Visualising Complex Linked Data

Student: Quanwei Han
Supervisor: Assoc. Prof. Peter Christen, Mr. Jeffrey Fisher
Motivation

- **Data linkage**[1]
  - Allows enrichment of data
  - Helps improve data quality
  - Enables data mining on multiple databases

- **Data visualisation**
  - Communicate information efficiently and intuitively[2]
  - Find hidden patterns and trends[3]

- **The visualisation of linked data**[4]
  - Better understand the characteristics of linked data
  - Identify potentially wrong or inconsistent links

---

[4] Source: individual project contract for this project.
Problem definition & Contribution

Problem definition

Research and develop a novel network- and graph-based visualisation technique for complex linked data.

Contribution

• A technique to visualise complex linked data
• A set of visualisations of demographic data

Source: individual project contract for this project.
Background about the data

A dataset which linked the 19th-century census data and civil registration data from the Isle of Skye in Scotland[1]

- A natural boundary
- Population: 21,000 (1861)


Background about the data(2)

civil registration data

birth

marriage

death

longitudinal

1861 census

1871 census

1881 census

1891 census

1901 census

census data
Solution

➤ Visualisation
  • The characteristics of linked demographic data
  • The potential inconsistencies generated by the data linkage process

➤ Quality
  • Interactive
  • Flexible
  • Configurable
Solution

- **Visualisation**
  - The characteristics of linked demographic data
  - The potential inconsistencies generated by the data linkage process

- **Quality**
  - Interactive
  - Flexible
  - Configurable
Conceptual graph

- Life track
- Family relationships
Single life segment visualisation
Single life segment visualisation (2)
Solution

- **Visualisation**
  - The characteristics of linked demographic data
  - The potential inconsistencies generated by the data linkage process

- **Quality**
  - Interactive
  - Flexible
  - Configurable
Data inconsistencies visualisation

- Visualised by coloured lines
  - error -> red
  - warning -> orange
  - single person -> life line
  - multiple persons -> relationship line
Data inconsistencies visualisation

![Graph showing data inconsistencies with timeline and markers for birth, death, birth of child, and marriage events.](image)
Data inconsistencies visualisation (3)

conflict information: more than 3 marriages in his/her life (level: warning)
Evaluation

- Self evaluation
  - Once a week
  - Compare the effect of design elements

- User evaluation
  - At the end of project
  - Get feedback from users

Self evaluation – Example

There might be more than one event in the same year

The stacked circles usually overlap the squares which represent spouses as the age gap between bride and groom is usually not too large
User evaluation – Summary

- The visualisation is a “clean and tidy” way to illustrate “messy” linked records.

- Showing potential wrong links via colouring lines is novel and useful.

- The usefulness of the models cannot be confirmed without practical application.

- It would be useful to visualise the data at a higher level and check more individuals simultaneously.

Source: feedback from the data linkage researchers in Cambridge and Scotland.
Conclusion

- The visualisation of complex linked databases
  - The characteristics of linked data
  - Potentially wrong or inconsistent links

- Multi-layer display
  - Graphical interface
  - Additional functions

- Further work
  The visualisation at higher levels
Thank you!
Additional functions

➢ Purpose
  • provide extra information
  • facilitate continuous research

➢ Functions
  • right click function
  • left click function
  • browse history
Right click function

(1871) To be displayed here. record 3328-7207021
Left click function & Browse history
Left click function & Browse history
Flexibility and reusability

- Configuration file
  - domain-specific -> domain-independent
  - easy to configure

- Data interface
  - event object
  - related person object
Configuration file

- Structure for each line

**in-built_rule:** `formula “shown_message” inconsistency_level`

- Examples

```plaintext
timelag(intra): birth - marriage > 0 "birth year after marriage year" 1

timelag(inter): main.birth - spouse.birth >= 20 "the age gap between bride and groom is larger than 20" 2

frequency(intra): marriage >= 3 "more than 3 marriages in his/her life" 2
```
### Data interface – Event object

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>distinguish</td>
</tr>
<tr>
<td>type</td>
<td>colour</td>
</tr>
<tr>
<td>event_year</td>
<td>position</td>
</tr>
<tr>
<td>birth_year</td>
<td>detailed information</td>
</tr>
<tr>
<td>description</td>
<td></td>
</tr>
</tbody>
</table>

**Diagram:**
- Year Of Event
- Year Of Birth Of A Person
- Colours represent different types of events:
  - Green: birth
  - Yellow: birth of child
  - Blue: marriage
  - Red: death
  - Black: census (1871)

**Note:** To be displayed here: record 3328-7207021
Data interface – Related person object

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>distinguish</td>
</tr>
<tr>
<td>type</td>
<td>colour</td>
</tr>
<tr>
<td>event_year</td>
<td></td>
</tr>
<tr>
<td>birth_year</td>
<td></td>
</tr>
<tr>
<td>description</td>
<td>detailed information</td>
</tr>
</tbody>
</table>

Year Of Birth Of A Person

Year Of Event

- birth
- birthofchild
- child
- census
- parent
- death
- spouse