PhD Confirmation of Candidature

Interoperable framework to integrate involuntary geospatial data in web based geoportals

Presenter:
Shirish Sharma

Affiliations:

Supervisors:
Assoc Prof Dr Peter Dahlhaus
(CeRDI, Federation University Australia)
Dr Iman Avazpour
(Data61, CSIRO)
Dr Angela Murphy
(CeRDI, Federation University Australia)
Introduction

Main research focus is Data Integration

On an underlying canvas of Smart Cities

“What are the parameters of a smart city?”

- Effective use of technology
- Effective citizen engagement
- Effective governance
- Decent living standards
- Cultural & economic development
- Environment friendly & sustainable
- Geospatial data visualisation
- Involuntary citizen data
Geospatial Data Visualisation

Data presented in visual form boosts cognitive processes thus providing a better ability to understand and analyse information that would otherwise be too complex to discern. (Tam and Song, 2016)

- Web based mapping services
  Proprietary (Google Maps) or VGI (OpenStreetMap)
  Other authoritative, private and community geoportals

- GIS, servers and database technologies
  e.g. QGIS, GeoServer, PostGRE
Geospatial Data Sources

Spatial data on the web are often heterogeneous in nature and as a result are challenging to share between organisational boundaries. (Paul and Ghosh, 2007)

- Authoritative (government) datasets
- Privately owned datasets
- Citizen datasets – voluntary or involuntary
Successful city management depends on reliable availability of data which, owing to an array of user preferences and technologies, must be accessible using interoperable solutions. (Amirian, et. al, 2007)

- **Interoperability**
  The ability of computer systems to exchange and make use of information

- **Open standards**
  OGC standards (WMS, WFS, WPS, GML)
Involuntary Citizen Data

Community feedback can come from "data mining" which, in today’s age of smartphones, are very effective in sensing everyday routine activities within complex social networks. (Batty et al., 2012)

- Harvesting citizen data
  Using relevant algorithms (for example Twitter APIs and RSS)

- Spatial data mining
  Finding spatial patterns from harvested citizen data
  Big area of research that is continually evolving
  An array of mature tools and algorithms available
In recent times, post disaster scenarios across the globe have seen a wide use of social media as a platform for affected citizens to share their data, which in turn aids in the improvement of disaster response and management services. (Goodchild, 2007)

- Geospatial data is widely available
- Geospatial visualisation technologies are mature and evolving
- Data mining technologies are mature and evolving
- Involuntary citizen data has huge potential
- However, this potential is relatively untapped
- A interoperable framework to facilitate the integration of such data into geoportals is imminent
Research Questions

Question 1

What are the best algorithms for harvesting and mining involuntary geospatial data from online sources?
Research Questions

Question 2

How can involuntary geospatial data, mined from the web, be integrated dynamically into web based geoportals?
How can such integration algorithms be combined into an interoperable framework?
Research Significance

Hypothesis

An interoperable framework that facilitates the integration of involuntary geospatial data in web based geoportals enriches available information and, as a result, improves decision making processes for smart cities.
Research Significance

The Use Cases

- Based across a global context of smart cities
- Real life cases with narratives

Example Use Case Scenarios:
- Disaster Management
- Heritage Conservation
- Election Result Reporting
- Road Traffic Reporting
Method - Stage One

What are the best algorithms for harvesting and mining involuntary geospatial data from online sources?

- Data sources
  - Twitter data
  - RSS news feeds

- Mining algorithms
  - Data harvesting
  - Data mining

- Use cases

- Analyse results

- Decision matrix
  - (MCDA)

- Final selection of data mining algorithms
How can involuntary geospatial data, mined from the web, be integrated dynamically into web based geoportals?
Method - Stage Three

How can such integration algorithms be combined into an interoperable framework?

- Easily adaptable to
  - A variety of geoportals
  - An array of online sources

- Will be tested on a variety of smart city based use cases
Conclusion

• Involuntary geospatial data on the web has huge potential
• Integrating harvested citizen data into geoportals is cumbersome
• An interoperable framework shall facilitate such integration
• Data integration solutions are critical in decision making for smart city applications
Thank you!

Questions?

Shirish Sharma
shirish.sharma@federation.edu.au
shirish.sharma@data61.csiro.au
Appendix A

Example Use Case

HUL Ballarat and Visualising Ballarat

- Heritage conservation initiative
- Citizen engagement through web portal and visualising tools
- Visualising Ballarat portal for geospatial visualisation
- Currently hosts authoritative data
- Enriching information using citizen data from social media would enhance the utility of the portal
Appendix B

Stage 1 Decision Matrix – Comparison of best data mining algorithms

Factors for building the decision criteria

• Total volume of data retrieved
• Context based clustering performance (spatial, temporal, lexical and sentimental)
• Context based association performance (evaluation rules e.g. support, confidence and lift)
• Computational time taken
• Ease of implementation and maintenance