



Universidad Simón Bolívar
Coordinación de Ciencias de la Computación
Maestría en Ciencias de la Computación
Web Semántica I

Informe del proyecto: Extendiendo LinkedCT

Fabiola Di Bartolo
09-87324

Diciembre 2010

Índice

| | |
|--|-----------|
| 1. Planteamiento del problema | 3 |
| 2. Problema | 4 |
| 3. Objetivos | 5 |
| 4. Solución | 5 |
| 4.1. Estadísticas del data set LinkedCT extendido | 6 |
| 5. Experimentos y resultados | 6 |
| 5.1. Experimento I | 7 |
| 5.1.1. <i>MeSH terms</i> utilizados para describir enfermedades . . . | 7 |
| 5.1.2. Enfermedades relacionadas según los <i>MeSH terms</i> . . . | 8 |
| 5.1.3. Enlaces entre enfermedades y <i>MeSH terms</i> | 10 |
| 5.1.4. Resumen de los resultados | 12 |
| 5.2. Experimento II | 12 |
| 5.2.1. <i>MeSH terms</i> de drogas utilizados para describir enferme- dades | 12 |
| 5.2.2. Enfermedades relacionadas según los <i>MeSH terms</i> de drogas | 14 |
| 5.2.3. Enlaces entre enfermedades y <i>MeSH terms</i> de drogas . . | 16 |
| 5.2.4. Resumen de los resultados | 17 |
| 5.3. Experimento III | 18 |
| 5.3.1. Drogas empleadas en el tratamiento de enfermedades . . | 18 |
| 5.3.2. Enfermedades relacionadas | 19 |
| 5.3.3. Resumen de los resultados | 21 |
| 5.4. Experimento IV | 21 |
| 5.4.1. Drogas empleadas | 22 |
| 5.4.2. Enfermedades relacionadas | 22 |
| 6. Conclusiones | 23 |

1. Planteamiento del problema

LinkedCT es un *data set* disponible en el *Cloud of linked data*, cuya información sobre los *Clinical trials* proviene de los datos publicados en ClinicalTrials.gov. Sin embargo, todavía existe información que podría ser relevante para la investigación médica, como relaciones entre drogas y enfermedades, que se encuentra disponible pero que no está anotada en LinkedCT.

El *Medical Subject Headings* (MeSH) es un vocabulario controlado empleado comúnmente para indexar o describir artículos o libros científicos y en particular es utilizada para describir *clinical trials*. Es por esto que se desea ampliar el conjunto de datos anotados en rdf de LinkedCT con *MeSH terms*, y así proporcionar un nuevo conocimiento que pueda ser útil.

Debido a que los *MeSH terms* pueden ser muy generales, para evitar introducir datos irrelevantes al *data set* es necesario que sean filtrados según algún criterio. Para descartar información irrelevante en este proyecto se utilizó *Unified Medical Language System* (UMLS) el cual es un compendio de vocabularios controlados en las ciencias biomédicas, y permite la conexión entre los diversos vocabularios que comprende, entre ellos MeSH. Utilizando los *UMLS types* se definen las categorías que serán consideradas para la traducción a RDF de los *MeSH terms*. Los *UMLS types* que fueron tomados en cuenta son los referentes a enfermedades, partes del cuerpo humano y drogas, estos son:

- ★ Body Part, Organ, or Organ Component
- ★ Disease or Syndrome
- ★ Neoplastic Process
- ★ Amino Acid, Peptide, or Protein
- ★ Antibiotic
- ★ Biologically Active Substance
- ★ Biomedical or Dental Material
- ★ Carbohydrate
- ★ Chemical

- ★ Chemical Viewed Functionally
- ★ Chemical Viewed Structurally
- ★ Clinical Drug
- ★ Eicosanoid
- ★ Element, Ion, or Isotope
- ★ Enzyme
- ★ Hazardous or Poisonous Substance
- ★ Hormone
- ★ Immunologic Factor
- ★ Indicator, Reagent, or Diagnostic Aid
- ★ Inorganic Chemical
- ★ Lipid
- ★ Neuroreactive Substance or Biogenic Amine
- ★ Nucleic Acid, Nucleoside, or Nucleotide
- ★ Organic Chemical
- ★ Organophosphorus Compound
- ★ Pharmacologic Substance
- ★ Receptor
- ★ Steroid
- ★ Vitamin

2. Problema

Extender el *data set* LinkedCT con *MeSH terms* asociados a los diferentes *Clinical trials*.

3. Objetivos

- ★ Extraer los *MeSH terms* presentes en los *clinical trials* de ClinicalTrials.gov.
- ★ Extraer los *MeSH terms* presentes en los artículos de MedLine Plus relacionados a los *clinical trials* de ClinicalTrials.gov.
- ★ Filtrar los *MeSH terms* que estén relacionados a los *UMLS types* mencionados anteriormente.
- ★ Anotar en RDF los *MeSH terms* considerados interesantes para este proyecto para cada *clinical trial*.
- ★ Realizar experimentos sobre los nuevos datos semánticos generados.

4. Solución

Para dar solución al problema planteado, se siguieron los siguientes pasos:

- ★ Descarga LinkedCT para manipularlo de forma local. Por cada *clinical trial* en el *data set* se almacenó el artículo de ClinicalTrials.gov al cual hace referencia el predicado `< http : //xmlns.com/foaf/0,1/page >`.
- ★ Reconocimiento por patrones de los *MeSH terms* y los enlaces a Medline Plus presentes en los artículos de ClinicalTrials.gov descargados.
- ★ Extracción y almacenamiento de las páginas Web de Medline Plus referenciadas en los archivos de ClinicalTrials.gov y se realizó reconocimiento por patrones de los *MeSH terms* presentes.
- ★ Integración de los *MeSH terms* obtenidos de ClinicalTrials.gov y Medline Plus para cada *clinical trial*.
- ★ Tomando los enlaces entre los *UMLS types* y los *MeSH terms*, se formó el subconjunto de *MeSH terms* relacionados a los *UMLS types* referentes a enfermedades, partes del cuerpo humano y drogas. Los cuales fueron utilizados como criterio de selección para filtrar aquellos *MeSH terms* interesantes de cada *clinical trial*.

- ★ Traducción a RDF en formato n-triples del conjunto de *MeSH terms* recolectados e integración con el *data set* de LinkedCT local.
- ★ Carga en el manejador RDF-3X del nuevo *data set* y realización experimentos.

4.1. Estadísticas del data set LinkedCT extendido

Algunas características de los datos son presentadas a continuación.

| Estadística | Número |
|---|------------|
| Tripletas de LinkedCT | 9.809.330 |
| Tripletas de LinkedCT extendido | 12.635.552 |
| <i>Clinical trials</i> | 80.237 |
| <i>Clinical trials</i> descritos con <i>MeSH terms</i> | 75.086 |
| <i>Clinical trials</i> en ClinicalTrials.gov con referencias a MedLine Plus | 68.533 |
| <i>MeSH terms</i> que fueron anotados en RDF | 942.074 |
| Tripletas generadas por los <i>MeSH terms</i> | 2.826.222 |

5. Experimentos y resultados

Para observar las nuevas relaciones generadas a partir de los *MeSH terms*, se diseñó un conjunto de consultas a ser ejecutadas en RDF-3X.

Se experimentó con tres *data sets*, en el primero son anotados los *MeSH terms* como elementos del tipo $\langle \text{http} : // \text{purl.org/dc/terms/MESH} \rangle$ y es el utilizado para el experimento I; en el segundo *data set* son anotados los *MeSH terms* diferenciando los términos referentes a drogas para las enfermedades como elementos del tipo $\langle \text{http} : // \text{purl.org/dc/terms/MESH/drugs} \rangle$ y el resto de los términos como $\langle \text{http} : // \text{purl.org/dc/terms/MESH} \rangle$, este último *data set* es empleado en el experimento II; el tercer *data set* es de LinkedCT sin la extensión y es el utilizado en el experimento III.

Se tomó el tiempo de ejecución y el número de resultados obtenidos de las consultas. Cada consulta se ejecutó 10 veces. Para la ejecución de los experimentos se utilizó una computadora Sun Fire X4100 M2 (Maryknoll.bd.cesma.usb.ve), la cual tiene 6GB de memoria RAM y sistema operativo SunOS 5.10 i386.

5.1. Experimento I

Fueron ejecutadas 9 consultas SPARQL distintas, cada una se corrió 10 veces. Las cuales son presentadas a continuación.

5.1.1. *MeSH terms* utilizados para describir enfermedades

Las siguientes consultas tienen como finalidad encontrar los *MeSH terms* asociados a los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 1

Obtener los *MeSH terms* que describen los *clinical trials* relacionados con el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?mesh WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH> .  
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh }
```

★ Tiempo promedio: 0.2525 segundos.

★ Resultados: 708 *MeSH terms*.

Consulta 2

Obtener los *MeSH terms* que describen los *clinical trials* relacionados con el *Colorectal cancer*.

★ Consulta en SPARQL:

```

SELECT DISTINCT ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh }

```

★ Tiempo promedio: 0.2505 segundos.

★ Resultados: 462 *MeSH terms*.

Consulta 3

Obtener los *MeSH terms* que describen los *clinical trials* relacionados con el *Prostate cancer*.

★ Consulta en SPARQL:

```

SELECT DISTINCT ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh }

```

★ Tiempo promedio: 0.251 segundos.

★ Resultados: 574 *MeSH terms*.

5.1.2. Enfermedades relacionadas según los *MeSH terms*

Las siguientes consultas tienen como finalidad encontrar los vínculos entre enfermedades mediante la obtención de las enfermedades asociadas a los *MeSH terms* que referencian los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 4

Obtener las enfermedades que están descritas por los *MeSH terms* que describen a los *clinical trials* relacionados con el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH> .  
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

★ Tiempo promedio: 0.2709 segundos.

★ Resultados: 377 enfermedades.

Consulta 5

Obtener las enfermedades que están descritas por los *MeSH terms* que describen a los *clinical trials* relacionados con el *Colorectal cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH> .  
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

★ Tiempo promedio: 0.2687 segundos.

★ Resultados: 187 enfermedades.

Consulta 6

Obtener las enfermedades que están descritas por los *MeSH terms* que describen a los *clinical trials* relacionados con el *Prostate cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH> .  
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

★ Tiempo promedio: 0.269 segundos.

★ Resultados: 227 enfermedades.

5.1.3. Enlaces entre enfermedades y *MeSH terms*

Las siguientes consultas tienen como finalidad encontrar los vínculos entre enfermedades y los *MeSH terms* que referencian los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 7

Obtener las enfermedades con sus *MeSH terms* que describen a su vez los *clinical trials* relacionados con el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition ?mesh WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH> .
```

```
?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh .
?v <http://purl.org/dc/elements/1.1/subject> ?m .
?v a <http://data.linkedct.org/resource/linkedct/trials> .
?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

★ Tiempo promedio: 0.5802 segundos.

★ Resultados: 16.818 enlaces.

Consulta 8

Obtener las enfermedades con sus *MeSH terms* que describen a su vez los *clinical trials* relacionados con el *Colorectal cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

★ Tiempo promedio: 0.4679 segundos.

★ Resultados: 9.441 enlaces.

Consulta 9

Obtener las enfermedades con sus *MeSH terms* que describen a su vez los *clinical trials* relacionados con el *Prostate cancer*.

★ Consulta en SPARQL:

```

SELECT DISTINCT ?condition ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }

```

★ Tiempo promedio: 0.471 segundos.

★ Resultados: 10.041 enlaces.

5.1.4. Resumen de los resultados

A continuación se muestra una tabla resumen de los resultados obtenidos sobre las asociaciones entre los *MeSH terms* y los distintos *clinical trials*.

| Enfermedad | <i>MeSH terms</i> | Enfermedades asociadas | Enlaces enfermedades y <i>MeSH terms</i> |
|-------------------|-------------------|------------------------|--|
| Breast Cancer | 705 | 377 | 16.818 |
| Colorectal Cancer | 462 | 187 | 9.441 |
| Prostate Cancer | 574 | 227 | 10.041 |

5.2. Experimento II

Con la finalidad de encontrar relaciones entre las enfermedades y las drogas utilizadas en su tratamiento. Fueron ejecutadas 9 consultas SPARQL distintas, cada una se corrió 10 veces. Las cuales son presentadas a continuación.

5.2.1. *MeSH terms* de drogas utilizados para describir enfermedades

Las siguientes consultas tienen como finalidad encontrar los *MeSH terms* referentes a drogas asociados a los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 1

Obtener los *MeSH terms* referentes a drogas que describen los *clinical trials* relacionados con el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?mesh WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH/drugs> .  
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh }
```

★ Tiempo promedio: 0.1906 segundos.

★ Resultados: 361 *MeSH terms* de drogas.

Consulta 2

Obtener los *MeSH terms* referentes a drogas que describen los *clinical trials* relacionados con el *Colorectal cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?mesh WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .  
  ?m a <http://purl.org/dc/terms/MESH/drugs> .  
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh }
```

★ Tiempo promedio: 0.1892 segundos.

★ Resultados: 192 *MeSH terms* de drogas.

Consulta 3

Obtener los *MeSH terms* referentes a drogas que describen los *clinical trials* relacionados con el *Prostate cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh }
```

★ Tiempo promedio: 0.1893 segundos.

★ Resultados: 298 *MeSH terms* de drogas.

5.2.2. Enfermedades relacionadas según los *MeSH terms* de drogas

Las siguientes consultas tienen como finalidad encontrar los vínculos entre enfermedades mediante la obtención de las enfermedades asociadas a los *MeSH terms* de drogas que referencian los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 4

Obtener las enfermedades que están descritas por los *MeSH terms* referentes a drogas que describen a los *clinical trials* relacionados con el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

- ★ Tiempo promedio: 0.2064 segundos.
- ★ Resultados: 278 enfermedades.

Consulta 5

Obtener las enfermedades que están descritas por los *MeSH terms* referentes a drogas que describen a los *clinical trials* relacionados con el *Colorectal cancer*.

- ★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

- ★ Tiempo promedio: 0.2043 segundos.
- ★ Resultados: 121 enfermedades.

Consulta 6

Obtener las enfermedades que están descritas por los *MeSH terms* referentes a drogas que describen a los *clinical trials* relacionados con el *Prostate cancer*.

- ★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

- ★ Tiempo promedio: 0.2046 segundos.
- ★ Resultados: 162 enfermedades.

5.2.3. Enlaces entre enfermedades y *MeSH terms* de drogas

Las siguientes consultas tienen como finalidad encontrar los vínculos entre enfermedades y los *MeSH terms* de drogas que referencian los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 7

Obtener las enfermedades con sus *MeSH terms* referentes a drogas que describen a su vez los *clinical trials* relacionados con el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }
```

★ Tiempo promedio: 0.4088 segundos.

★ Resultados: 3.345 enlaces.

Consulta 8

Obtener las enfermedades con sus *MeSH terms* referentes a drogas que describen a su vez los *clinical trials* relacionados con el *Colorectal cancer*.

★ Consulta en SPARQL:


```

SELECT DISTINCT ?condition ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }

```

★ Tiempo promedio: 0.3511 segundos.

★ Resultados: 1.263 enlaces.

Consulta 9

Obtener las enfermedades con sus *MeSH terms* referentes a drogas que describen a su vez los *clinical trials* relacionados con el *Prostate cancer*.

★ Consulta en SPARQL:

```

SELECT DISTINCT ?condition ?mesh WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://purl.org/dc/elements/1.1/subject> ?m .
  ?m a <http://purl.org/dc/terms/MESH/drugs> .
  ?m <http://www.w3.org/2000/01/rdf-schema#label> ?mesh .
  ?v <http://purl.org/dc/elements/1.1/subject> ?m .
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/condition_name> ?condition }

```

★ Tiempo promedio: 0.3483 segundos.

★ Resultados: 1.377 enlaces.

5.2.4. Resumen de los resultados

A continuación se muestra una tabla resumen de los resultados obtenidos sobre las asociaciones entre los *MeSH terms* de drogas y los distintos *clinical trials*.

| Enfermedad | <i>MeSH/drugs</i> | Enfermedades asociadas | Enlaces enfermedades y <i>MeSH/drugs</i> |
|-------------------|-------------------|------------------------|--|
| Breast Cancer | 361 | 278 | 3.345 |
| Colorectal Cancer | 192 | 121 | 1.263 |
| Prostate Cancer | 298 | 162 | 1.377 |

5.3. Experimento III

Con la finalidad de contrastar los resultados de los experimentos anteriores, se ejecutaron 6 consultas SPARQL distintas sobre el *data set* de LinkedCT, cada una ejecutada 10 veces. Las cuales son presentadas a continuación.

5.3.1. Drogas empleadas en el tratamiento de enfermedades

Las siguientes consultas tienen como finalidad encontrar las drogas asociados a los *clinical trials* que documentan el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 1

Obtener las drogas utilizadas en el tratamiento del *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?droga WHERE {
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .
  ?u <http://data.linkedct.org/resource/linkedct/intervention> ?y .
  ?y <http://data.linkedct.org/resource/linkedct/intervention_type> "Drug" .
  ?y <http://data.linkedct.org/resource/linkedct/intervention_name> ?droga}
```

★ Tiempo promedio: 0.028 segundos.

★ Resultados: 462 drogas.

Consulta 2

Obtener las drogas utilizadas en el tratamiento del *Colorectal cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?droga WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_type> "Drug" .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_name> ?droga}
```

★ Tiempo promedio: 0.027 segundos.

★ Resultados: 194 drogas.

Consulta 3

Obtener las drogas utilizadas en el tratamiento del *Prostate cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?droga WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_type> "Drug" .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_name> ?droga}
```

★ Tiempo promedio: 0.027 segundos.

★ Resultados: 294 drogas.

5.3.2. Enfermedades relacionadas

Las siguientes consultas tienen como finalidad encontrar los vínculos entre enfermedades mediante la obtención de las enfermedades tratadas con las drogas empleadas en el tratamiento del *Breast cancer*, el *Colorectal cancer* y el *Prostate cancer*.

Consulta 4

Obtener las enfermedades que han sido tratadas con las drogas que se trata el *Breast cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Breast Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_type> "Drug" .  
  ?v <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?x2 .  
  ?x2 <http://data.linkedct.org/resource/linkedct/condition_name> ?condition}
```

★ Tiempo promedio: 0.050 segundos.

★ Resultados: 528 enfermedades.

Consulta 5

Obtener las enfermedades que han sido tratadas con las drogas que se trata el *Colorectal cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Colorectal Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_type> "Drug" .  
  ?v <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?x2 .  
  ?x2 <http://data.linkedct.org/resource/linkedct/condition_name> ?condition}
```

★ Tiempo promedio: 0.0489 segundos.

★ Resultados: 398 enfermedades.

Consulta 6

Obtener las enfermedades que han sido tratadas con las drogas que se trata el *Prostate cancer*.

★ Consulta en SPARQL:

```
SELECT DISTINCT ?condition WHERE {  
  ?x <http://data.linkedct.org/resource/linkedct/condition_name> "Prostate Cancer" .  
  ?u <http://data.linkedct.org/resource/linkedct/condition> ?x .  
  ?u a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?u <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?y <http://data.linkedct.org/resource/linkedct/intervention_type> "Drug" .  
  ?v <http://data.linkedct.org/resource/linkedct/intervention> ?y .  
  ?v a <http://data.linkedct.org/resource/linkedct/trials> .  
  ?v <http://data.linkedct.org/resource/linkedct/condition> ?x2 .  
  ?x2 <http://data.linkedct.org/resource/linkedct/condition_name> ?condition}
```

★ Tiempo promedio: 0.049 segundos.

★ Resultados: 493 enfermedades.

5.3.3. Resumen de los resultados

A continuación se muestra una tabla resumen de los resultados obtenidos sobre las asociaciones entre los *MeSH terms* de drogas y los distintos *clinical trials*.

| Enfermedad | Drogas | Enfermedades asociadas |
|-------------------|--------|------------------------|
| Breast Cancer | 462 | 528 |
| Colorectal Cancer | 194 | 398 |
| Prostate Cancer | 294 | 493 |

5.4. Experimento IV

Este experimento consiste en comparar los datos obtenidos con el *data set* de LinkedCT y el *data set* extendido.

5.4.1. Drogas empleadas

De los resultados obtenidos en los experimentos II y III, se compararon las drogas obtenidas por los *MeSH terms* con las utilizadas en las intervenciones, para identificar si existen otras posibles drogas que no han sido estudiadas en esos *clinical trials*.

Debido a que las drogas no fueron escritas siguiendo el mismo estándar para los MeSH/drugs y los intervention_name, podrían existir drogas iguales cuyos nombres hayan sido escrito de forma diferente que no se hayan tomado como iguales, sin embargo, es evidente que existe una gran diferencia entre los MeSH/drugs y los intervention_name, lo que comprueba que nuevas relaciones entre enfermedades y drogas han sido añadidas en el *data set* LinkedCT extendido.

| Enfermedad | MESH/drugs | intervention_name | Drogas en común |
|-------------------|------------|-------------------|-----------------|
| Breast Cancer | 361 | 462 | 31 |
| Colorectal Cancer | 192 | 194 | 12 |
| Prostate Cancer | 298 | 294 | 30 |

5.4.2. Enfermedades relacionadas

De los resultados obtenidos en los experimentos II y III, se compararon las enfermedades asociadas a los *MeSH terms* de drogas que describen el *Breast cancer*, *Colorectal cancer* y *Prostate cancer* con las enfermedades tratadas con las drogas utilizadas en el tratamiento del *Breast cancer*, *Colorectal cancer* y *Prostate cancer*, para identificar si existen nuevas relaciones entre enfermedades.

Nótese que no son las mismas enfermedades que se encontraron en las consultas, lo que comprueba que nuevas relaciones entre enfermedades han sido añadidas en el *data set* LinkedCT extendido.

| Enfermedad | Asociadas a los <i>MeSH terms</i> | Asociadas a los intervention_name | Enfermedades en común |
|-------------------|--------------------------------------|--------------------------------------|--------------------------|
| Breast Cancer | 278 | 528 | 187 |
| Colorectal Cancer | 121 | 398 | 87 |
| Prostate Cancer | 162 | 493 | 100 |

6. Conclusiones

Según los experimentos efectuados, nuevas relaciones entre los datos han sido creadas en la extensión del *data set* LinkedCT, las cuales pueden ser utilizadas para diversas investigaciones científicas en la búsqueda de curas de patologías. Sin embargo, queda a criterio del investigador identificar si la nueva información producida es útil para el estudio de los distintos *clinical trials*.

Sería interesante encontrar relaciones entre las diversas drogas empleadas para el tratamiento de una enfermedad, las cuales podrían ser probadas para patologías relacionadas. Asimismo, es recomendable encontrar porcentajes de efectividad y de uso de las drogas en los tratamientos de estas enfermedades, para documentarlos en RDF y complementar el conocimiento generado a partir de los *MeSH* terms.