

Organización de Datos
Segundo Cuatrimestre de 2010
Trabajo Practico

Grados de Separación.

Conociendo las relaciones establecidas entre un grupo de personas se puede establecer la cantidad de grados de separación que hay entre ellas.

Los grados de separación se miden teniendo en cuenta la distancia de personas conocidas que hay entre unos y otros.

Objetivo

Se debe construir una aplicación que permita establecer la cantidad de grados de separación de un grupo de personas a partir de un conjunto de información. El conjunto de datos a considerar es el provisto por la catedra a partir de DBpedia.org considerando:

- **Personas:** Actores, Directores y Escritores de films.
- **Relaciones entre personas:** Estarán establecidas por los films en los que las personas propuestas hayan estado involucradas.

Se deberán implementar los siguientes casos de uso:

Calculo de grados de separación:

La aplicación recibirá el nombre (origen y fin) de cada una de las personas a calcular los grados de separación.

La respuesta esperada del mismo consistirá en el numero de grados de separación y una descripción detallada de como se logra este camino (incluyendo cada uno de las el nombre y profesión de cada una de las personas a partir del cual se establece un grado de separación y partir de que film se estableció).

Personas a cierto grado de separación:

La aplicación recibirá un nombre de una persona y un numero de grados de de separación a considerar.

La respuesta esperada sera un listado de las personas que se encuentran a esa cantidad de grados de separación.

Detalles de Implementación

Base de Conocimientos

La información sera provista por la cátedra la cual ha sido extraída de DBpedia.org. La misma sera entregada en archivos de formato XML conteniendo varios films, por ejemplo:

```

<films>
  <film>
    <id>http://dbpedia.org/resource/Mulholland_Drive_(film)</id>
    <titles>
      <title>Mulholland Dr.</title>
    </titles>
    <directors>
      <director>http://dbpedia.org/resource/David_Lynch</
director>
    </directors>
    <writers>
      <writer>http://dbpedia.org/resource/David_Lynch</writer>
    </writers>
    <actors>
      <actor>http://dbpedia.org/resource/Laura_Harring</actor>
      <actor>http://dbpedia.org/resource/Naomi_Watts</actor>
      <actor>http://dbpedia.org/resource/Melissa_George</actor>
      <actor>http://dbpedia.org/resource/Ann_Miller</actor>
      <actor>http://dbpedia.org/resource/Robert_Forster</actor>
      <actor>http://dbpedia.org/resource/Dan_Hedaya</actor>
      <actor>http://dbpedia.org/resource/Justin_Theroux</actor>
    </actors>
  </film>
</films>

```

Cada uri especificada por recurso, ya sea un film o actor, puede ser considerada como un indentificador.

Esta informacion debera ser procesada y almacenada localmente en un formato que permita resolver el objetivo del trabajo practico.

La información provista permitirá realizar el trabajo practico (tanto los nombres de los films como los nombres de los actores pueden ser extraídos de las uris y limpiados eventualmente). Si el grupo considera necesario acceder a mayor información para desarrollarlo podrá hacerlo accediendo a DBPedia.org usando Sparql.

El procesamiento del xml provisto por la cátedra podra realizarse C o C++ o en algun otro lenguaje como python o ruby.

Motor de calculo de grados de separación.

Consistirá en una aplicaron de consola, desarrollada en C o C++ bajo entorno Unix que hara uso de la base de conocimientos.

Ejemplos de ejecución:

```
user@machine:~$ ./degree --from "Kevin Bacon" --to "Wes Anderson"
```

degree of separation: 2

* Kevin Bacon was in The River Wild with Meryl Streep.
* Meryl Streep was in The Fantastic Mr. Fox with Wes Anderson.

user@machine:~\$./degree --from "Kevin Bacon" --to "Tony Leung"
degree of separation: 3

* Kevin Bacon was in Frost Nixon with Arne Starr.
* Arne Starr was in Rush Hour 3 with Jackie Chan.
* Jackie Chang was in Xin jing cha gu shi with Tony Leung.

user@machine:~\$./degree --from "Kevin Bacon" --degree 2

* Jackie Chang.
* Wes Anderson.
* Noah Baumbauch.
... (mas resultados)

Condiciones de Entrega

El trabajo práctico tendrá dos entregas, las cuales se encuentran indicadas en el calendario de la materia.

Entrega de Diseño:

Esta primer entrega deberá ser pensada como una especificación de cómo se resolverá el problema planteado en la sección anterior.

Como mínimo esta entrega deberá contener:

- Detalle de la estrategia a seguir para realizar la importación de los datos al entorno local.
- Detalle de la estrategia de resolución del problema de grados de separación especificando algoritmos utilizados.
- Todas las estructuras de datos tanto de almacenamiento físico como en memoria a utilizar.

A su vez se podrán incluir como opcionales:

- Diagramas de arquitectura de la aplicación o sus distintas partes.
- Ejemplos de como se realizan los casos de uso del trabajo practico sobre las estructuras planteadas para tener una mejor comprensión de los procesos planteados.
- Otro tipo de documentación que el grupo considere pertinente para la entrega.

El formato de la entrega consistirá en una carpeta en cuya carátula figurará el numero de grupo de trabajo practico y cada uno de los integrantes del mismo junto con su padrón y email.

Pueden usarse gráficos o diagramas para ilustrar los puntos requeridos.

Entrega Final:

Las condiciones de entrega de fijarán luego de la entrega de diseño.

Links de interes

<http://wiki.dbpedia.org/>

<http://wiki.dbpedia.org/OnlineAccess>

<http://www.w3.org/TR/rdf-sparql-query/>

http://en.wikipedia.org/wiki/Shortest_path_problem

http://en.wikipedia.org/wiki/Small_world_experiment

http://en.wikipedia.org/wiki/Six_degrees_of_separation