PROPOSING RICH VIEWS OF LINKED OPEN DATA SETS
THE S-PATHS PROTOTYPE AND THE VISUALIZATION OF FRBR-IZED DATA IN DATA.BNF

27/11/2019
Raphaëlle Lapôtre, Marie Destandau & Emmanuel Pietriga
UNPREDICTABILITY

IRREGULARITY

VOLUME


FACETED BROWSERS


**SET-BASED VISUALISATIONS**

---


**Mazumdar, Suvodeep, Daniela Petrelli, and Fabio Ciravegna.** "Exploring user and system requirements of linked data visualization through a visual dashboard approach." *Semantic Web* 5.3 (2014): 203-220.

---

**Fig. 2.** Map View with selected restaurants


S-PATHS

• A generic approach to continuously explore (from overview to detail) sets of RDF data with no a-priori knowledge of the model

• Offer a readable default view at any stage, and let the user explore other configurations at will

• Enable advanced selection
Jacques Bertin “Useful information is a cluster”
La Graphique et le traitement de l'information
Jacques Bertin “Useful information is a cluster”
La Graphique et le traitement de l'information
FOLLOW THE PATHS

http://data.nobelprize.org/terms/laureateAward

http://data.nobelprize.org/terms/category

http://www.w3.org/2000/01/rdf-schema#label
ANALYSIS OF THE PATHS

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>category</td>
<td>one of: datetime, geographical coordinate, image, number, text or URI</td>
</tr>
<tr>
<td>depth</td>
<td>number of hops (statements) from the subject to the final property</td>
</tr>
<tr>
<td>coverage</td>
<td>percentage of resources in the set for which this path actually exists</td>
</tr>
<tr>
<td>count</td>
<td>total number of values for the property at the end of the path, over all resources</td>
</tr>
<tr>
<td>unique count</td>
<td>number of unique values for the property at the end of the path, over all resources</td>
</tr>
</tbody>
</table>

```
SELECT DISTINCT
  ?p1 ?p2 ?pn
  ?datatype
  (COUNT(?totalEntities) as ?nbEntities)
  (COUNT(?values) as ?uniqueValues)
  (COUNT(DISTINCT ?entities) as ?nbCoveredEntities)
  (COUNT(?values) as ?totalValues)
WHERE {
  ?totalEntities rdf:type <TYPE_URI> .
  ?coveredEntities rdf:type <TYPE_URI> .
  BIND(?datatype(VALUES) as ?datatype)
}
```

Fig. 1. Theoretical query to retrieve paths characteristics. S-Paths splits this query into multiple queries performed recursively.
SET OF VIEWS
SET OF VIEWS

```javascript
let views = [
  {
    id: '2DDensityPlot',
    name: '2ddensityplot',
    thumb: '/images/2ddensityplot.svg',
    constraints: [
      {
        category: 'datetime', unique: { min: 2 }
      },
      {
        category: 'text', avg: { max: 70, optimal: [10, 40] }, unique: { min: 2, max: 150 }
      },
      {
        category: 'url', unique: { min: 2, max: 150 }
      }
    ]
  }
];
```
### SET OF VIEWS

<table>
<thead>
<tr>
<th>View</th>
<th>Type</th>
<th>Weight</th>
<th>Number of resources</th>
<th>dimension 1</th>
<th>dimension 2</th>
<th>dimension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>density plot</td>
<td>aggregate</td>
<td>0.5</td>
<td></td>
<td>datetime, text or uri</td>
<td>text or uri</td>
<td></td>
</tr>
<tr>
<td>treemap</td>
<td>aggregate</td>
<td>0.3</td>
<td></td>
<td>single path: text or uri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stacked chart</td>
<td>multiple distinct</td>
<td>0.9</td>
<td>min/max: [2, 1000]</td>
<td>single path:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>optimal: [4, 200]</td>
<td>text or uri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>timeline</td>
<td>multiple distinct</td>
<td>0.85</td>
<td>min/max: [2, 50]</td>
<td>single path:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>optimal: [10, 20]</td>
<td>text or uri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>URI wheel</td>
<td>multiple distinct</td>
<td>0.4</td>
<td>min: 2</td>
<td>all datetime paths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>map</td>
<td>multiple distinct</td>
<td>0.85</td>
<td>min/max: [2,1000]</td>
<td>geo</td>
<td>geo</td>
<td>text</td>
</tr>
<tr>
<td>breakdown by values</td>
<td>multiple distinct</td>
<td>0.7</td>
<td>max: 50, optimal: [1, 30]</td>
<td>single path: any category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>images</td>
<td>multiple distinct</td>
<td>0.8</td>
<td>min/max: [2, 1000]</td>
<td>single path: image</td>
<td>single path: text</td>
<td></td>
</tr>
<tr>
<td>info card</td>
<td>single entity</td>
<td>1</td>
<td>min/max: [1,1]</td>
<td>all paths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>node link diagram</td>
<td>single entity</td>
<td>0.5</td>
<td>min/max: [1.1]</td>
<td>all paths</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**

Default configuration of view templates as configured for the Nobel dataset.
A MATCHING ALGORITHM
TRANSITIONS, BRUSHING & LINKING
A SAMPLE OF DATA.BNF

Nobel : ≈ 87 000 triplets

Data: representative sample = 33 118 586 triplets
≈ 10 %

Main entities:

<table>
<thead>
<tr>
<th>skos:Concept (395297)</th>
</tr>
</thead>
<tbody>
<tr>
<td>frbr-rda:Work (80991)</td>
</tr>
<tr>
<td>bnffr:expositionVirtuelle (3818)</td>
</tr>
<tr>
<td>Person (165722)</td>
</tr>
<tr>
<td>Organization (38450)</td>
</tr>
<tr>
<td>Document (4218)</td>
</tr>
</tbody>
</table>
You are visualizing 105539 entities belonging to the class 'Person' according to 2 property paths traversing 1 graph.

Focus on collection:

Plot from selection to...
S-PATHS

You are visualizing 2937 entities belonging to the class of resources Document according to 2 property paths traversing 1 graph.

Focus on selection

Subject / focus / Language of the person (Descrete/d)
Sample of entities in the current selection

Path 1:

Path 2:
Http://ark.bnf.fr/ark:12345/6789/0123456789 / Subject / http://cata.berkley.edu/ark:12345/6789/0123456789 / Focus / http://data.berkley.edu/ark:12345/6789/0123456789 / Language of the person (Deprecated) / 0
`Abd Allāh ibn `Umar ibn Muḥammad ibn `Alī al- Bāyḍāwī (12...-1286?)

L'ongue : arabe
Sexe : masculin
Naissance : Al-Bayḍā (Iran) 12...
Mort : 1286
Note : Savant spécialiste du droit Islamique
Autres formes du nom :
Abd Allāh Bayḍāwī (12...-1286?)
Nūsīr al-Dīn `Abd Allāh ibn `Umar al- Bāyḍāwī (12...-1286?)

ISNI : ISNI 0000 0001 1897 9016

Ses activités
(16 documents)
Documents à propos de cet auteur
(2 pages)
Pages dans data.bnf.fr
Sources et références

Voir tous les documents (16)
Voir les documents numérisés (*)

Auteur du texte (14)
Auteur du commentaire (1)
Autre (1)

Auteur du texte
14 documents

Description : Note : Tafsīr al-Qorān. Commentaire du Coran en marge du texte sacré
[catalogue]

Description : Note : Trad. anglais. coulo. Index
Édition : Leiden Boston Brill, 2002
[catalogue]
DIFFICULTIES

• Recursive analysis of paths

• Binning in the query

• Number of paths => select relevant branches to explore for subselections
IN PRINCIPLE

S-Paths can handle a deep model
IN REALITY

Number of entities of a type

Depth of relevant paths

Cost of query
APPLICATIONS

- reveal defects in data sources
- visualize modeling specificities
- show trends in the data that can be used for communication towards end users.
EVALUATION

- On the applications
- About readability / understanding
http://s-paths.lri.fr

marie.destandau@inria.fr

@ marie_ototoi