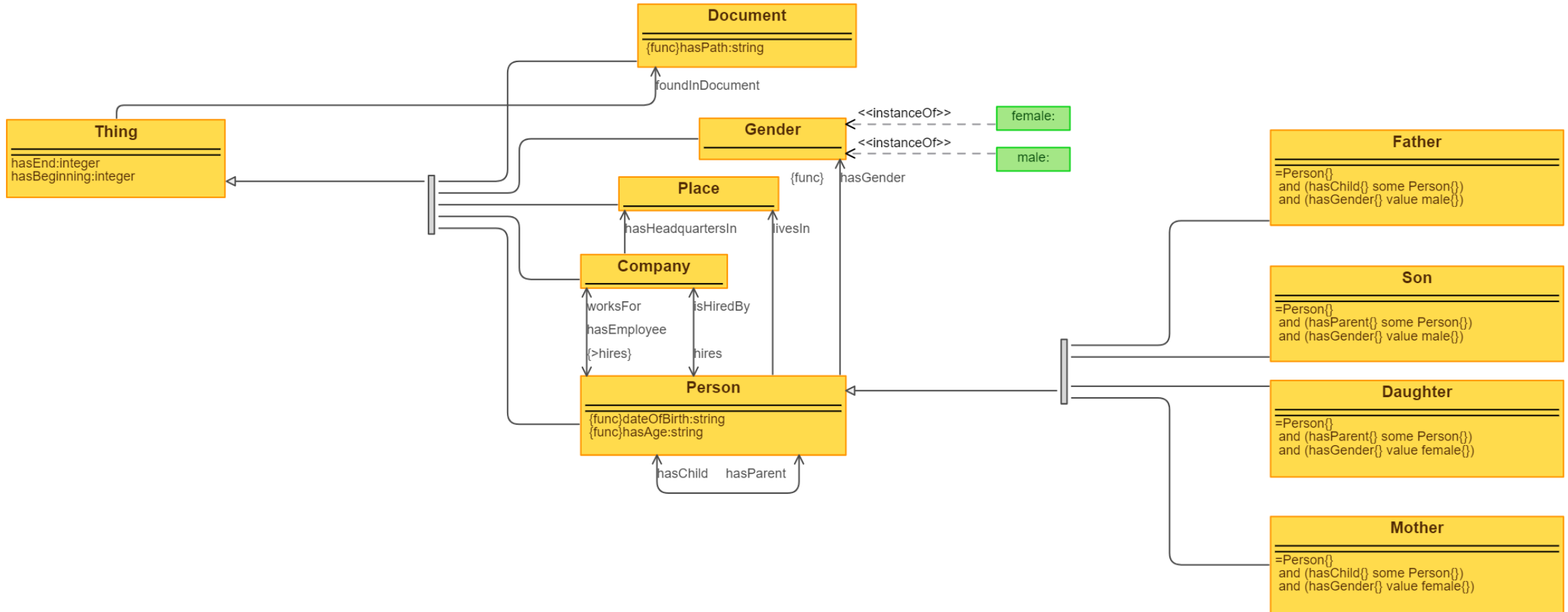
Ontologies, Taxonomies and Knowledge Bases

Some definitions

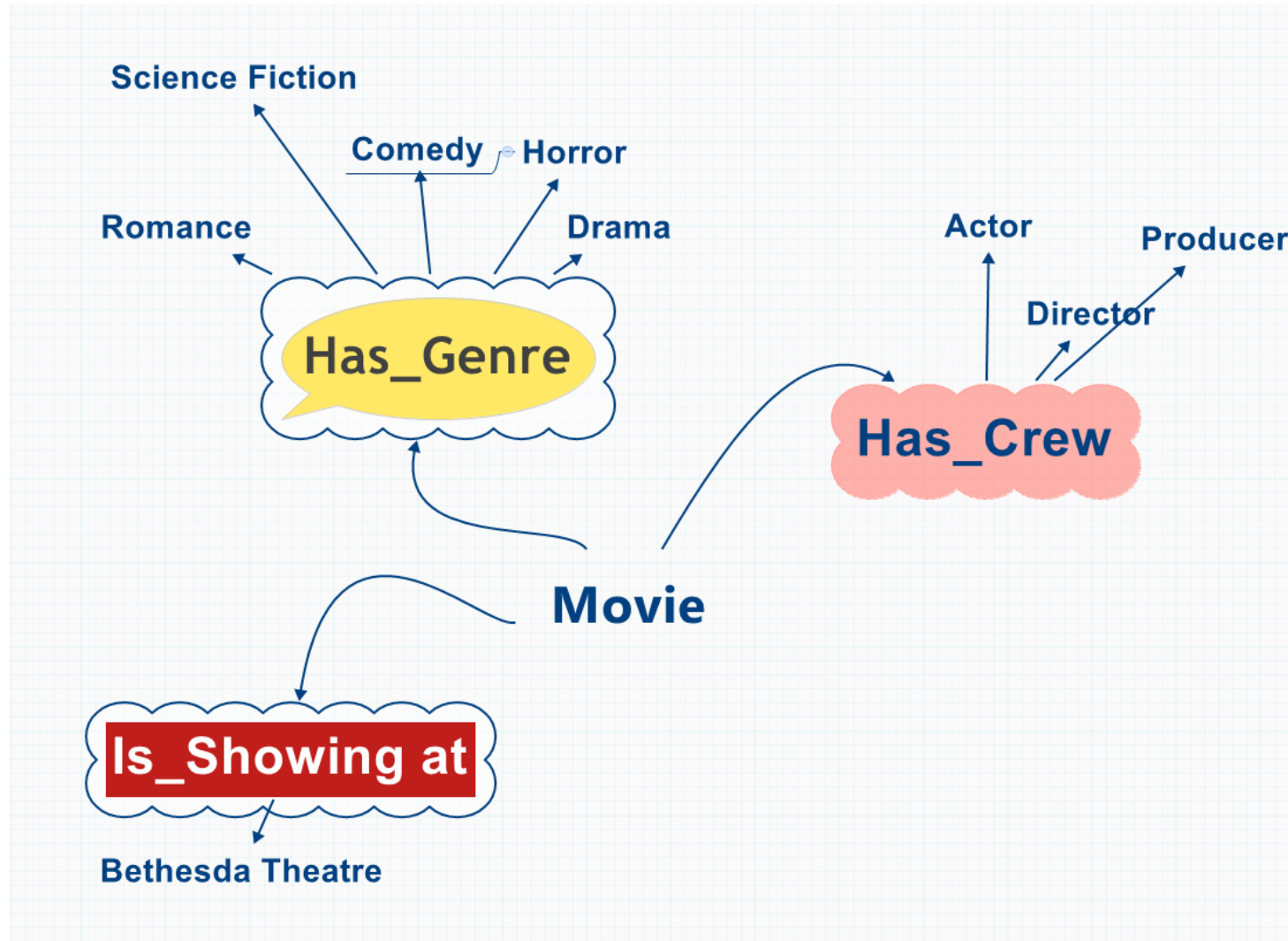
- **Ontology** : a conceptual model representing a specific field of information
 - Classes (e.g. “person”, “organization”, “dog”)
 - Attributes of classes (“age”, “birth date”, “foundation date”)
 - Relationships between classes (“marry”, “own”, “hire”)
- **Taxonomy** : is the science of classification according to a predetermined system
 - Taxonomy is the process of naming and classifying things such as animals and plants into groups within a larger system, according to their similarities and differences.
- **Knowledge Base (KB)** : contains the ontology model plus the instances of
 - Classes (“John Smith”, “Expert System”, “Buddy”),
 - Properties (“John Smith is 25 years old”)
 - Relationships (“Expert System hired John Smith”)
 - Adding instances to an ontology model = “populating” the ontology
- Ontologies and KBs use RDF/RDFS as the representation language
- OWL enables more complex definitions of classes and relationships
- An OWL ontology can include axioms to leverage automated reasoning over the model

Your Ontology before its diet

Empty ontology model



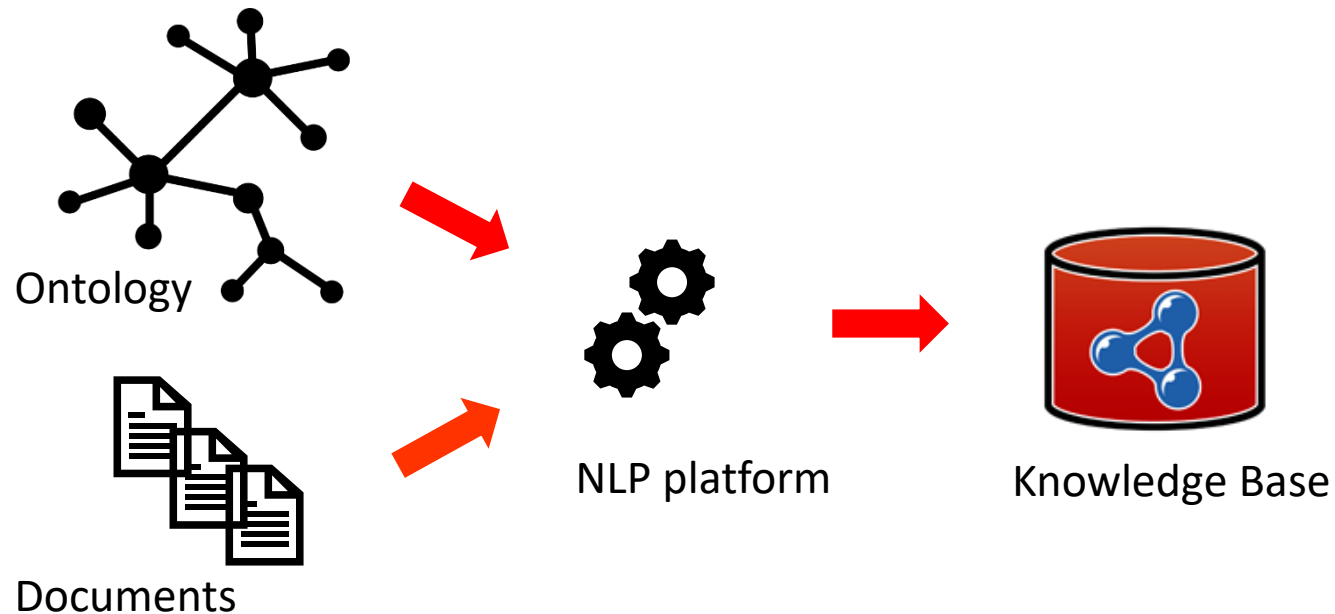
Your Movie Ontology before its diet



Populating an Ontology

A challenging task

- Transforming an ontology into a rich and accurate KB is quite challenging
 - Requires access to (and validation of) a large amount of information
 - NLP helps automate this process, extracting valuable information from documents and framing it in an ontology model



The Diet for your Ontology is NLP

What features?

- <https://www.youtube.com/watch?v=Kzt7pd-YfMU>
- Part of Speech recognition
- Tokenization, Lemmatization, Morphologic Analysis
- Syntactic Processing
- Full Sentence and inter-sentence parsing
- **Co-reference and anaphora resolution**
- **Entity extraction and disambiguation**
- **Entity linking using Linked Open Data**
- **Rule-based extraction and categorization**
- **Machine learning**
- **Ontology and taxonomy management**
- **Automatic enhancement for ontology and taxonomy**

The Diet for your Ontology is NLP

Disambiguation based on understanding the context

Which “gas” means “gasoline” ?

My jaguar eats gas when I step on the gas.

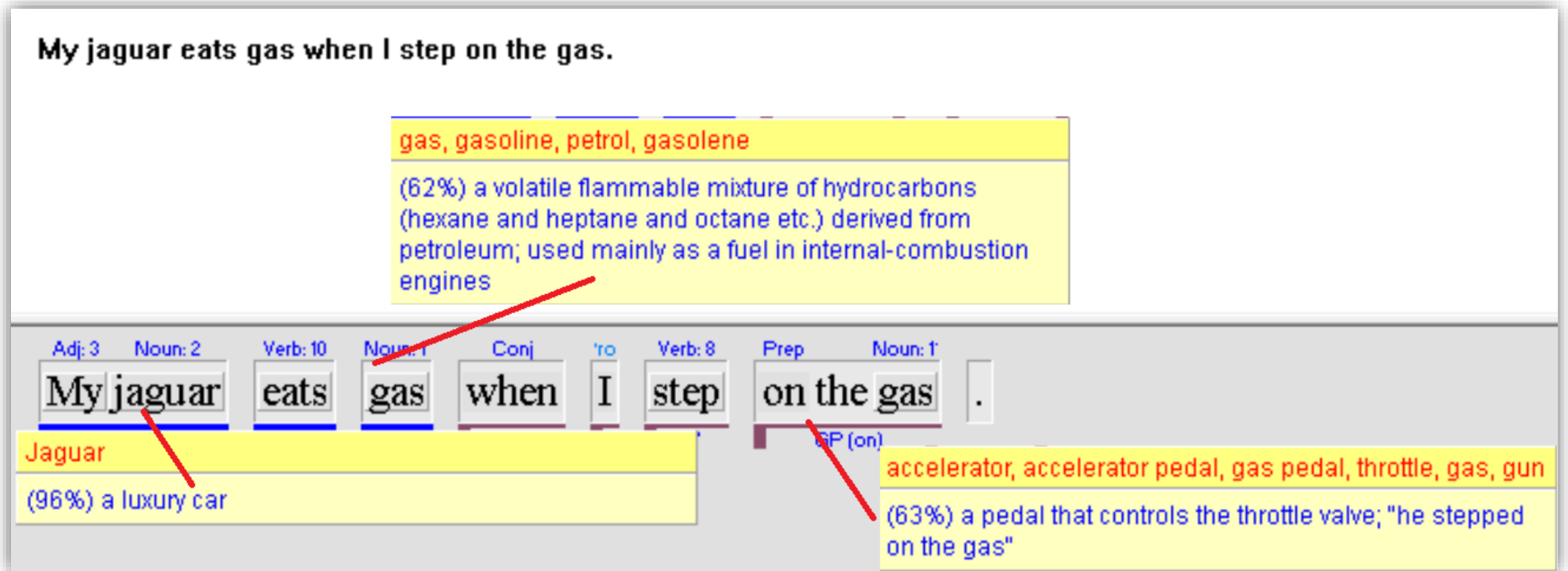
gas, gasoline, petrol, gasolene
(62%) a volatile flammable mixture of hydrocarbons (hexane and heptane and octane etc.) derived from petroleum; used mainly as a fuel in internal-combustion engines

Adj: 3 Noun: 2 Verb: 10 Noun: 1 Conj to Verb: 8 Prep Noun: 1

My jaguar eats gas when I step on the gas .

Jaguar
(96%) a luxury car

GP (on)
accelerator, accelerator pedal, gas pedal, throttle, gas, gun
(63%) a pedal that controls the throttle valve; "he stepped on the gas"



The Diet for your Ontology is NLP

Disambiguation based on context and external properties (LODs)

Which “Paul Smith” am I talking about ?

The screenshot displays the COGITO NLP interface. On the left, a central node 'Paul Smith' is connected to 'Herman Miller' (via 'brand'), 'Maharam' (via 'furniture'), and 'collection' (via 'stripe'). Below this are sections for 'Main sentences', 'Search concept', 'Main topics' (with a bar chart for fabric, architecture, textile industry, and fashion), and 'Sentiment' (with a color scale). The main area shows a list of text snippets with a pop-up window for 'Paul Smith' (Human proper noun) featuring a photo and a description: 'Sir Paul Smith, CBE, RDI, (born 5 July 1946) is a British fashion designer and businessman, whose business and reputation is founded upon his men's fashion with a special focus on tailoring. He is kno...'. The right side of the interface shows a list of text snippets related to the search.

The Diet for your Ontology is NLP

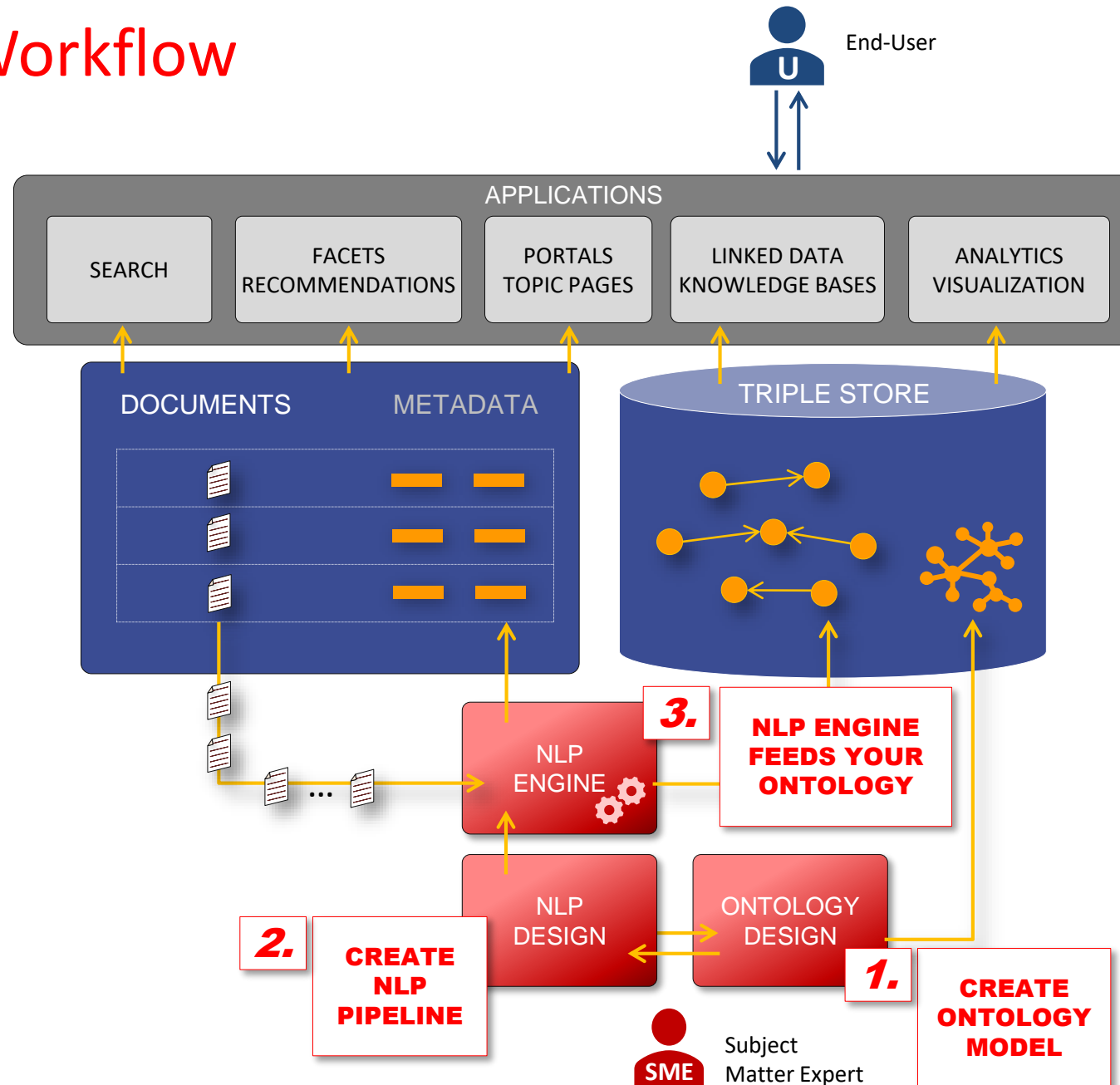
Disambiguation based on context and external properties

Which “Paul Smith” am I talking about ?

The screenshot displays the COGITO NLP interface. On the left, a central node 'Paul Smith' is connected to several related nodes: 'pianist', 'Plays Steve Allen', 'Ziggy Elman', 'Annette Warren', and 'Ella Fitzgerald'. Below this, there are sections for 'Main sentences', 'Search concept', 'Main topics' (with sliders for 'music' and 'singing'), and 'Sentiment' (with a color-coded scale).

On the right, a 'Text' pane shows a snippet of text about Paul Thatcher Smith, with a tooltip for the 'Human proper noun' entity. The tooltip lists 'Paul Thatcher Smith, Smith, Thatcher Smith' and provides a description: 'Paul Thatcher Smith (April 17, 1922 - June 29, 2013) was an American jazz pianist. He performed in various genres of jazz, most typically bebop, but is best known as an accompanist of singers, especially Ella Fitzgerald.' The text in the pane also includes a sentence: 'Paul Thatcher Smith, pianist: born San Diego 17 April 1922; married Annette Warren (two sons, one daughter); died Torrance, California 29 June 2013.'

High-Level Workflow



Preparing Your Ontology's Diet

Ingredients for your ontology

- NLP that enables advanced cognitive rule writing
 - Supports effective extraction of class instances, properties & relationships

The screenshot displays the Expert System interface. On the left, a 'Template' list includes 'PERSON_CHILD_OF_PERSON' with sub-templates 'NAME_OF_CHILD', 'NAME_OF_PARENT', and 'RELATE', each with a count of 10. The main text area contains a paragraph about Google and Mike Holson, with the sentence 'Mike is the child of David Holson' highlighted in green. Below this, a 'SCOPE SENTENCE' box shows a rule definition for 'PERSON_CHILD_OF_PERSON' using NLP syntax. At the bottom, an 'Extraction' table shows the results for the 'PERSON_CHILD_OF_PERSON' rule, listing 'NAME_OF_CHILD' as 'Mike Holson', 'NAME_OF_PARENT' as 'David Holson', and 'RELATE' as 'child'.

Template	Type	Rules
COMPANY_EMPLOYS_PERSON		
NAME_OF_COMPANY		2
NAME_OF_PERSON		2
RELATE		2
COMPANY_LOCATED_PLACE		
NAME_OF_COMPANY		2
PLACE		2
RELATE		2
COMPANY_NAME		
PERSON_AGE		
PERSON_CHILD_OF_PERSON		
NAME_OF_CHILD		10
NAME_OF_PARENT		10
RELATE		10
PERSON_DOB		
PERSON_GENDER		
PERSON_LIVES_IN_PLACE		
PERSON_NAME		
PLACE		

Google Inc. just fired Mike Holson without any notice. Since our founding in 1998, Google has grown by leaps and bounds. Google is located in Mountain View, California. Mike Holson was hired by IBM in 2015. Mike is the child of David Holson. Mike was born in 9/12/1981. From offering search in a single language we now offer dozens of products and services - including various forms of advertising and web applications for all kinds of tasks - in scores of languages. And starting from two computer science students in a university dorm room, we now have thousands of employees and offices around the world. A lot has changed since the first Google search engine appeared. IBM's headquarters are in NYC. David Holson is a very rich man, mentioned in the list of Forbes 2014.

```
SCOPE SENTENCE
{
  IDENTIFY (PERSON_CHILD_OF_PERSON)
  {
    @NAME_OF_CHILD[TYPE (NPH)] // Person ; human being; individual
    &SV
    LEMMA ("be")
    <0:3>
    @RELATE[KEYWORD ("son", "daughter", "child")]
    <0:2>
    @NAME_OF_PARENT[TYPE (NPH)] // Person ; human being; individual
  }
}
```

Record / File	Author ID
COMPANY_NAME	
PERSON_CHILD_OF_PERSON	
NAME_OF_CHILD	Mike Holson
NAME_OF_PARENT	David Holson
RELATE	child

Start feeding Your Ontology

The diet turns the ontology in a Knowledge Base

Mike Holson is 30 years old and lives in New York. He's the child of David Holson, founder of a real estate company. Mike was born on September 12, 1981.

```
SCOPE SENTENCE
{
  IDENTIFY (PERSON_CHILD_OF_PERSON)
  {
    @NAME_OF_CHILD [TYPE (NPH)] // Person ; human being; individual
    &SV
    LEMMA ("be")
    <0:3>
    @RELATE [KEYWORD ("son", "daughter", "child")]
    <0:2>
    @NAME_OF_PARENT [TYPE (NPH)] // Person ; human being; individual
  }
}
```

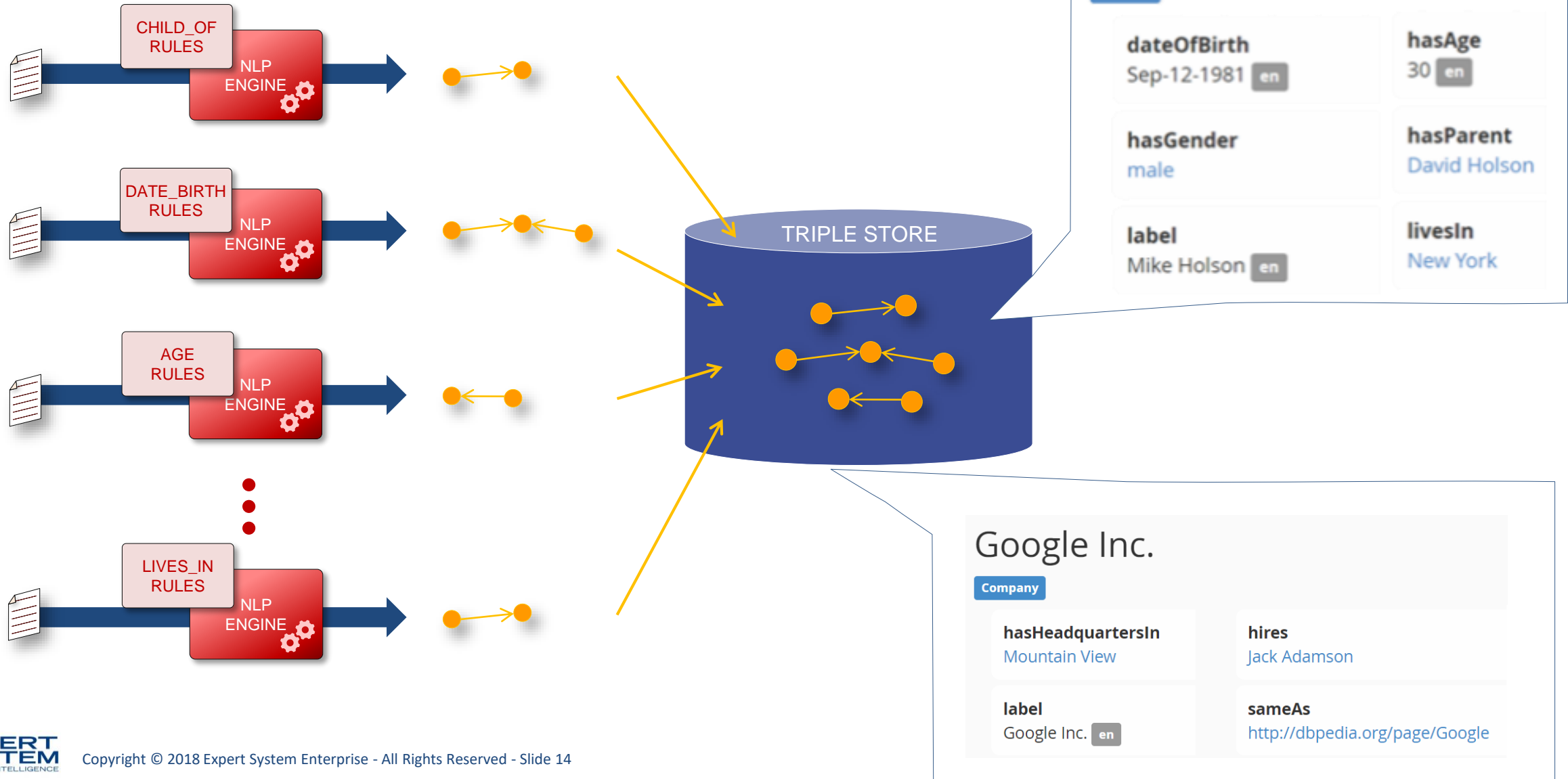
PERSON_CHILD_OF_PERSON	
NAME_OF_CHILD	Mike Holson
NAME_OF_PARENT	David Holson
RELATE	child

Mike Holson

Person

dateOfBirth	Sep-12-1981 en	hasAge	30 en
hasGender	male	hasParent	David Holson
label	Mike Holson en	livesIn	New York

Ontology becomes a Knowledge Base



It's Served Hot !

Well-fed ontology leverages information applications

- On top of the Triple Store, an information application allows end users to interact with the well-fed ontology and content by asking questions in natural language
- A Natural Language-to-SPARQL module translates the query to SPARQL and runs the query against the Triple store endpoint, retrieving the results in RDF
- The answer to the question is displayed to the end user

In which city are the headquarters of the company which hired David Holson's son?

Ask

SPARQL Query:

```
SELECT distinct ?statement ?s ?p ?o ?label ?toNext
WHERE { ?statement rdf:predicate :hasParent . ?statement rdf:subject :David_Holson . ?statement rdf:predicate ?p . ?statement rdf:subject ?s . ?statement rdf:object ?toNext . ?o rdfs:label ?label . ?statement rdf:object ?o . }
SELECT ?statement ?s ?p ?o ?toNext ?label
WHERE { ?company :hires <...> . ?statement rdf:subject ?company . ?statement rdf:predicate :hires . ?statement rdf:object <...> . ?statement rdf:subject ?s . ?statement rdf:predicate ?p . ?statement rdf:object ?o . ?statement rdf:subject ?toNext . ?company rdfs:label ?label . }
SELECT ?statement ?s ?p ?o ?label
WHERE { <...> :hasHeadquartersIn ?city . ?statement rdf:subject <...> . ?statement rdf:predicate :hasHeadquartersIn . ?statement rdf:object ?city . ?statement rdf:subject ?s . ?statement rdf:predicate ?p . ?statement rdf:object ?o . ?city rdfs:label ?label . }
```

It's Served Hot !

Well-fed ontology leverages information applications

Query result:

[New York](#)

Show / Hide documents

Label: Mike Holson

Google Inc. just fired Mike Holson without any notice. Since our founding in 1998, Google has grown by leaps and bounds. Google is located in Mountain View, California. Mike Holson was hired by IBM in 2015. Mike is the child of David Holson. Mike was born in 9/12/1981. From offering search in a single language we now offer dozens of products and services including various forms of advertising and web applications for all kinds of tasks in scores of languages. And starting from two computer science students in a university dorm room, we now have thousands of employees and offices around the world. A lot has changed since the first Google search engine

Label: IBM

Google Inc. just fired Mike Holson without any notice. Since our founding in 1998, Google has grown by leaps and bounds. Google is located in Mountain View, California. Mike Holson was hired by IBM in 2015. Mike is the child of David Holson. Mike was born in 9/12/1981. From offering search in a single language we now offer dozens of products and services including various forms of advertising and web applications for all kinds of tasks in scores of languages. And starting from two computer science students in a university dorm room, we now have thousands of employees and offices around the world. A lot has changed since the first Google search engine

Label: New York

Mike Holson was hired by IBM in 2015. Mike is the child of David Holson. Mike was born in 9/12/1981. From offering search in a single language we now offer dozens of products and services including various forms of advertising and web applications for all kinds of tasks in scores of languages. And starting from two computer science students in a university dorm room, we now have thousands of employees and offices around the world. A lot has changed since the first Google search engine appeared. IBM's headquarters are in NYC. David Holson is a very rich man, mentioned in the list of Forbes 2014.

Wrap-up

- Feeding your ontology w/ instances & relationships **is essential**
- AI/Cognitive technology leverages semantic reasoning to :
 - **Automatically feed** your ontology based on your content
 - **Resolve ambiguity** for a comprehensive view of each entity
 - Maintain **up-to date knowledge**
- The result ? A **well-fed ontology**
 - Provides end-users with **relevant insights based on semantic triples extracted from your content**



- Largest pure-play vendor of **AI/Cognitive Computing & Text Analytics** solutions
- 7 Countries / 250+ employees
- **COGITO**
Award-winning, patented technology
Text Analytics/NLP & Ontology
Management

Gartner



- Diversified Business
 - Publishing & Media
 - Finance & Insurance
 - Life Sciences
 - Federal



References





Thank You

Ramona Pauna - rpauna@expertsystem.com – 646- 276-2021

Ian Cameron – icameron@expertsystem.com – 443-979-3860