Semi-Automated Metadata Detection for Assessing the Credibility of Map Mashup

Sr. Dr. Nurul Hawani Idris
Universiti Teknologi Malaysia (UTM)

Commission No 3

Co-Authors

Prof. Mike J. Jackson, FRICS
Nottingham Geospatial Institute,
University of Nottingham UK

Sr. Mohamad Nor Said (Assoc. Prof.)
Geoinformation Department
Universiti Teknologi Malaysia

Zamri Ismail
Geoinformation Department
Universiti Teknologi Malaysia

Mohamad Ghazali Hashim
Geoinformation Department
Universiti Teknologi Malaysia

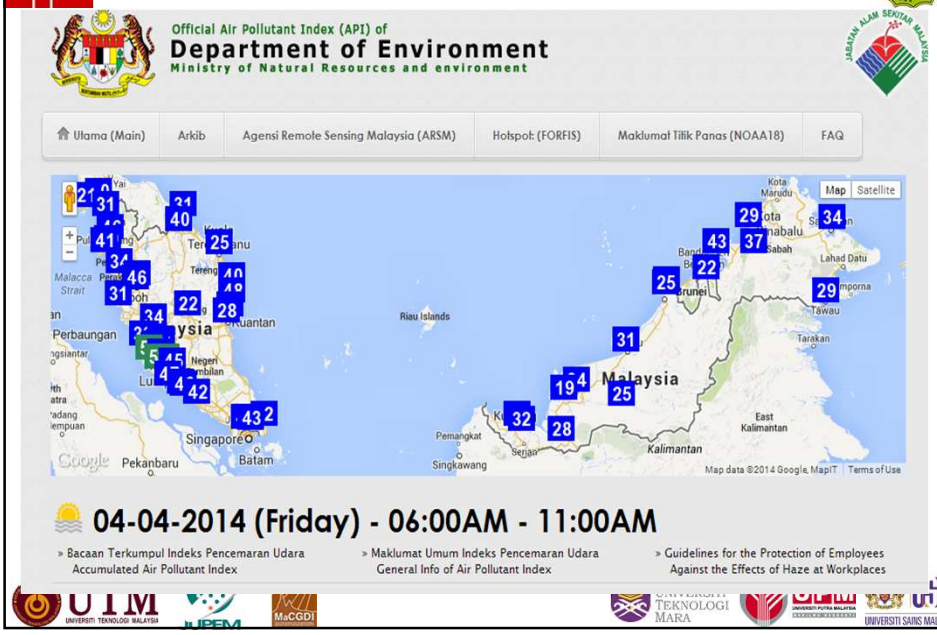
Dr. Mohamad Hafis Izran Ishak
Infocomm Research Alliance
Universiti Teknologi Malaysia

Fathin Nazeera Nazri
MSc. Student
Universiti Teknologi Malaysia

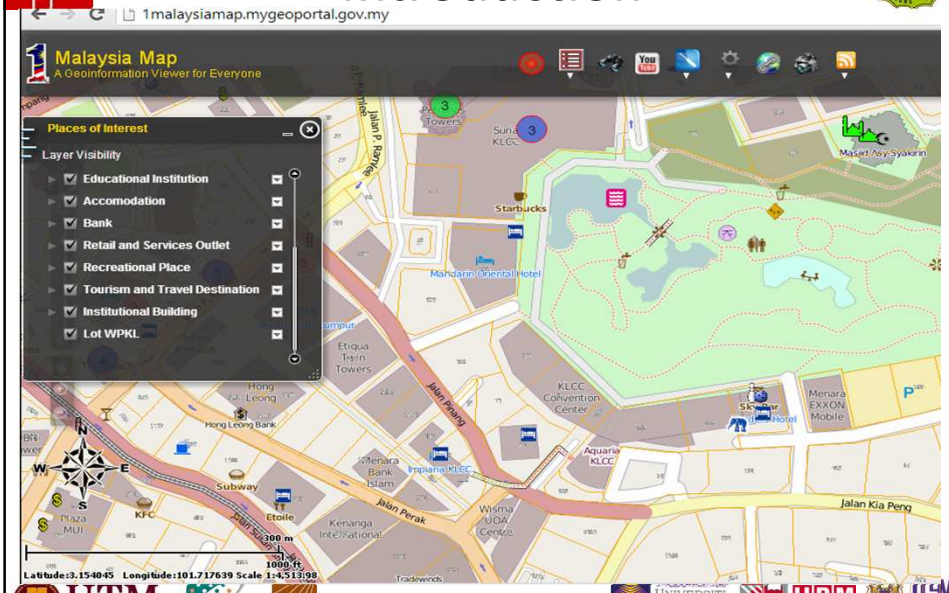
Introduction

- Web 2.0 and the release of Google Map APIs (2005) have made big impact on the culture of mapping
- Neogeography -
  - Neogeography consists of a set of techniques and tools that fall outside the realm of traditional GIS (Turner, 2006)
- Volunteered Geographic Information (VGI)
  - The widespread engagement of large number of private citizen, often with little in the way of formal qualification. They are largely untrained and their actions are always almost voluntary and the results may or may not be accurate (Goodchild, 2007)

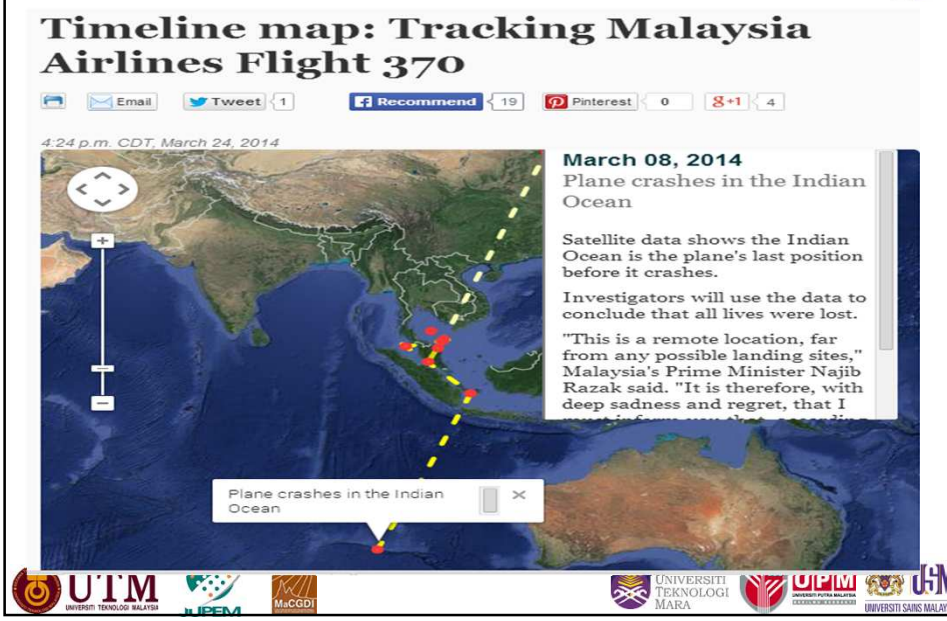
Introduction



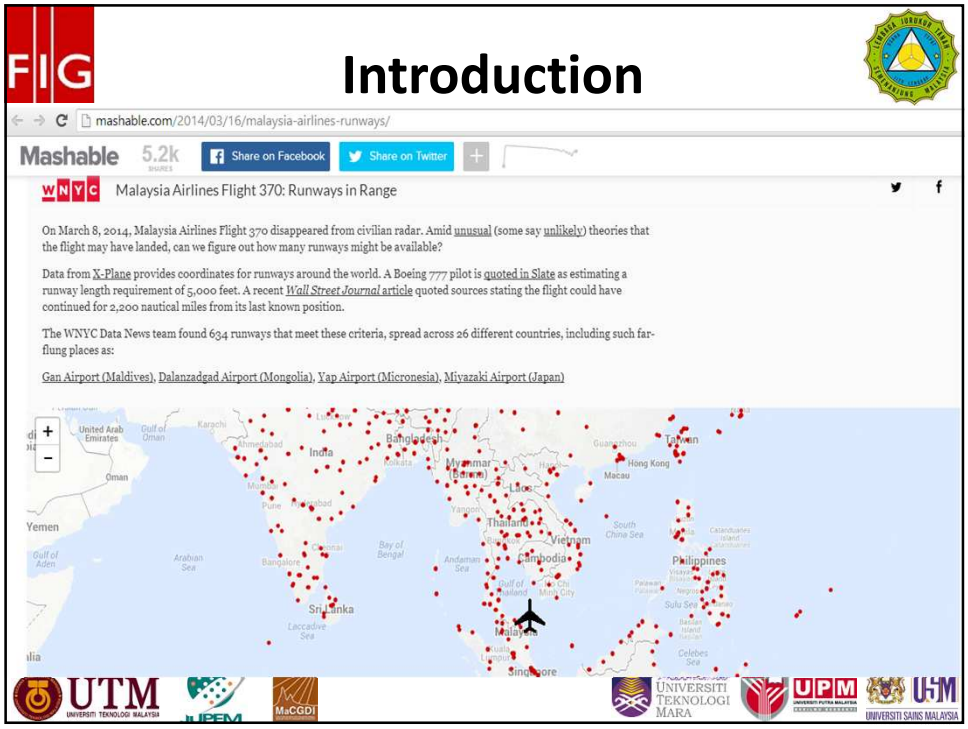
Introduction

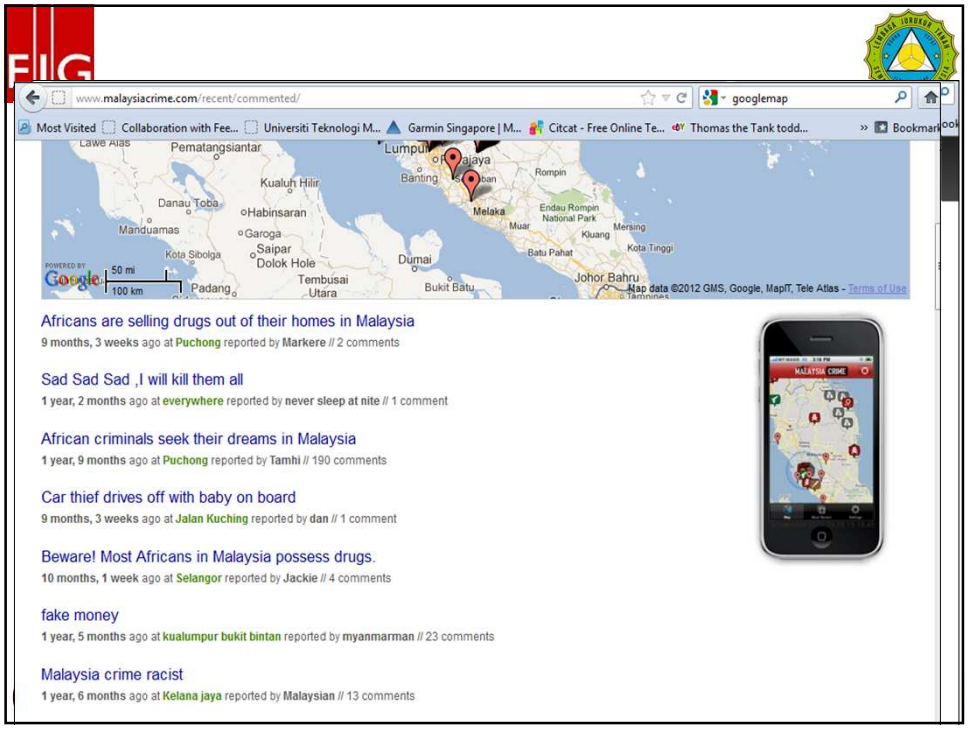


Introduction



Introduction





Research Background

BEFORE

- Professional developers
- Data supplied by governments / commercial organisations
- Standardised procedure – capture, quality control, dissemination

AFTER

- Non-professionals / citizens
- Personal/localised data/various sources data
- Low cost tools
- Low programming skills
- Low experiences/prior knowledge – mapping design/data capture

Issues – truth, quality, accuracy !

Produce own maps

Previous Studies

1. Simply tagging metadata on map mashup –low influence on users' perceived credibility
Nurul Hawani Idris, Mike J.Jackson & Robert J. Abrahart (2011), Map Mashups: What looks good must be good?, GIS Research UK Conference (GISRUK), Portsmouth UK, 27-29 April 2011

2.High influence of Credibility Labelling on users perceived credibility– Colour Coded Traffic Light labelling
Nurul Hawani Idris, Mike J.Jackson & Robert J. Abrahart (2011), Credibility Labelling in Map Mashup Design: Influence on Users' Judgement, EuroSDR/ISPRS Workshop on Web Cartography, Lund Sweden, 5th-6th May 2011

3. Colour Coded Traffic Light Labelling: An Approach to Assist Users in Judging Data Credibility in Map Mashup Applications. Nurul Hawani Idris, Mike J. Jackson, and Robert J. Abrahart (2011). In Cidalia C. Fonte, Luisa Goncalves & Gil Goncalves (Eds), Proceedings of the 7th International Symposium on Spatial Data Quality (pp. 201-206). Coimbra Portugal: INESC Coimbra

Previous studies

- Successful implementation of 'seal of approvals' in other domains

- Visual quality indicators on GIS maps (Devillers et al. (2002)) - a response to the difficulty of communicating metadata to professional and lay users. Devillers et al. (2007) - a practical model to implement a quality rating system in GIS for use by experts to give advice about the quality of a dataset.

Problem Statements

- Issue 1 : the insufficient of simply tagging metadata on map mash up
- Issue 2: A crucial need of a gatekeeper control scheme on online maps, particularly on map mashups

Purpose of study

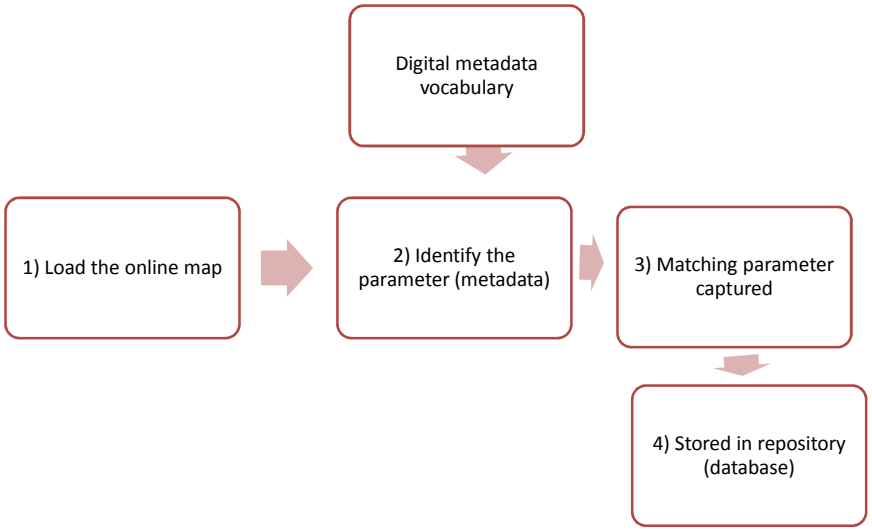
- To demonstrate the ability of web crawler (scraper) to detect metadata criteria towards the implementation of credibility labelling on map mashup

Related studies

Metadata component

1. Rule based line classification – a set of rules to detect metadata concept (Wang and Richard 2007)
2. Supervised machine learning – similar technique to automate recognition and filtering spam emails. A good corpus of training data from a huge collection of metadata used in mapping application. Patterns of texts to specific categories of metadata used to detect and rate the indicators (Gaudinat et al., 2007)

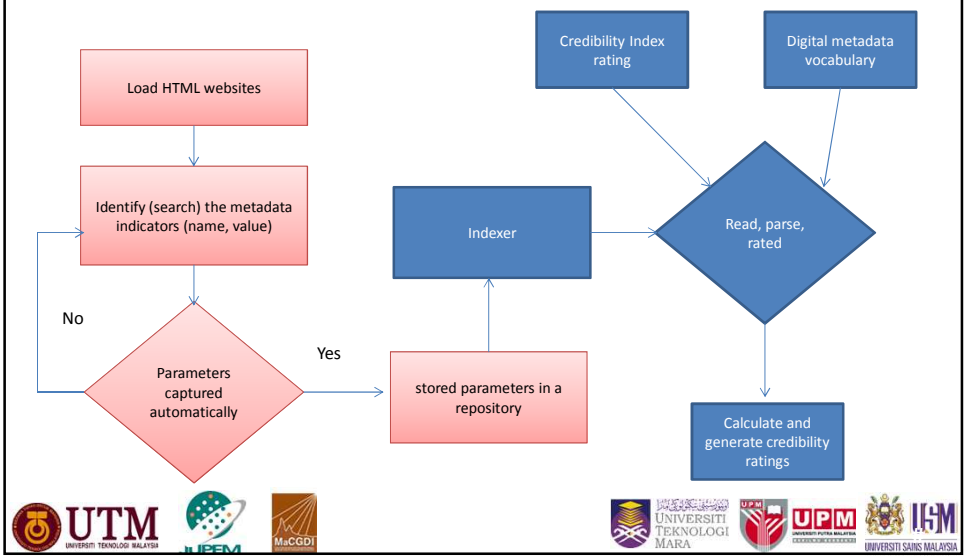
Metadata detection algorithm



Digital Metadata vocabulary

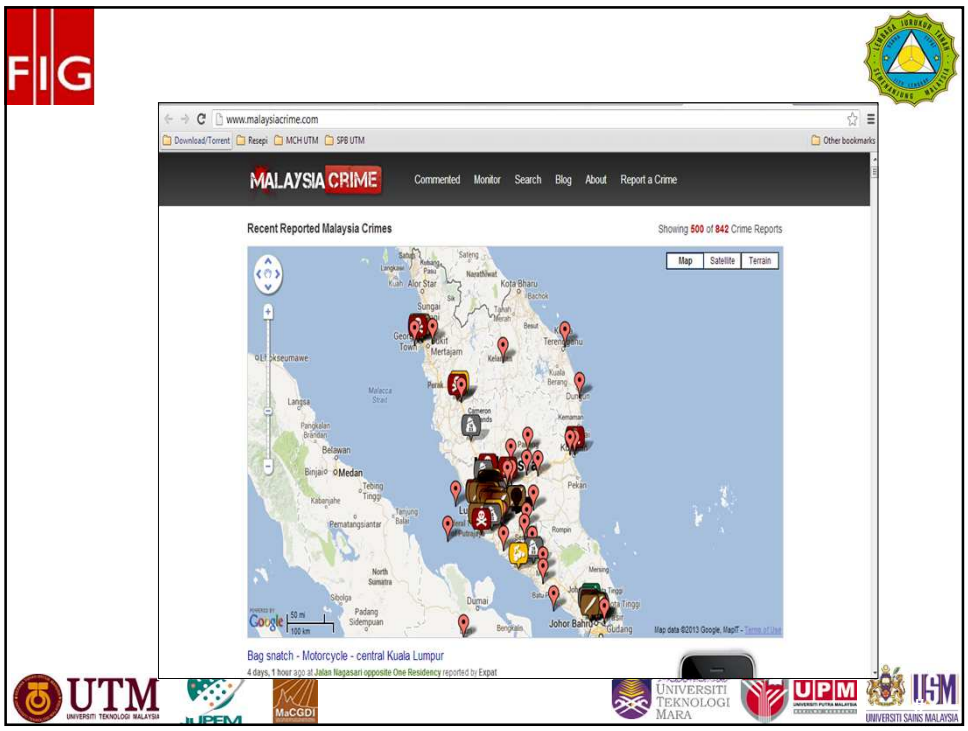
Name (metadata indicator)	Common code expressions	Possible values (keywords)
Currency – publication date	structured	Date, published data
Currency – last updated date	structured	Date, last updated date
Currency for the period under consideration	unstructured	Timeline, temporal, period
Disclose author/creator map mashup	structured	Published by, created by, developed by, author,
Contact information	structured/ unstructured	About us, email, organization, telephone (phone), fax number,
Disclose identity of background data	unstructured	Copyright
Disclose identity of foreground data supplier	structured	Source, supplied by, origin,
Reputation domain URL	structured	URL http://
Affiliation/association	structured	Hyperlinks
Seal approval	unstructured	Approved by,
Sponsorship	unstructured	Sponsored by, funded by,
Mission/purposes/ motive	unstructured	About us, Objectives, purposes, goal

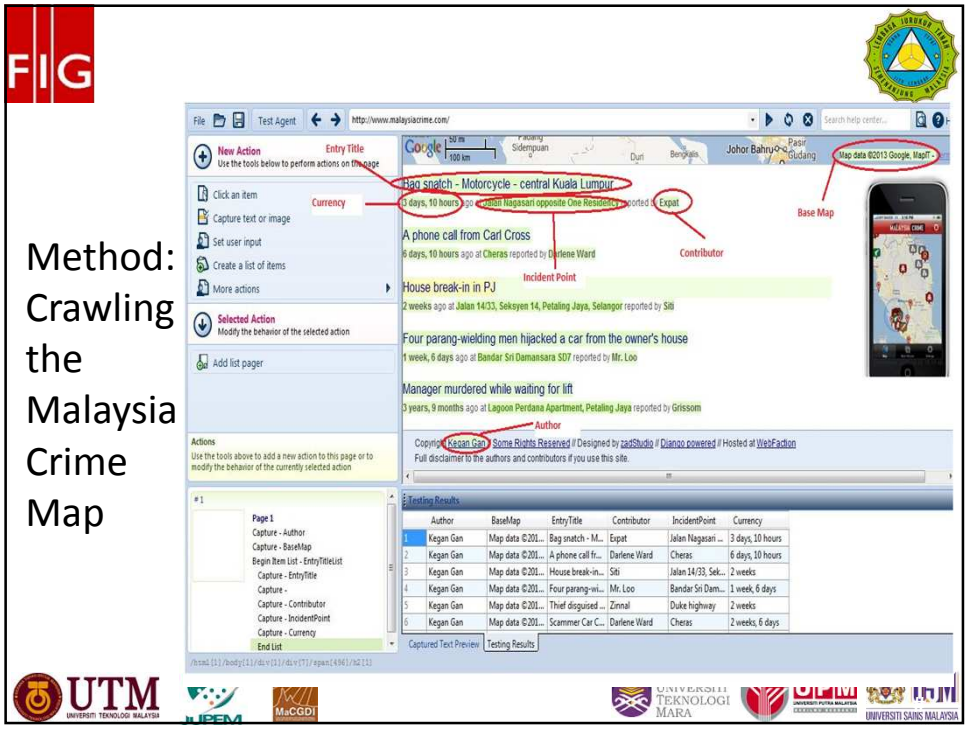
Framework of credibility rating index



Case study

Map title	Parameters values					
	Domain URL	currency	Identity of author	Disclose foreground data supplier	Disclose background data supplier	Affiliation / association
Malaysia Crime Map Mashup [1]	http://www.malaysiacrime.com/	age	Copyright	Reported by	Map data	-
Tropical Cyclone Phailin Map Mashup [2]	http://google.org/crisismap/ 2013-phailin	Last updated	Published by	Source	Map Data	-
Mapping South Africa with dot distribution) [3]	http://dotmap.adrianfrith.com/		By	Data from	Map data	-
2013 Colorado Floods Map Mashup [4]	http://google.org/crisismap/2013-boulder-floods	Last updated	Published by	Source	Map data	-
The Ellis Act Map Mashup [5]	http://www.antievictionmappingproject.net/ellis.html	-	Created by	-	Powered by	-
London Cycling Census Map Mashup [6]	http://casa.oobrien.com/traffic/	Direct detection	Produced by	Source	Basemap data	-
UTM Campus Map [7]	http://web.utm.my/campus_map/	Since	Contact	Copyright	Imagery	Direct detection
Oak Mapper Map Mahup [8]	http://www.oakmapper.org	Recent Submissions	By	-	-	Links
Typhoon Pablo (Bopha) [9]	http://google.org/crisismap/2012-pablo	Last updated	Published by	Source	Map data	-
Operational UK Renewable Electricity Sites Map Mashup [10]	http://ukdataexplorer.com/renewables/	Downloaded on	Published by	-	Powered by	-


www.malaysiacrime.com
MALAYSIA CRIME
Commented
Monitor
Search
Blog
About
Report a Crime
Recent Reported Malaysia Crimes
Showing 500 of 842 Crime Reports
Map
Satellite
Terrain
Bag snatch - Motorcycle - central Kuala Lumpur
4 days, 1 hour ago at Jalan Nagasari opposite One Residency reported by Expat


Method: Crawling the Malaysia Crime Map
New Action
Use the tools below to perform actions on the page
Entry Title
Currency
Expat
Base Map
Contributor
Incident Point
Author
Bag snatch - Motorcycle - central Kuala Lumpur
A phone call from Carl Cross
House break-in in PJ
Four parang-wielding men hijacked a car from the owner's house
Manager murdered while waiting for lift
Testing Results

Results

Author (creator) of mashup	Point cases	Data Contributor	Currency	Base Map	Incident Point
structured	structured	structured	structured	Unstructured	structured
Kegan Gan	Bag snatch - Motorcycle -	Expat	3 days, 9 hours	Map data Â©2013 Google, MapIT -	Jalan Nagasari
Kegan Gan	A phone call from Carl Cross	Darlene	6 days, 9 hours	Map data Â©2013 Google, MapIT -	Cheras
Kegan Gan	House break-in in PJ	Siti	2 weeks	Map data Â©2013 Google, MapIT -	Jalan 14/33, Seksyen 14

Detection accuracies

Map ID [See Table 3]	Values (count)				
	currency	Identity of author	Disclose foreground data supplier	Disclose background data supplier	Affiliation / association
[1]	500/500	1/1	500/500	1/1	-
[2]	2/3 *one date not captured	1/1	6/7 *one source incorrectly captured	1/1	-
[3]	-	1/1	1/1	1/3	-
[4]	3/4 *one date not captured	1/1	8/9 *one source incorrectly captured	1/1	-
[5]	-	1/1	-	3/3	-
[6]	1/1	1/1	1/1	1/2*one data supplier not captured	-
[7]	1/1	1/1	1/1	1/1	1/7
[8]	-	1/1	-	-	5/5
[9]	1/3	1/1	14/14	1/1	-
[10]	1/1	3/3	1/1	1/1	-

Remarks

This study demonstrates the ability of a web crawler to detect metadata indicators and values before it can be indexed to generate a credibility rating of a map mashup application.

The possible techniques to detect the structured and unstructured metadata by using a web crawler

The results indicate the efficiency of metadata indicators and values to be detected in an automated manner.

This could reduce the cost of labor and additional tasks required to generate a list of metadata indicators before it could be indexed to generate credibility ratings.

Remarks

The relevance of this study to Surveying Community

- Current tools available for the automated creation of metadata for the assessment of credibility and trust are not yet adequate for professional purposes.
- As crowd-data is increasingly seen as providing a significant data contribution to authoritative data sets and especially spatial data infrastructures (SDI's)
- This study represent a step towards a formal quality assessment framework within which crowd-data may be accompanied by metadata

Limitations

- The ideal way for the web crawler (harvester) to assess the credibility of map mashup is through a fully intelligent automated mechanism
- The current web crawler tools used in this study still require human assistance.
- The less intelligent of the web crawler tools used in this study to learn the similar expression of metadata

Future works

- To develop a digital metadata vocabulary – large sample of dataset
- To develop a credibility rating indexer component in order to support the framework of assessing credibility of map mashup application in semi-automated manner.

Acknowledgement

- Financial assistance from Malaysian Land Surveyors Board (LJT).
- This project was funded by the Universiti Teknologi Malaysia (UTM) and Ministry of Higher Education Malaysia (MOHE) under Research University Grant vote no. 08J94.

Thank you for your attention!

Corresponding author:
hawani@utm.my
Geoinformation Department
Faculty Geoinformation and Real Estate
Universiti Teknologi Malaysia (UTM)