To trust, or not to trust:
Highlighting the need for data provenance in mobile apps for smart cities

Mikel Emaldi[1], Oscar Peña[1], **Jon Lázaro**[1], Diego López-de-Ipiña[1]
Sacha Vanhecke[2], Erik Mannens[2]

[1] DeustoTech - Deusto Institute of Technology, University of Deusto

[2] iMinds - Multimedia Lab, Ghent University

University of Deusto
www.deusto.es

What is a *Smart City?*

A city that makes available all the services and applications enabled by ICT to citizens, companies and authorities that are part of a city's system.

It aims to increase all citizens' quality of life and improve the efficiency and quality of the services provided by governing entities and businesses.

Smartphones in Smart Cities

Smartphones are an essential part of Smart Cities

- Enable a way to access smart cities' services

- Enable the participation of the citizens in city governance

- Ease the way users share they own opinion, data...

biased information

uncompleteness

incorrect
data

contradictions

outdated
information

conflicts

partial data

biased information

uncompleteness

incorrect
data

contradictions

outdated
information

conflicts

partial data

Solution? Trust!

We need trust mechanisms to...

- Measure if submitted data is reliable.

- Identify a given user and measure his/her reputation.

Solution? Trust!

We need trust mechanisms to...

- Measure if submitted data is reliable.

- Identify a given user and measure his/her reputation.

Build user-centric smartphone apps for smart cities that keep provenance information

Solution? Trust!

We need trust mechanisms to...

- Measure if submitted data is reliable.

- Identify a given user and measure his/her reputation.

Build user-centric smartphone apps for smart cities that keep provenance information

Provide trust mechanisms that help authorities and users deciding if the data is reliable or not

IES Cities: Internet Enabled Services for cities across Europe

Motivations:

- Citizens must be heard & empowered

- The information gathered and provided by both cities and citizens must be linked and processed

Goals:

- Create a multi-device dataset and application marketplace based on standard and accessible web technologies, exploiting data shared between citizens and councils, and providing an enhanced experience to municipalities

Use case: 311 Bilbao

Consuming & reporting of complaints concerning public infrastructures
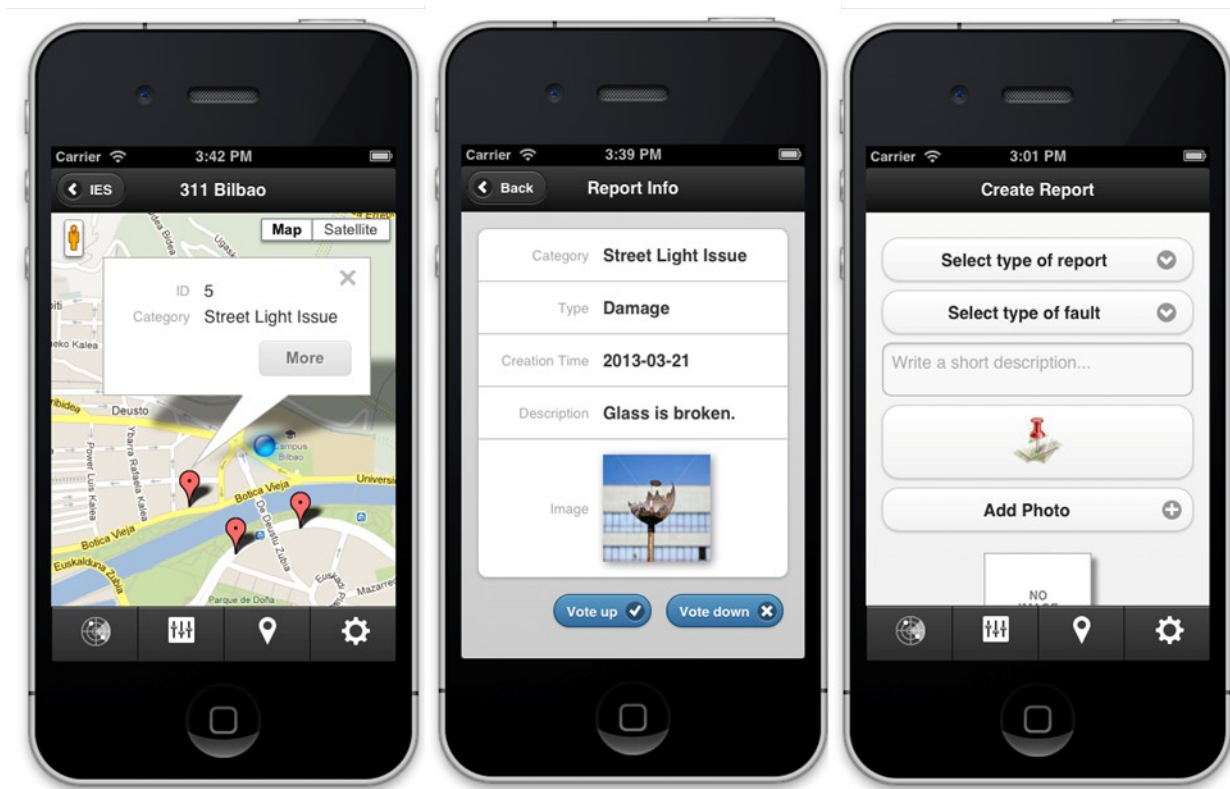
Introduction

Related work

Semantic representation of provenance

Provenance-based trust model

Conclusions and future work
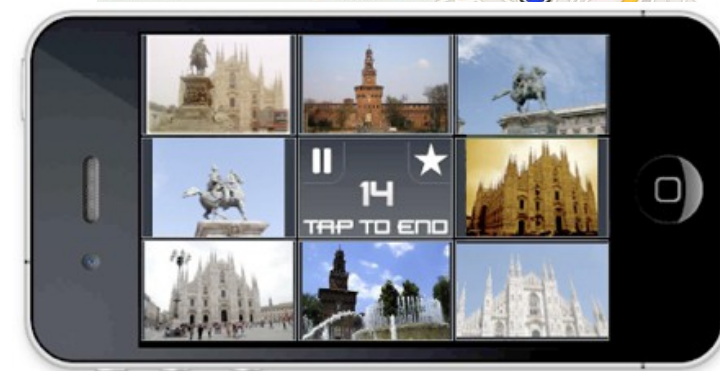
Apps for Smart Cities

Urbanopoly

- Human Computation + Linked Data + Gamification to verify and correct data about POIs

Urban Match

- User compares photos to link smart-cities' datasets

csxPOI

- Collaborative creation, edition and share of semantic POIs

Apps for Smart Cities

Different ways to "trust" user data:

- Popularity (Urbanopoly)

- User historical reputation (Urbanopoly)

- Clustering of properties of reports (Urban Match)

- Combine known real and false data (csxPOI)

Apps for Smart Cities

Different ways to "trust" user data:

- Popularity (Urbanopoly)

- User historical reputation (Urbanopoly)

- Clustering of properties of reports (Urban Match)

- Combine known real and false data (csxPOI)

Our proposal: Keep provenance information and use it in trust algoithms

Outline

Introduction

Related work

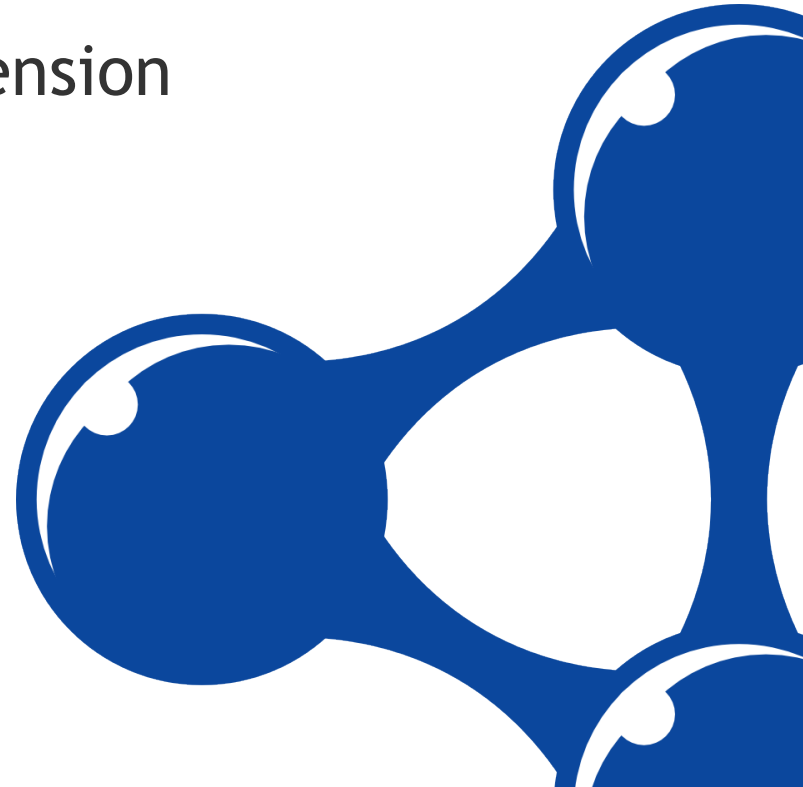Semantic representation of provenance

Provenance-based trust model

Conclusions and future work

Representation of provenance

W3C PROV Data Model

- RDF model for provenance representation

- PROV-O ontology

- Extensible: Uncertainty extension

Representation of a report

User John Doe sends
a report

```
entity(:report_23456, [prov:value="The
paper bin is broken"])

wasAttributedTo(:report_23456, :jdoe)

agent(:jdoe,
[ prov:type='prov:Person',
foaf:name="John Doe",
foaf:mbox='<mailto:jdoe@example.org>'
])

wasGeneratedBy(:report_23456,
:reportActivity_23456)

activity(:reportActivity_23456,
2013-07-22T01:01:01, geo:lat=43.25,
geo:long=-2.93)
```

Representation of a report

User John Doe sends
a report

Report

```
entity(:report_23456, [prov:value="The
paper bin is broken"])

wasAttributedTo(:report_23456, :jdoe)

agent(:jdoe,
[ prov:type='prov:Person',
foaf:name="John Doe",
foaf:mbox='<mailto:jdoe@example.org>'
])

wasGeneratedBy(:report_23456,
:reportActivity_23456)

activity(:reportActivity_23456,
2013-07-22T01:01:01, geo:lat=43.25,
geo:long=-2.93)
```

Representation of a report

User John Doe sends
a report

Report

User

```
entity(:report_23456, [prov:value="The
paper bin is broken"])
```

```
wasAttributedTo(:report_23456, :jdoe)

agent(:jdoe,
[ prov:type='prov:Person',
foaf:name="John Doe",
foaf:mbox='<mailto:jdoe@example.org>'
])
```

```
wasGeneratedBy(:report_23456,
:reportActivity_23456)
```

```
activity(:reportActivity_23456,
2013-07-22T01:01:01, geo:lat=43.25,
geo:long=-2.93)
```

Representation of a report

User John Doe sends
a report

Report

User

Context

```
entity(:report_23456, [prov:value="The
paper bin is broken"])
```

```
wasAttributedTo(:report_23456, :jdoe)

agent(:jdoe,
[ prov:type='prov:Person',
foaf:name="John Doe",
foaf:mbox='<mailto:jdoe@example.org>'
])
```

```
wasGeneratedBy(:report_23456,
:reportActivity_23456)

activity(:reportActivity_23456,
2013-07-22T01:01:01, geo:lat=43.25,
geo:long=-2.93)
```

Representation of a revision report

A worker of the council invalidates John's report

```
wasInvalidatedBy(:report_23456, :invActivity_639,
2013-07-22T03:05:03)

activity(:invActivity_639, 2013-07-22T02:58:01,
2013-07-22T03:04:47)

wasAssociatedWith(:invActivity_639, :jane)

entity(:report_23457, [ prov:value="It is incorrect,
another paper bin has replaced the old one, but 2
meters beyond" ])

wasDerivedFrom(:report_23457, :report_23456,
:invActivity_639, -, -,
[ prov:type='prov:Revision' ])

wasAttributedTo(:report_23457, :jane)

agent(:jane, [ prov:type='prov:Person',
foaf:name="Jane",
foaf:mbox='<mailto:jane@bilbao.iescities.org>' ])

actedOnBehalfOf(:jane, :bilbao_city_council)

agent(:bilbao_city_concil,
[ prov:type='prov:Organization', foaf:name="Bilbao
City Council"])
```

Representation of a revision report

A worker of the council invalidates John's report

Invalidation

```
wasInvalidatedBy(:report_23456, :invActivity_639,
2013-07-22T03:05:03)

activity(:invActivity_639, 2013-07-22T02:58:01,
2013-07-22T03:04:47)

wasAssociatedWith(:invActivity_639, :jane)
```

```
entity(:report_23457, [ prov:value="It is incorrect,
another paper bin has replaced the old one, but 2
meters beyond" ])

wasDerivedFrom(:report_23457, :report_23456,
:invActivity_639, -, -,
[ prov:type='prov:Revision' ])

wasAttributedTo(:report_23457, :jane)

agent(:jane, [ prov:type='prov:Person',
foaf:name="Jane",
foaf:mbox='<mailto:jane@bilbao.iescities.org>' ])

actedOnBehalfOf(:jane, :bilbao_city_council)

agent(:bilbao_city_concil,
[ prov:type='prov:Organization', foaf:name="Bilbao
City Council"])
```

Representation of a revision report

A worker of the council invalidates John's report

Invalidation

Revision report

```
wasInvalidatedBy(:report_23456, :invActivity_639,
2013-07-22T03:05:03)

activity(:invActivity_639, 2013-07-22T02:58:01,
2013-07-22T03:04:47)

wasAssociatedWith(:invActivity_639, :jane)
```

```
entity(:report_23457, [ prov:value="It is incorrect,
another paper bin has replaced the old one, but 2
meters beyond" ])

wasDerivedFrom(:report_23457, :report_23456,
:invActivity_639, -, -,
[ prov:type='prov:Revision' ])

wasAttributedTo(:report_23457, :jane)
```

```
agent(:jane, [ prov:type='prov:Person',
foaf:name="Jane",
foaf:mbox='<mailto:jane@bilbao.iescities.org>' ])

actedOnBehalfOf(:jane, :bilbao_city_council)

agent(:bilbao_city_concil,
[ prov:type='prov:Organization', foaf:name="Bilbao
City Council"])
```

Representation of a revision report

A worker of the council invalidates John's report

Invalidation

Revision report

Worker

```
wasInvalidatedBy(:report_23456, :invActivity_639,
2013-07-22T03:05:03)

activity(:invActivity_639, 2013-07-22T02:58:01,
2013-07-22T03:04:47)

wasAssociatedWith(:invActivity_639, :jane)
```

```
entity(:report_23457, [ prov:value="It is incorrect,
another paper bin has replaced the old one, but 2
meters beyond" ])

wasDerivedFrom(:report_23457, :report_23456,
:invActivity_639, -, -,
[ prov:type='prov:Revision' ])

wasAttributedTo(:report_23457, :jane)
```

```
agent(:jane, [ prov:type='prov:Person',
foaf:name="Jane",
foaf:mbox='<mailto:jane@bilbao.iescities.org>' ])
```

```
actedOnBehalfOf(:jane, :bilbao_city_council)

agent(:bilbao_city_concil,
[ prov:type='prov:Organization', foaf:name="Bilbao
City Council"])
```

Representation of a revision report

A worker of the council invalidates John's report

Invalidation

Revision report

Worker

Council

```
wasInvalidatedBy(:report_23456, :invActivity_639,
2013-07-22T03:05:03)

activity(:invActivity_639, 2013-07-22T02:58:01,
2013-07-22T03:04:47)

wasAssociatedWith(:invActivity_639, :jane)
```

```
entity(:report_23457, [ prov:value="It is incorrect,
another paper bin has replaced the old one, but 2
meters beyond" ])

wasDerivedFrom(:report_23457, :report_23456,
:invActivity_639, -, -,
[ prov:type='prov:Revision' ])

wasAttributedTo(:report_23457, :jane)
```

```
agent(:jane, [ prov:type='prov:Person',
foaf:name="Jane",
foaf:mbox='<mailto:jane@bilbao.iescities.org>' ])
```

```
actedOnBehalfOf(:jane, :bilbao_city_council)

agent(:bilbao_city_concil,
[ prov:type='prov:Organization', foaf:name="Bilbao
City Council"])
```

Introduction

Related work

Semantic representation of provenance

Provenance-based trust model

Conclusions and future work

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency…

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust parameters

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust parameters

Relevance of a given property [0,1]

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...
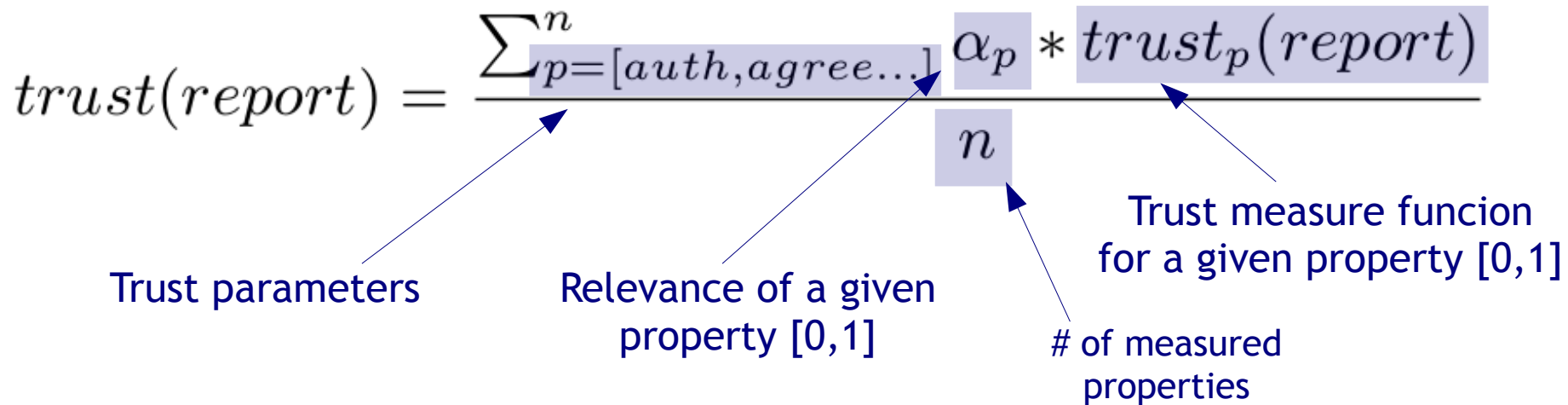
$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust parameters

Relevance of a given property [0,1]

Trust measure funcion for a given property [0,1]

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust parameters

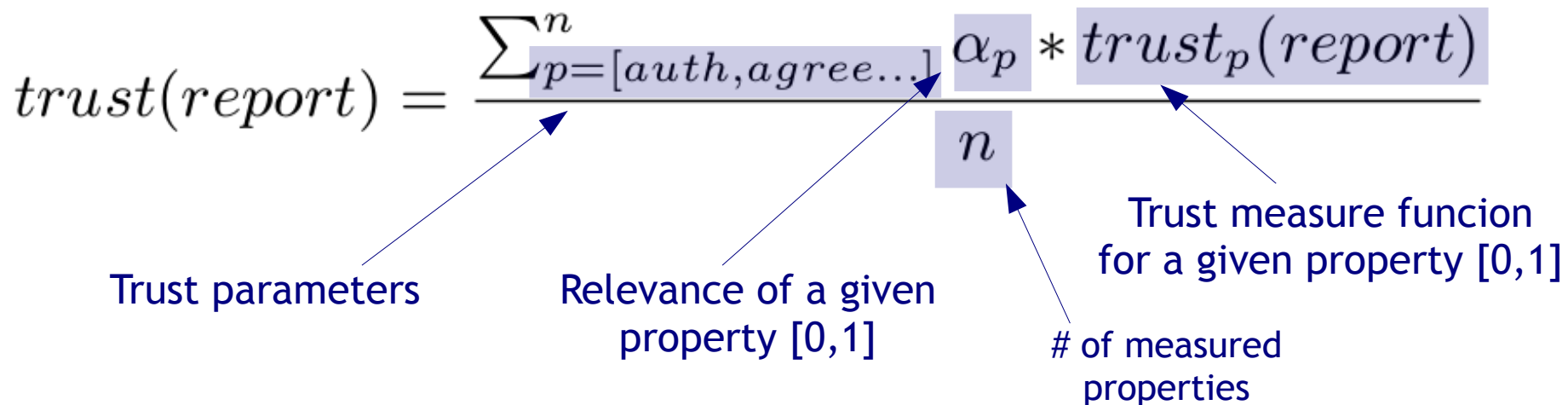Relevance of a given property [0,1]

# of measured properties

Trust measure funcion for a given property [0,1]

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...

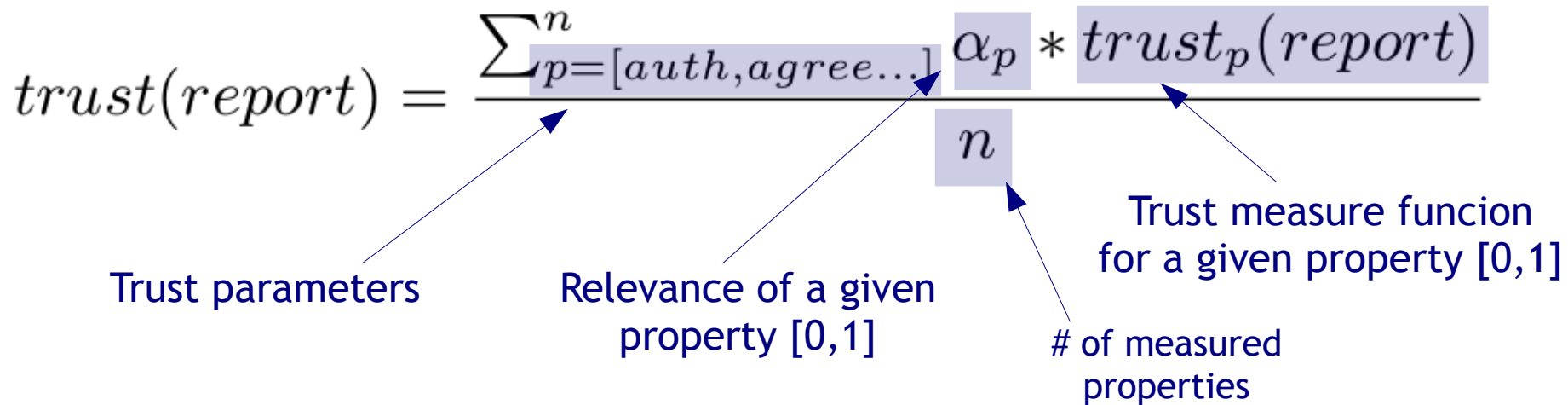$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust parameters

Relevance of a given property [0,1]

# of measured properties

Trust measure funcion for a given property [0,1]

- Generic and flexible model

- and defined by developers or councils

Trust model

Gil et al. identify 19 parameters on how to determine trust in web content: authority, popularity, recency...

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n}$$

Trust parameters

Relevance of a given property [0,1]

# of measured properties

Trust measure funcion for a given property [0,1]

- Generic and flexible model
- α_p and $trust_p$ defined by developers or councils

Authority

Was the data created by an authority in a given context?

Authority

Was the data created by an authority in a given context?

$$trust_{authority} = \begin{cases} 0 & if\, user \neq authority \\ 1 & if\, user = authority \end{cases}$$

In Bilbao 311 and for John Doe, an SPARQL query like…

```
ASK { :jdoe prov:actedOnBehalfOf :bilbao_city_concil }
```

would have no results, so

$$trust_{authority} = 0$$

Popularity

Is the piece of information referenced by other users?

Popularity

Is the piece of information referenced by other users?

In Bilbao 311, we measure *popularity* in terms of visits that a report receive:

$$trust_{popularity} = \frac{visits_{report}}{visits_{open\ reports}}$$

Given John Does report has 100 visits and there has been 550 visits to all open reports...

$$trust_{popularity} = \frac{100}{550} = 0.18$$

Recommendation

Is the data relevant for other users?

Recommendation

Is the data relevant for other users?

In Bilbao 311, we measure *recommendation* with a simple voting system (+1/-1):

$$trust_{recommendation} = \frac{positive\ votes_{report}}{total\ votes_{report}}$$

Given John Does report has 10 of 12 positive votes...

$$trust_{recommendation} = \frac{10}{12} = 0.83$$

Provenance / Reputation

Can I trust the user who generated the data? What's his/her reputation?

Provenance / Reputation

Can I trust the user who generated the data? What's his/her reputation?

In Bilbao 311, we use the historical reputation algorithm by *Ceolin et al.* 3 steps:

1. *Evidence selection*: Get every report of the user.

2. *Evidence weighting*: Apply *recommendation* function for every report, taking in account invalidation / validation.

3. *Evidence aggregation*: Subjective logic algorithm

$$trust_{provenance} = 0.65$$

Recency / Timeliness

Is the data up-to-date?

Recency / Timeliness

Is the data up-to-date?

In Bilbao 311, we measure *recency* with the model by *Hartig et al.*

$$trust_{recency} = (max(1 - \frac{currency}{volatility}), 0)^{sensitivity}$$

where...

$$currency = DeliveryTime - InputTime$$

$$volatility = ExpiryTime - InputTime$$

$$sensitivity = \begin{cases} high & if\ report\ constantly\ updated \\ low & if\ report\ not\ changing \end{cases}$$

Recency / Timeliness

$$currency = 1day$$

$$volatility = 7day$$

$$sensitivity = 1$$

Recency / Timeliness

Given...

$$currency = 1day$$

$$volatility = 7day$$

$$sensitivity = 1$$

Recency / Timeliness

Given...

$$currency = 1day$$

$$volatility = 7day$$

$$sensitivity = 1$$

Then...

$$trust_{recency} = (max(1 - \frac{1\ day}{7\ day}), 0)^1 = max(0.86, 0) = 0.86$$

Other trust factors

The model is flexible enough to include other factors depending on the context of the application.

e.g. in Bilbao 311, the distance between reporting and reported places is a key trust factor:

$$trust_{distance} = \frac{1}{geodistance(loc_{report}, loc_{reportedplace})}$$

Given John Doe is very close to reporting place…

$$trust_{distance} = 0.95$$

Results of the trust model

After applying our model we will get a trust value between 0 and 1.

Results of the trust model

After applying our model we will get a trust value between 0 and 1.

If we assume that α_p is equal for every property, John's report would have a trust value of...

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n} = \frac{0 + 0.18 + 0.83 + 0.65 + 0.86 + 0.95}{6} = 0.58$$

Results of the trust model

After applying our model we will get a trust value between 0 and 1.

If we assume that α_p is equal for every property, John's report would have a trust value of...

$$trust(report) = \frac{\sum_{p=[auth,agree...]}^{n} \alpha_p * trust_p(report)}{n} = \frac{0 + 0.18 + 0.83 + 0.65 + 0.86 + 0.95}{6} = 0.58$$

That would be represented in the provenance graph as:

```
:report_23456 up:contentConfidence "0.58"
```

Enabling SPARQL queries like:

```
SELECT ?report WHERE {
    ?report up:contentConfidence ?confidence .
    FILTER (?confidence < 0.5)
}
```

Outline

Introduction

Related work

Semantic representation of provenance

Provenance-based trust model

Conclusions and future work

Conclusions

- An approach that allow to represent and evaluate the provenance of user-submitted data in IES Cities' platform.

- Proposed metrics measure data trustworthiness level, providing an extra confidence layer in the project's framework.

- Flexible model allowing developers and admins to develop own functions and alphas.

- Using a model based on RDF makes simple the publication and querying (trough SPARQL) of these data over the web.

Future work

- Work-in-progress: Finish the implementation of the trust layer in IES Cities platform.

- Evaluation and validation of the proposed model against other implementations using RDF provenance models.

- Research the posibility of adding new relevant metrics.

Bibliography

I. Celino, S. Contessa, M. Corubolo, D. Dell'Aglio, E. D. Valle, S. Fumeo, and T. Krüger. *UrbanMatch - linking and improving smart cities data*

Y. Gil and D. Artz. *Towards content trust of web resources*

O. Hartig and J. Zhao. *Using web data provenance for quality assessment*

D. Ceolin, P. T. Groth, W. R. van Hage, A. Nottamkandath, and W. Fokkink. *Trust evaluation through user reputation and provenance analysis*

M. Braun, A. Scherp and S. Staab. *Collaborative creation of semantic points of interest as linked data on the mobile phone*

I. Celino, D. Cerizza, S. Contessa, M. Corubolo, D. DellAglio, E. D. Valle, and S. Fumeo. *Urbanopoly - a social and location-based game with a purpose to crowdsource your urban data.*

Thank you

Jon Lázaro – jlazaro@deusto.es
Mikel Emaldi – memaldi@deusto.es
Oscar Peña – oscar.pena@deusto.es
Diego López-de-Ipiña – dipina@deusto.es
Sacha Vanhecke - sacha.vanhecke@ugent.be
Erik Mannens- erik.mannens@ugent.be

University of Deusto
www.deusto.es