

Técnicas para Big Data

Clase 06 - Web Semántica

Outline

Introducción a Bases de Datos de Grafos

Web Semántica

SPARQL

NoSQL

NoSQL

- Menos restricciones que un esquema relacional
- “Schemaless”
- Facilidad para distribuir

NoSQL

Tipos de NoSQL

NoSQL

Tipos de NoSQL

- Bases de datos de Documentos
- Bases de datos Key - Value
- Bases de datos de Grafos
- Bases de datos orientadas a columnas

Bases de Datos de Grafos

Bases de Datos de Grafos



Bases de Datos de Grafos



Bases de Datos de Grafos



Bases de Datos de Grafos

Casos de Uso

Bases de Datos de Grafos

Casos de Uso

- Redes sociales
- Provenance
- Algoritmos de rutas
- Muchos otros...

Bases de Datos de Grafos

Modelo

Bases de Datos de Grafos

Modelo

- Conjunto de Nodos V
- Conjunto de aristas $E \subseteq V \times V$
- Nodos y aristas pueden tener atributos
- Consultas en base a patrones

Bases de Datos de Grafos

Modelo

Si se puede modelar con una base de datos relacional, ¿por qué existen las bases de datos de grafos?

Bases de Datos de Grafos

Modelo

- Están orientadas a encontrar relaciones rápidamente.
- Bases de datos relacionales son muy generales, las bases de datos de grafos son un subconjunto “optimizado”

Grafos vs JSON / XML

JSON / XML	Grafos
Información jerárquica	Información sin jerarquía
Lectura humana y por computador	Principalmente lectura por computador
Comunicación entre aplicaciones	Pensada para consultas de relaciones
Acoplamiento a la Web y HTML	Se puede modelar sobre BD relacional

Outline

NoSQL

Web Semántica

SPARQL

Web Semántica

Web Semántica

¿Se puede consultar la Web como una base de datos? ¿De qué forma sería posible?

Web Semántica

- ¿Por qué consultar automáticamente la Web?
- ¿Qué es lo que hace Google?

Web Semántica

- Existen documentos legibles por humanos:
 - XML
 - HTML
- Queremos documentos legibles por computadores

Web Semántica

```
<html>
  <body>
    <h1>My webpage!</h1>
    <p>Hi! My name is Raymundo Lizama and this is
      my webpage.</p>
    <p>If you want to contact me,
      <a href=mailto:raymundo@uai.cl>send me an
      email</a></p>
    <p><a href="papers">Here</a> are some
      papers I've collaborated in in the context
      of my research with
      <a href="http://adriansoto.cl">
      Adrián Soto Suárez.</a></p>
    <p><a href="dogpictures">Here</a> are some
      pictures of my dog.</p>
    ...
```

Web Semántica

- Queremos que la web tenga datos además de información HTML
- Los documentos hablan sobre **entidades**
- Las entidades se modelan como un **grafo**
- Las entidades tienen un identificador único (URI)
- Las entidades tienen relaciones entre sí, y también llevan una URI

Resource Description Framework (RDF)

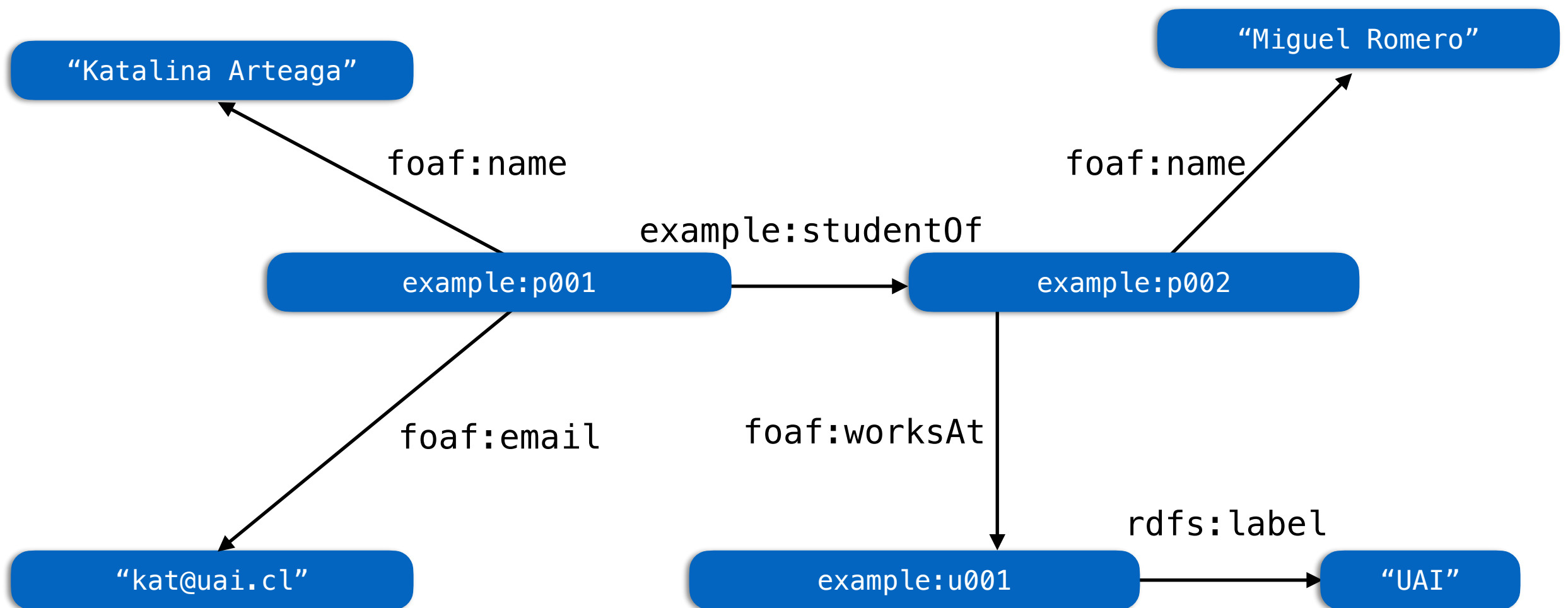
Representación de Triples

Sujeto	Predicado	Objeto
example:p001	foaf:name	"Katalina Arteaga"
example:p002	foaf:name	"Miguel Romero"
example:p001	example:studentOf	example:p002
example:p002	foaf:worksAt	example:u001
example:u001	rdfs:label	"UAI"

Obs example, foaf y rdfs son abreviaciones de prefijos

example:p001 = <http://example.org/p001>

Resource Description Framework (RDF)



Ontologías

Ontologías

- Vocabulario común para definir relaciones
- Ejemplo: FOAF (Friend of a Friend)
 - URI: `<http://xmlns.com/foaf/0.1/>` (Prefijo)
 - `foaf:name` (abreviación de `<http://xmlns.com/foaf/0.1/name>`)
 - `foaf:knows`
 - ...
- Otros ejemplos: OWL, BabelNet, BioPAX...

RDF

RDF

- Datos distribuidos
- Representación de la Web como un grafo
- Utilizadas por miles de personas

SPARQL

SPARQL

Lenguaje de consultas para Bases de Datos RDF

SPARQL

Lenguaje de consultas para Bases de Datos RDF

```
SELECT ?x ?y ?z WHERE { ?x ?y ?z }
```

SPARQL

Ejemplo

SPARQL

Ejemplo

```
SELECT ?name WHERE {  
    ?universidad rdfs:label "UAI" .  
    ?prof foaf:worksAt ?universidad .  
    ?prof foaf:name ?name  
}
```

SPARQL

Ejemplo

“Todos los profesores que hacen clases en la UAI”

```
SELECT ?name WHERE {  
    ?universidad rdfs:label "UAI" .  
    ?prof foaf:worksAt ?universidad .  
    ?prof foaf:name ?name  
}
```

Web Semántica

Web Semántica

- La idea es usar toda la información de la Web como si fuera un grafo.
- Esto implica:
 - Estándar para publicar datos (RDF)
 - Estándar para consultar datos
 - Lidar con la arquitectura y escala de la Web
- Tópico de investigación importante a nivel mundial

Outline

NoSQL

Web Semántica

SPARQL

SPARQL

SPARQL

- Es el lenguaje de consultas de la Web Semántica
- Se basa en Graph Patterns

SPARQL

```
SELECT ?var1 ... ?varn WHERE {  
    <Graph Pattern>  
}
```

Se retornan todos los mappings que cumplan con la expresión señalada en <Graph Pattern>

SPARQL graph patterns

- Conjunciones

```
SELECT * WHERE {  
    ?U :name ?N .  
    ?U :email ?E .  
    ?U :phoneNumber ?P  
}
```

SPARQL graph patterns

- Conjunciones
- Grupos

```
SELECT * WHERE {  
    ?U :name ?N .  
    {  
        ?U :email ?E .  
        ?U :phoneNumber ?P  
    }  
}
```

SPARQL graph patterns

- Conjunciones
- Grupos
- Opcionales

```
SELECT * WHERE {  
    ?U :name ?N .  
    OPTIONAL {  
        ?U :email ?E .  
        ?U :phoneNumber ?P  
    }  
}
```

SPARQL graph patterns

- Conjunciones
- Grupos
- Opcionales
- Anidación de opcionales

```
SELECT * WHERE {  
    ?U :name ?N .  
    OPTIONAL {  
        ?U :email ?E .  
        OPTIONAL {  
            ?U :phoneNumber ?P  
        }  
    }  
}
```

SPARQL graph patterns

- Conjunciones
- Grupos
- Opcionales
- Anidación de opcionales
- Filtrados (selección)
- Unión
- ...

```
SELECT * WHERE {  
  ?U :name ?N .  
  OPTIONAL {  
    ?U :email ?E .  
    OPTIONAL {  
      ?U :phoneNumber ?P  
    }  
  }  
  FILTER (contains(?N,  
    "Letelier")  
}
```

Basic Graph Patterns

Basic Graph Patterns

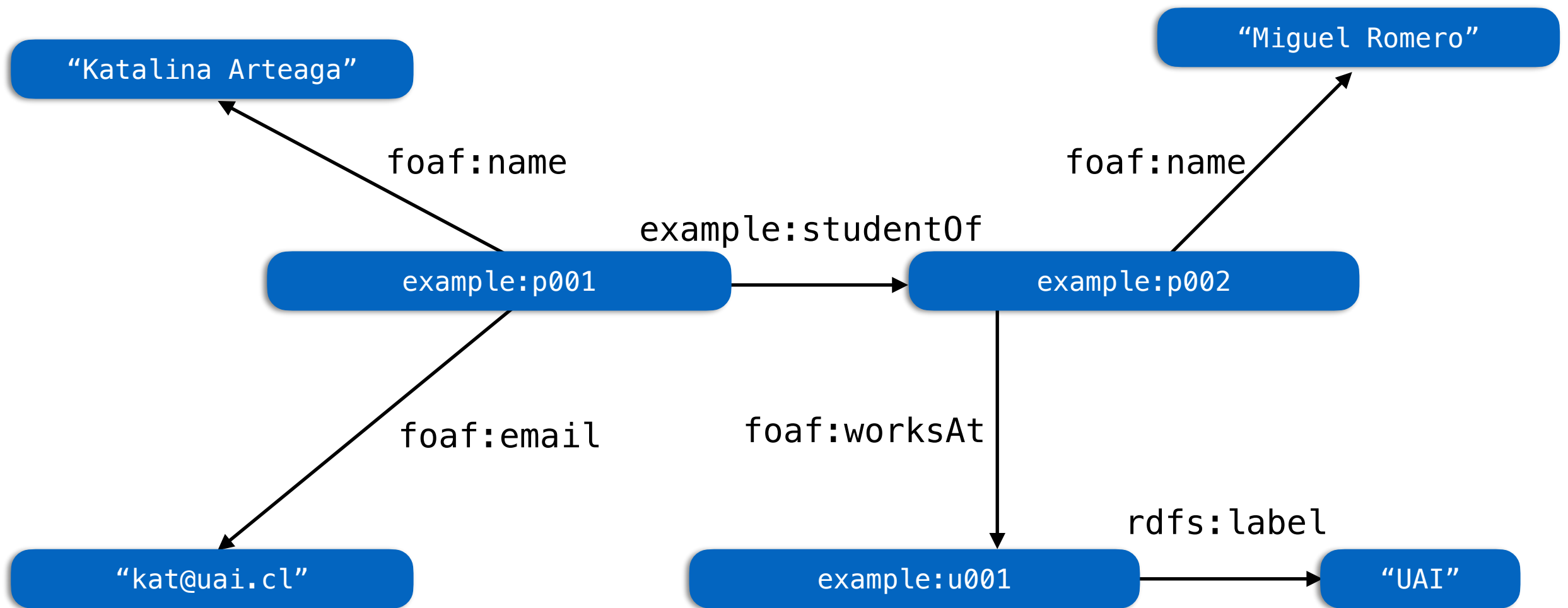
Se hace match de una expresión con los triples de la base de datos

Basic Graph Patterns

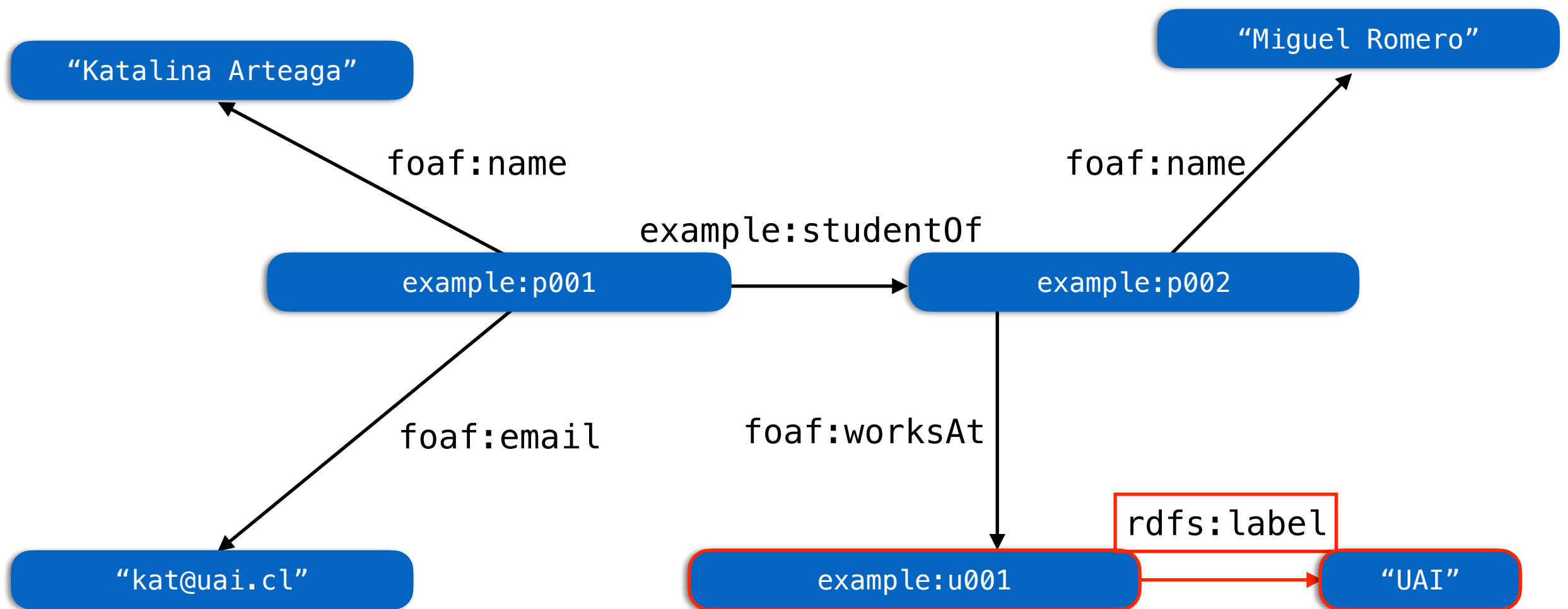
Basic Graph Patterns

Observación: en SPARQL se consultan Grafos, pero una query con SELECT retorna **Mappings**!

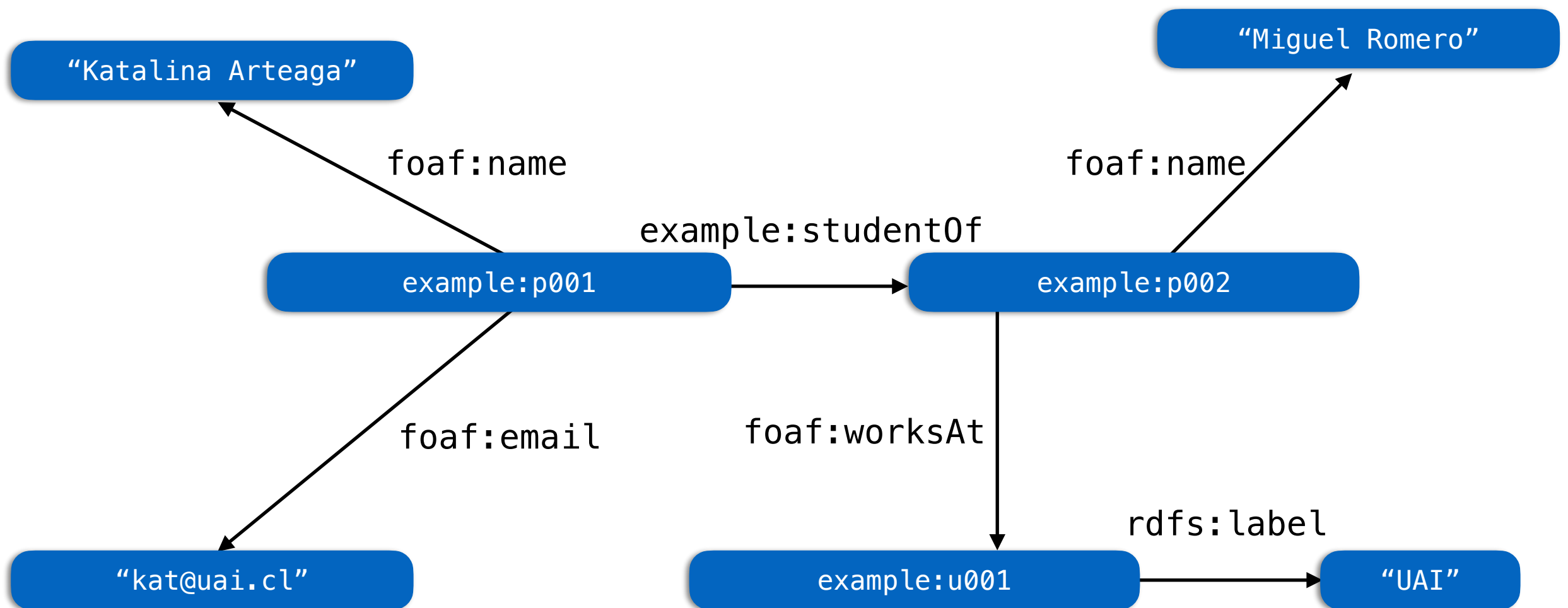
?universidad **rdfs:label** "UAI" .



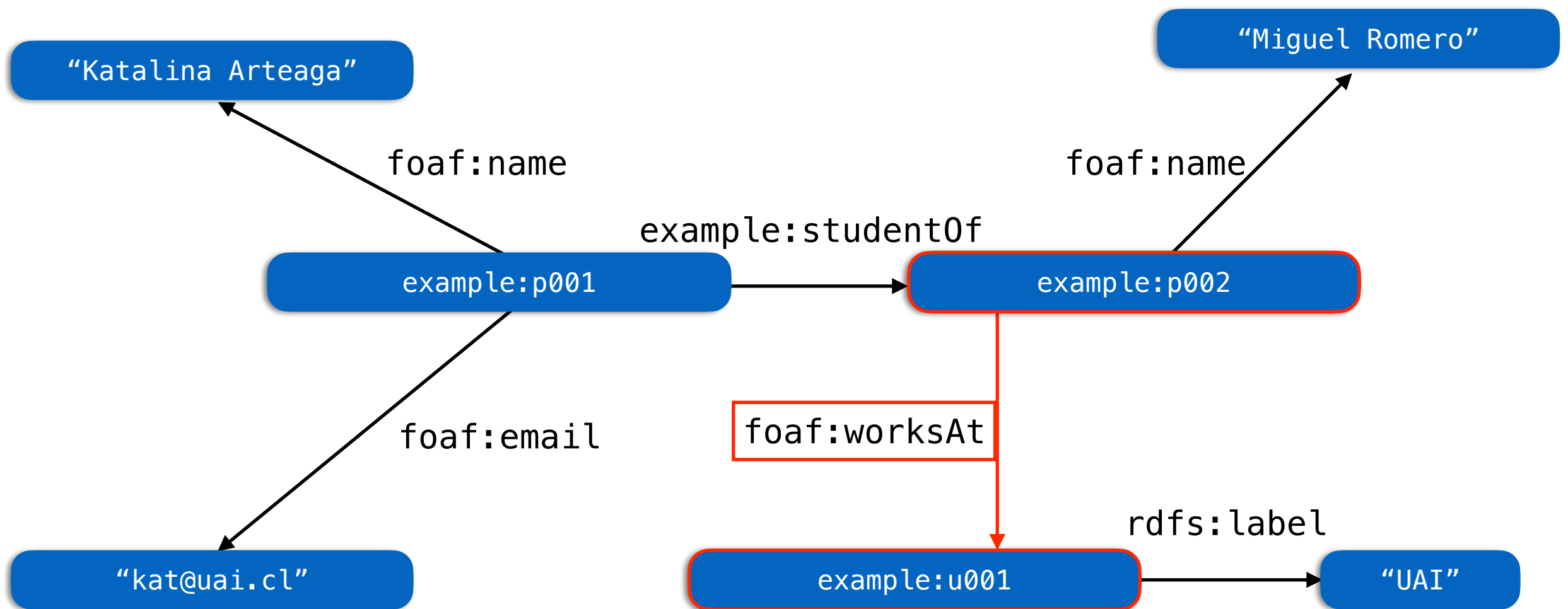
?universidad **rdfs:label** "UAI" .



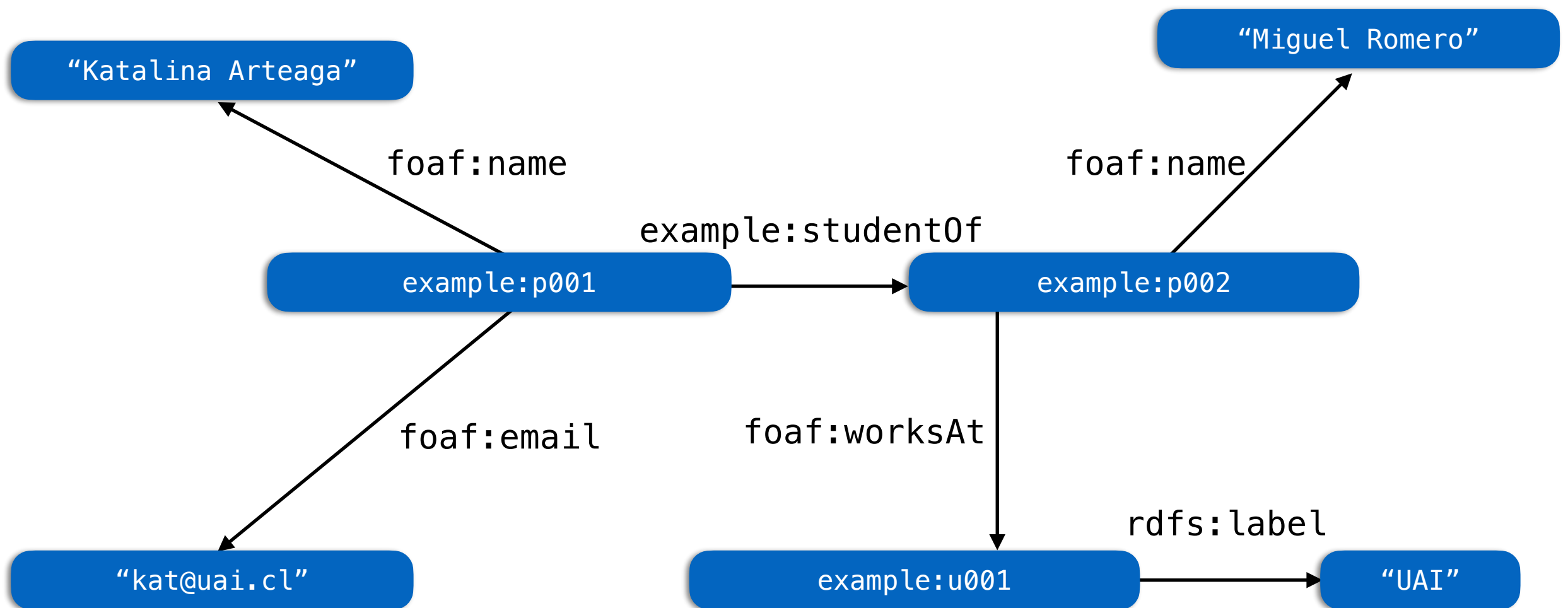
?prof foaf:worksAt ?universidad .



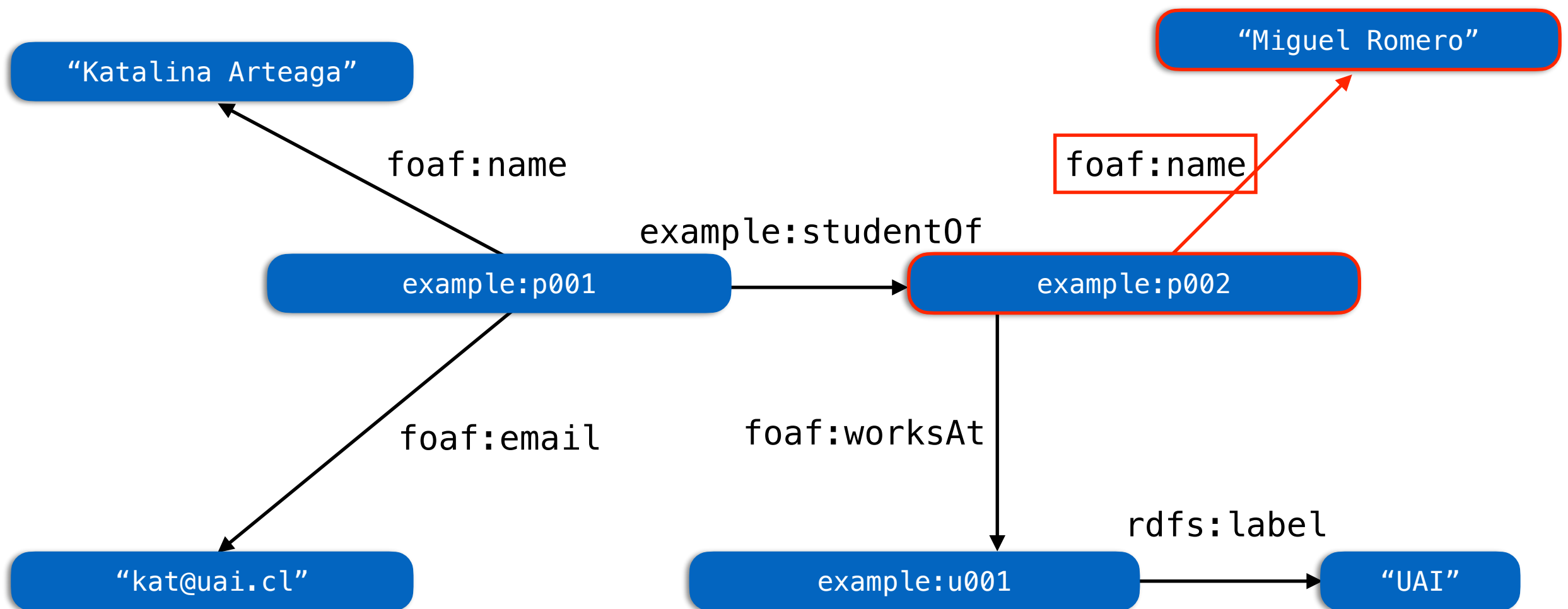
?prof foaf:worksAt ?universidad .



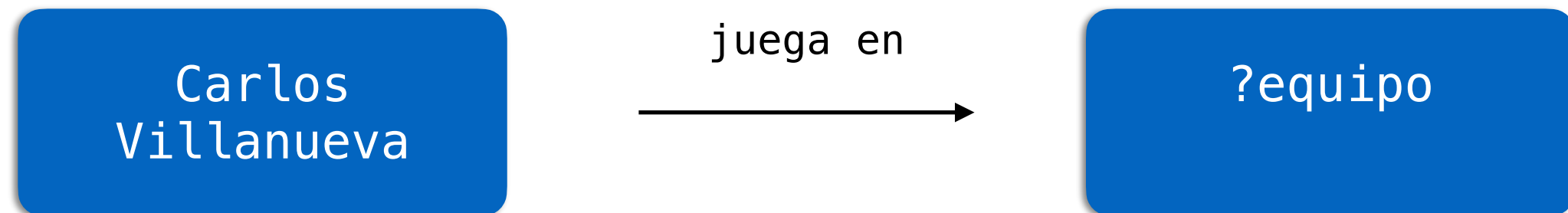
?prof foaf:name ?name



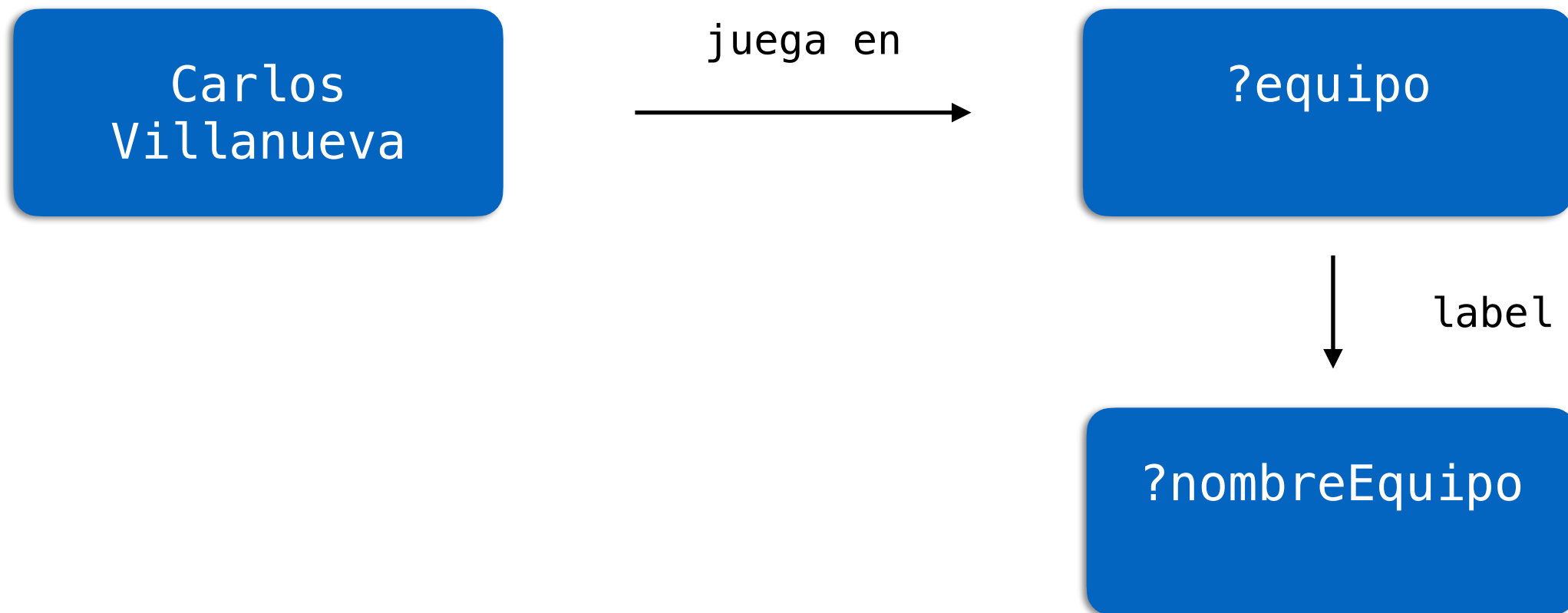
?prof foaf:name ?name



Basic Graph Patterns



Basic Graph Patterns



Basic Graph Patterns

Ejemplo

Basic Graph Patterns

Ejemplo

Todos los equipos de Carlos Villanueva Rolland

```
SELECT ?nombreequipo WHERE {
```

```
<http://www.wikidata.org/entity/Q716965> <http://  
www.wikidata.org/prop/direct/P54> ?equipo .
```

```
?equipo <http://www.w3.org/2000/01/rdf-  
schema#label> ?nombreequipo
```

```
}
```

Basic Graph Patterns

Ejemplo

Basic Graph Patterns

Ejemplo

A partir de ahora, los ejemplos usarán los prefijos respectivos

Basic Graph Patterns

Ejemplo

A partir de ahora, los ejemplos usarán los prefijos respectivos

Todos los equipos de Carlos Villanueva Rolland

```
SELECT ?nombreequipo
WHERE
{
    wd:Q716965 wdt:P54 ?equipo .
    ?equipo rdfs:label ?nombreequipo
}
```


Basic Graph Patterns

Otro ejemplo

Basic Graph Patterns

Otro ejemplo

Todas las comunas de Chile

```
SELECT ?label WHERE
{
    ?comunasdechile rdfs:label "commune of Chile"@en .
    ?comuna wdt:P31 ?comunasdechile .
    ?comuna rdfs:label ?label
}
```

Filter

Filter

Se filtran los mappings arrojados por un Graph Pattern

Filter

Ejemplo

Filter

Ejemplo

Todas las comunas de Chile, pero solamente retornamos los labels en inglés

```
SELECT ?label
WHERE
{
    ?comunasdechile rdfs:label "commune of
Chile"@en .
    ?comuna wdt:P31 ?comunasdechile .
    ?comuna rdfs:label ?label
    FILTER(lang(?label) = "en")
}
```

Filter

Otro ejemplo

Filter

Otro ejemplo

Países con una población mayor a 10000000

```
SELECT ?label ?poblacion
WHERE
{
    ?pais wdt:P31 wd:Q6256 .
    ?pais rdfs:label ?label .
    ?pais wdt:P1082 ?poblacion
    FILTER(lang(?label) = "en" && ?poblacion >
10000000)
}
```


Optional

Optional

- Se añade una cláusula al graph pattern
- Si hay mappings que no hacen match dentro de la cláusula opcional, no se elimina el resultado

Optional

Las selecciones nacionales de fútbol junto a su cuenta de twitter, si es que tienen

```
SELECT ?seleccion ?label ?twitter WHERE {  
    ?seleccion wdt:P31 wd:Q6979593 .  
    ?seleccion rdfs:label ?label .  
    OPTIONAL {?seleccion wdt:P2002 ?twitter}  
    FILTER(lang(?label) = "en")  
}
```

Union

Se unen resultados de dos Graph Patterns

Union

Ejemplo

Union

Ejemplo

Comunas de Chile y Argentina

```
SELECT ?comuna ?label WHERE
{
  {
    ?comunasdechile rdfs:label "commune of Chile"@en .
    ?comuna wdt:P31 ?comunasdechile .
    ?comuna rdfs:label ?label
    FILTER(lang(?label) = "en")
  } UNION {
    ?comuna wdt:P31 wd:Q515 .
    ?comuna wdt:P17 wd:Q414 .
    ?comuna rdfs:label ?label
    FILTER(lang(?label) = "en")
  }
}
```

Construct

Construct

En vez de retornar mappings, se retornan grafos

Construct

Ejemplo

Construct

Ejemplo

Grafo con todas las comunas de Chile

```
CONSTRUCT { ?comuna ex:nombre_comuna ?label } WHERE
{
    ?comunasdechile rdfs:label "commune of Chile"@en .
    ?comuna wdt:P31 ?comunasdechile .
    ?comuna rdfs:label ?label
}
```

Más SPARQL

Más SPARQL

Existen muchos otros operadores de SPARQL:

- SERVICE
- BIND
- COUNT
- DISTINCT
- FROM, FROM GRAPH
- Property Paths
- ...

SPARQL y SQL

Comparación

SPARQL y SQL

Comparación

SPARQL	SQL
Basic Graph Patterns	Similar a hacer un Join
SELECT ?v1 ... ?vn	SELECT v1, ..., vn
Operador OPTIONAL	Operador OUTER JOIN
FILTER	Selección del álgebra relacional
DISTINCT	DISTINCT
UNION	UNION

SPARQL Endpoints

SPARQL Endpoints

- Un Endpoint de SPARQL es un servicio que permite recibir consultas en SPARQL y retornar los resultados.
- Permiten distribuir las bases de datos en muchos servidores.
- Algunos endpoints:
 - DBPedia
 - Yago
 - DBLP
 - WikiData
 - ...

Service

Service

Operador que permite consultar varios endpoint en una consulta

Service

Ejemplo

Service

Ejemplo

```
SELECT ?name WHERE
{
  <http://example.org/myfoaf/I> foaf:knows ?person .
  SERVICE <http://people.example.org/sparql> {
    ?person foaf:name ?name
  }
}
```

Más acerca de SPARQL

- SPARQL 1.1: <https://www.w3.org/TR/sparql11-query/>
- Semántica de SPARQL: http://users.dcc.uchile.cl/~cgutierrez/ftp/sparql_semantics.pdf
- Endpoint de WikiData: <https://query.wikidata.org/>

Técnicas para Big Data

Clase 06 - Web Semántica