Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Semantic integration of heterogeneous big data sources

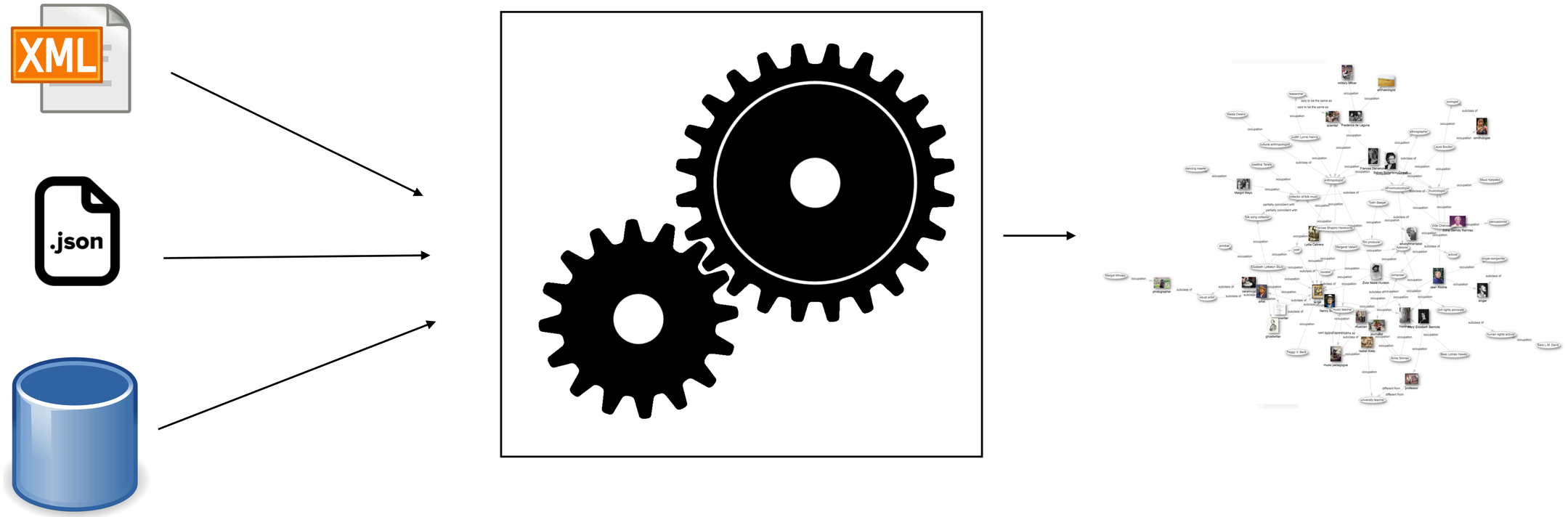
Herminio García González

Supervised by Prof. José Emilio Labra Gayo & Prof. Juan Manuel Cueva Lovelle

Introduction & Motivation

What is data integration?

- Data integration is the problem of aggregating different sources so they can be used through a single interface



(Photos extracted from Wikimedia Commons and thenounproject.com)

Introduction & Motivation

Why data integration?

- Data is growing and it’s becoming difficult to process and handle
  - Advancements in:
    - Artificial Intelligence
    - Internet of Things
    - Big data
- However, still: Information silos

Introduction & Motivation

OK, let’s integrate but how?

- Need for an integration format and environment
  - Links between datasets
  - Unambiguous entities
  - Defined semantics

} **Semantic Web**

Introduction & Motivation

Why the Semantic Web?

- Our advocated target format and environment
  - Interlinking foundation
  - IRI identifiers
  - Ontologies and vocabularies
- RDF compositional property (merging for free)

Introduction & Motivation

Data validation, why?

- Standardisation
- Normalisation
- Integrity
- Thus, more confidence
- Transfer validation qualities to integrated data

Introduction & Motivation

Data validation in data integration, how?

- Schema alignments
- Shape Expressions (ShEx)
- Shapes Constraint Language (SHACL)
- Implications?

Introduction & Motivation

And this is important because…?

- Two use cases
- E-Learning
- Digital Humanities
- What can Semantic Web and data integration make for other fields?

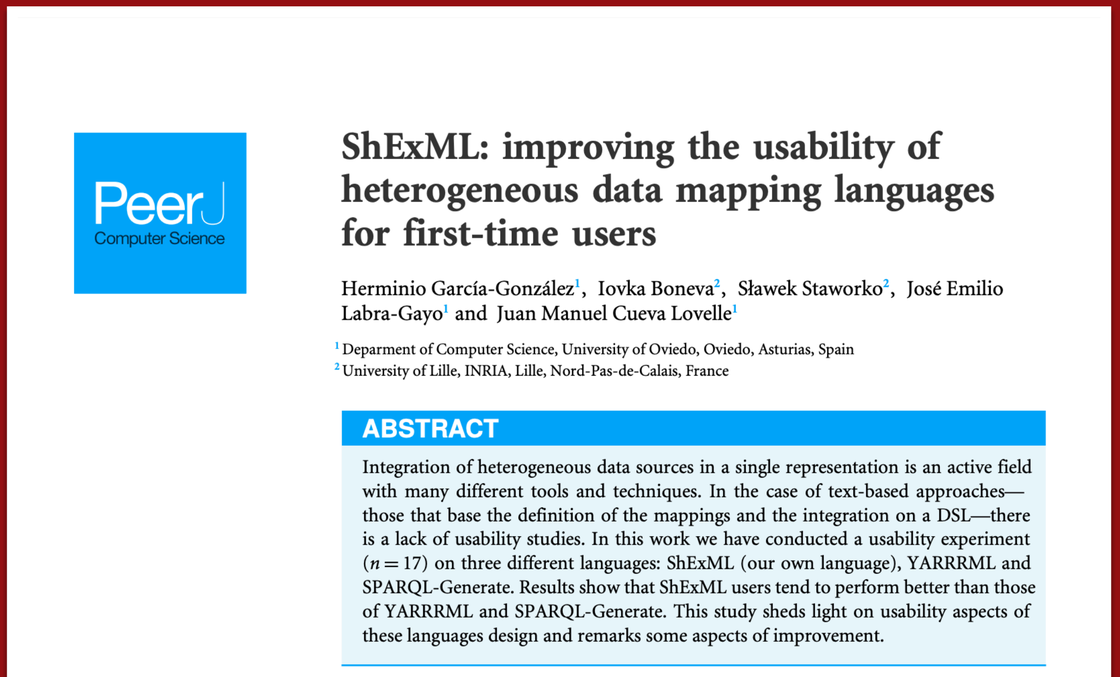
Introduction & Motivation

A thesis by published works

- Data integration language & usability study (@ PeerJ Computer Science)
- XML Schema to ShEx alignment (@ Semantic Web Journal)
- Using Semantic Web to enhance e-Learning content (@ IEEE Trans. Learning Technologies)
- Asturian Notaries integration methodology (@ WHISE III)

ShExML: improving the usability of heterogeneous data mapping languages for first-time users

Herminio García-González, Iovka Boneva, Sławek Staworko, José Emilio Labra-Gayo and Juan Manuel Cueva Lovelle



PeerJ Computer Science

ShExML: improving the usability of heterogeneous data mapping languages for first-time users

Herminio García-González[1], Iovka Boneva[2], Sławek Staworko[2], José Emilio Labra-Gayo[1] and Juan Manuel Cueva Lovelle[1]

[1] Deparment of Computer Science, University of Oviedo, Oviedo, Asturias, Spain
[2] University of Lille, INRIA, Lille, Nord-Pas-de-Calais, France

ABSTRACT

Integration of heterogeneous data sources in a single representation is an active field with many different tools and techniques. In the case of text-based approaches—those that base the definition of the mappings and the integration on a DSL—there is a lack of usability studies. In this work we have conducted a usability experiment ($n = 17$) on three different languages: ShExML (our own language), YARRRML and SPARQL-Generate. Results show that ShExML users tend to perform better than those of YARRRML and SPARQL-Generate. This study sheds light on usability aspects of these languages design and remarks some aspects of improvement.

ShExML usability study

Introduction

- Heterogeneous data integration in RDF
  - Text-based approaches (“expert users”)
  - Graphical approaches (“non-expert users”)
    - RML Editor, Karma…
- We are going to focus in text-based approaches
- Let’s make a recap

ShExML usability study

Related work

- First approaches (extending R2RML to heterogeneous data)
  - RML
  - xR2RML
- Inherited RDF based syntax
  - Verbose!!!
  - Seems like a serialisation / intermediate representation

ShExML usability study

RML minimum example

Result

```
@prefix :       <http://example.com/> .
@prefix schema: <http://schema.org/> .
@prefix dbr:    <http://dbpedia.org/resource/> .

:2      schema:name   "Interstellar" .

:1      schema:name   "Dunkirk" .

```

RML mapping rules

```
@prefix :       <http://example.com/> .
@prefix schema: <http://schema.org/> .
@prefix rml:    <http://semweb.mmlab.be/ns/rml#> .
@prefix rr:     <http://www.w3.org/ns/r2rml#> .
@prefix d2rq:   <http://www.wiwiss.fu-
berlin.de/suhl/bizer/D2RQ/0.1#> .
@prefix dbr:    <http://dbpedia.org/resource/> .
@prefix ql:     <http://semweb.mmlab.be/ns/ql#> .
@prefix map:    <http://mapping.example.com/> .

map:p_1  a            rr:predicateMap ;
         rr:constant  schema:name .

map:550785378  a                    rml:LogicalSource ;
         rml:iterator               "//film" ;
         rml:referenceFormulation   ql:XPath ;
         rml:source
 "http://shexml.herminiogarcia.com/files/films.xml" .

map:m_1  a                      rr:TriplesMap ;
         rml:logicalSource      map:550785378 ;
         rr:predicateObjectMap  map:po_1 ;
         rr:subjectMap          map:s_1 .

map:o_1  a            rr:ObjectMap ;
         rr:template  "{name}" ;
         rr:termType  rr:Literal .

map:po_1  a               rr:PredicateObjectMap ;
         rr:objectMap     map:o_1 ;
         rr:predicateMap  map:p_1 .

map:s_1  a            rr:SubjectMap ;
         rr:template  "http://example.com/{@id}" .

```

ShExML usability study

Related work

- Other languages try to solve this
  - SPARQL-Generate
    - SPARQL based syntax
    - Great functionality and flexibility
    - Difficult to learn
    - A lot of options

ShExML usability study

SPARQL-Generate minimum example

Result

```
@prefix :       <http://example.com/> .
@prefix schema: <http://schema.org/> .
@prefix dbr:    <http://dbpedia.org/resource/> .

:2      schema:name  "Interstellar" .

:1      schema:name  "Dunkirk" .

```

SPARQL-Generate mapping rules

```
PREFIX fun: <http://w3id.org/sparql-generate/fn/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mail: <http://example.com/mail#>
PREFIX note: <http://example.com/note/>
PREFIX ex: <http://example.com/>
PREFIX schema: <http://schema.org/>

GENERATE {
    <http://example.com/{?id}> schema:name ?name .
}
SOURCE <http://shexml.herminiogarcia.com/files/films.xml> AS ?source
ITERATOR ite:XPath(?source, "//film") AS ?film
WHERE {
    BIND( fun:XPath(?film,"/film/@id") AS ?id )
    BIND( fun:XPath(?film,"/film/name/text()") AS ?name )
}
```

ShExML usability study

Related work

- Other languages try to solve this
  - YARRRML
    - YAML based syntax (very compact syntax)
    - “Compiles” to RML
    - Totally dependent on RML

ShExML usability study

YARRRML minimum example

Result

```
@prefix :       <http://example.com/> .
@prefix schema: <http://schema.org/> .
@prefix dbr:    <http://dbpedia.org/resource/> .

:2      schema:name  "Interstellar" .

:1      schema:name  "Dunkirk" .

```

YARRRML mapping rules

Input: YARRRML

```
prefixes:
  ex: "http://example.com/"
  schema: "http://schema.org/"

mappings:
  person:
    sources:
      - ['films.xml~xpath', '//film']
    s: http://example.com/$(@id)
    po:
      - [schema:name, $(name)]
```

ShExML usability study

Why to create ShExML?

- Shape Expressions (ShEx) defines structure information of an RDF graph
  - ShEx based syntax to define output format
- Separation of concerns
  - Declarations (how to extract / transform / generate content)
  - Shapes (how to output content) (one shape to rule them all!)
- Increase usability with these precepts (main goal)
- KG generation & validation at the same time?

ShExML usability study

ShExML minimum example

Result

```
@prefix :       <http://example.com/> .
@prefix schema: <http://schema.org/> .
@prefix dbr:    <http://dbpedia.org/resource/> .

:2      schema:name  "Interstellar" .

:1      schema:name  "Dunkirk" .

```

ShExML mapping rules

```
PREFIX : <http://example.com/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX schema: <http://schema.org/>
SOURCE films_xml_file
<http://shexml.herminiogarcia.com/files/films.xml>
ITERATOR film_xml <xpath: //film> {
    FIELD id <@id>
    FIELD name <name>
}
EXPRESSION films <films_xml_file.film_xml>

:Films :[films.id] {
    schema:name [films.name] ;
}

```

ShExML usability study

The reason for a usability study

- YARRRML & SPARQL-Generate claim to be user-friendly
- No users studies were run on this topic
- Need to check these claims
- Need to check if ShExML was well designed
- Comparison between the three languages

ShExML usability study

Research questions

- RQ1: Is ShExML more usable for first-time users over other languages?
- RQ2: If true, can a relation be established between features support and usability for first-time users?
- RQ3: Which parts of ShExML—and of other languages—can be improved to increase usability?

ShExML usability study

Methodology - Sample

- 20 students of MSc in Web Engineering first course
- Most of them have a Bachelor (240 ECTS credits) in computer science or similar fields
- They were receiving a Semantic Web course of two weeks: total of 30 hours, 3 hours per day
- They were introduced to semantic technologies like: RDF, SPARQL, ShEx, etc.
- Experiment hosted the final day of the course
- Therefore, first-time users with some background knowledge

ShExML usability study

Methodology - Conduction

- Two tasks
  - Create a mapping from the scratch
  - Modify the created mapping
- Mousotron used to capture behavioural/objective measures (keystrokes, clicks, etc.)
- Office 365 used to capture achieved mapping rules, elapsed time and subjective impressions

ShExML usability study

Results

- First task (17 out of 20 submitted results)
- Second task (7 out of 20 submitted results)
- Statistical analysis of results
  - One Way ANOVA for normal distributions and continuous variables
  - Kruskal-Wallis for categorical data and non-normal distributions
    - Pair-wise comparison using Post-hoc tests

ShExML usability study

Results - Objective variables (Task 1)

- No significant differences between keystrokes, left and right button clicks, mouse wheel scroll and meters traveled by the mouse
- Elapsed seconds ($F(2,14) = 6.00$, $p = .013$, $\omega = .60$)
  - ShExML and YARRRML ($p = .016$)
- Completeness percentage ($H(2) = 9.73$, $p = .008$)
  - ShExML and SPARQL-Generate ($p = .012$, $r = .87$)
- Precision ($H(2) = 9.68$, $p = .008$)
  - ShExML and SPARQL-Generate ($p = .012$, $r = .87$)

ShExML usability study

Results - Subjective variables (Task 1)

- No significant differences between applicability, intuitiveness, error proneness, error reporting usefulness
- General satisfaction ($H(2) = 6.28$, $p = .043$)
  - ShExML and YARRRML ($p = .039$, $r = .69$)
- Learnability ($H(2) = 8.63$, $p = .013$)
  - ShExML and SPARQL-Generate ($p = .042$, $r = .78$)
  - ShExML and YARRRML ($p = .040$, $r = .69$)
- Mapping definition easiness ($H(2) = 10.25$, $p = .006$)
  - ShExML and SPARQL-Generate ($p = .013$, $r = .90$)
  - ShExML and YARRRML ($p = .037$, $r = .69$)
- Easiness of use ($H(2) = 9.82$, $p = .007$)
  - ShExML and YARRRML ($p = .011$, $r = .81$)

ShExML usability study

Discussion - Task 1

- Similar environment
- SPARQL-Generate users were not able to find working solutions
- Faster working solutions achieved with ShExML wrt YARRRML
- Applicability, intuitiveness, error proneness and error reporting usefulness (bad scores for all, we have to deal with this in the community)
- General usability and learnability better in ShExML
- No differences between YARRRML and SPARQL-Generate (can be caused by low completeness percentage in SPARQL-Generate)

ShExML usability study

Results & Discussion - Task 2

- No significant differences in objective and subjective analysis
  - Due to very low sample size
- 6 for ShExML and 1 for YARRRML (0 for SPARQL-Generate)
- YARRRML users were not able to achieve task 2 solution
- Modifiability: 3 in YARRRML, ShExML 83% = 5

ShExML usability study

Alignment with features support

- More features and their design affecting usability of SPARQL-Generate
- YARRRML & ShExML similar features
  - Difference could be the syntax
    - ShExML use of keywords (self-explanatory) and iterators modularity (object-oriented similarity).
- Features support is not the problem
  - Rather their design and inclusion in the language

ShExML usability study

Conclusions

- Better solutions and speed on finding them: ShExML
- No solutions achieved: SPARQL-Generate (first-time users find it difficult to use)
- Partial solutions: YARRRML
- This study covers first-time users with some programming/Linked Data background
  - Need for bigger experiments
  - Need for other profiles (non-expert users, semantic web practitioners, etc.)

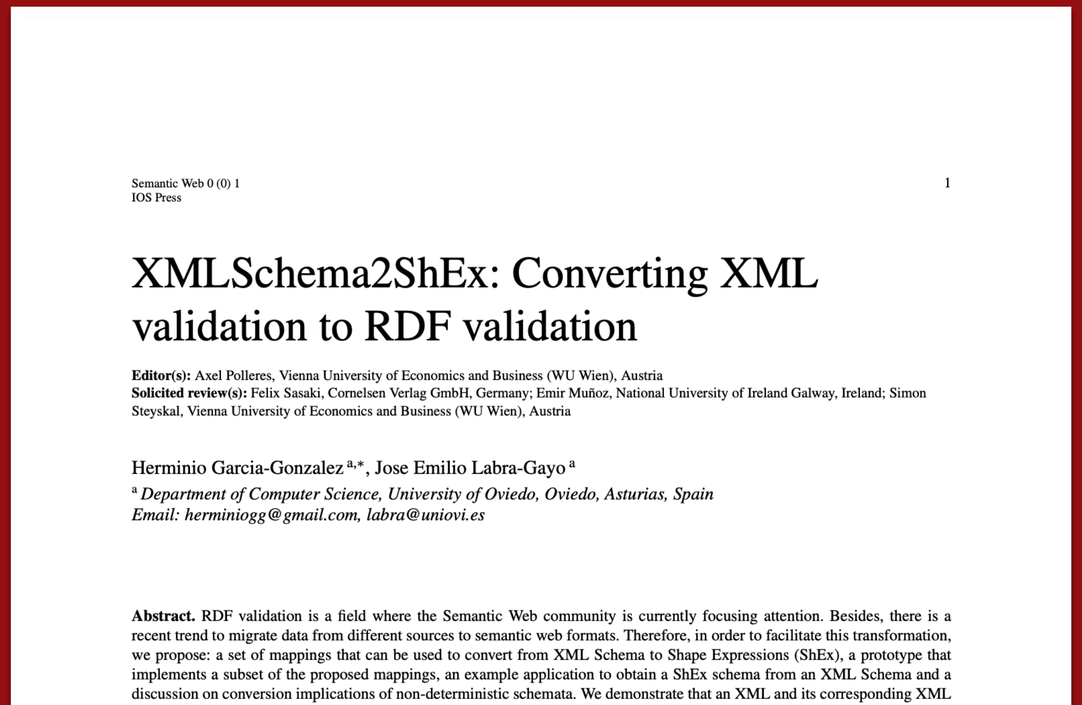
ShExML usability study

Conclusions

- Better usability -> Better accuracy of solutions and less time invested
- First-time users -> Adoption!!!
- Carefully design features support
- Future line -> Cognitive complexity Frameworks?

XMLSchema2ShEx: Converting XML validation to RDF validation

Herminio García-González and José Emilio Labra-Gayo



Semantic Web 0 (0) 1
IOS Press

XMLSchema2ShEx: Converting XML validation to RDF validation

Editor(s): Axel Polleres, Vienna University of Economics and Business (WU Wien), Austria
Solicited review(s): Felix Sasaki, Cornelsen Verlag GmbH, Germany; Emir Muñoz, National University of Ireland Galway, Ireland; Simon Steyskal, Vienna University of Economics and Business (WU Wien), Austria

Herminio Garcia-Gonzalez [a,*], Jose Emilio Labra-Gayo [a]
[a] *Department of Computer Science, University of Oviedo, Oviedo, Asturias, Spain*
Email: herminiogg@gmail.com, labra@uniovi.es

Abstract. RDF validation is a field where the Semantic Web community is currently focusing attention. Besides, there is a recent trend to migrate data from different sources to semantic web formats. Therefore, in order to facilitate this transformation, we propose: a set of mappings that can be used to convert from XML Schema to Shape Expressions (ShEx), a prototype that implements a subset of the proposed mappings, an example application to obtain a ShEx schema from an XML Schema and a discussion on conversion implications of non-deterministic schemata. We demonstrate that an XML and its corresponding XML

XMLSchema2ShEx

Introduction

- Data validation -> Normalisation, standardisation, data cleansing, reliability
- In XML world -> XML Schema (and other: Schematron, Relax NG)
- In Semantic Web
  - RDF Schema and OWL
    - Open World and Non-Unique Name Assumptions
  - Need for more specificity
    - ShEx and SHACL

XMLSchema2ShEx

Related work

- From XML Schema to other schemata (JSON Schema, relational)
- From XML Schema to RDF Schema and OWL
- FHIR approach (abstract domain model)
- No conversion to ShEx or SHACL
  - Validation of XML to RDF conversions
  - Transference of validation attributes to converted data

XMLSchema2ShEx

Research questions

- RQ1: What components should have a mapping from XML Schema to ShEx?
- RQ2: How to ensure that both schemata are equivalent?
- RQ3: Is it possible to ensure a backwards conversion in all cases?
- RQ4: Are non-deterministic schemata (i.e., ambiguous schemata) possible to translate and validate?

XMLSchema2ShEx

Proposed mappings - Easy ones

```
### XML Schema
<xs:element name="birthday" type="xs:date"/>

### ShEx
:birthday xs:date ;
```

Element mapping

```
### XML Schema
<xs:element name="item"
            minOccurs="0"
            maxOccurs="unbounded">
    <xs:complexType>
        ...
    </xs:complexType>
</xs:element>

### ShEx
:item @<item> * ;
```

Element with complex type mapping

XMLSchema2ShEx

Proposed mappings - More easy ones

```
### XML Schema
<xs:element name="nameZeroUnbounded"
            type="xs:string"
            minOccurs="0"
            maxOccurs="unbounded">
<xs:element name="nameOneUnbounded"
            type="xs:string"
            minOccurs="1"
            maxOccurs="unbounded">
<xs:element name="nameOptional"
            type="xs:string"
            minOccurs="0"
            maxOccurs="1">
<xs:element name="nameFourToTen"
            type="xs:string"
            minOccurs="4"
            maxOccurs="10">
```

Cardinality in XML Schema elements


```
:nameZeroUnbounded xs:string * ;
:nameOneUnbounded xs:string + ;
:nameOptional xs:string ? ;
:nameFourToTen xs:string {4, 10} ;
```

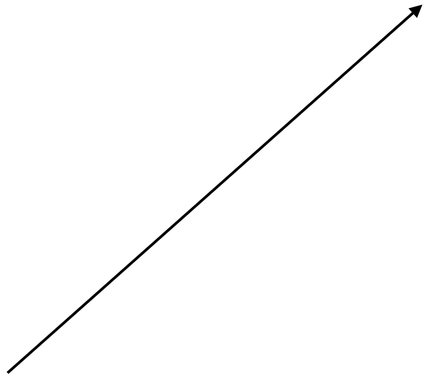
Conversion to ShEx (predicate + object)

XMLSchema2ShEx

Proposed mappings - First problems

- XML Schema allows to define an order inside complex types (xs:sequence)
- Difference in data model (tree vs graphs) complicates conversion
- Our proposed mapping: using RDF Lists

```
### XML Schema
<xs:complexType name="Address">
    <xs:sequence>
        <xs:element name="street"
            type="xs:string"/>
        <xs:element name="city"
            type="xs:string"/>
        <xs:element name="state"
            type="xs:string"/>
        <xs:element name="zip"
            type="xs:decimal"/>
    </xs:sequence>
</xs:complexType>
```



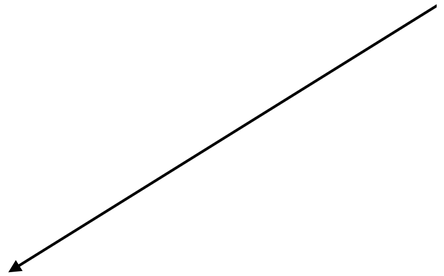
```
### ShEx
<address> {
    rdf:first @<street> ;
    rdf:rest @<i1> ;
}
<i1> {
    rdf:first @<city> ;
    rdf:rest @<i2> ;
}
<i2> {
    rdf:first @<state> ;
    rdf:rest @<i3> ;
}
<i3> {
    rdf:first @<zip> ;
    rdf:rest [ rdf:nil ] ;
}
<street> {
    :street xs:string ;
}
<city> {
    :city xs:string ;
}
<state> {
    :state xs:string ;
}
<zip> {
    :zip xs:decimal ;
}
```

XMLSchema2ShEx

Proposed mappings - Restrictions and extensions

- ShEx does not support restrictions, extensions or inheritance; though it is being discussed for inclusion
- Thus, resulting shape should be computed
- Loss of semantics! Not possible to perform a backwards conversion

```
### ShEx
<MountainBikeSize> ["small" "medium" "large"]

<FamilyMountainBikes> {
 :mountainBikeSize @<MountainBikeSize> ;
 :familyMember ["child" "male" "female"];
}
```



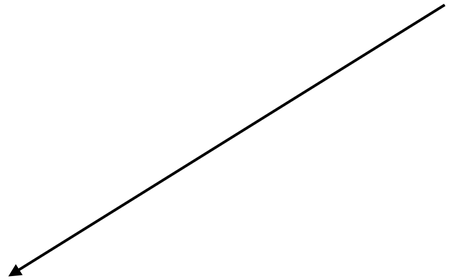
```
### XML Schema
<xs:simpleType name="mountainBikeSize">
  <xs:restriction base="xs:string">
    <xs:enumeration value="small" />
    <xs:enumeration value="medium" />
    <xs:enumeration value="large" />
  </xs:restriction>
</xs:simpleType>

<xs:complexType name="FamilyMountainBikes">
  <xs:simpleContent>
    <xs:extension base="mountainBikeSize">
      <xs:attribute name="familyMember">
       <xs:simpleType>
        <xs:restriction base="xs:string">
          <xs:enumeration value="child" />
          <xs:enumeration value="male" />
          <xs:enumeration value="female" />
        </xs:restriction>
       </xs:simpleType>
      </xs:attribute>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>
```

XMLSchema2ShEx

Proposed mappings - Unique

- ShEx does not support unique; though it is going to be included
- Possibility: use of semantic actions (ShEx extension mechanism)
- Backwards conversion will require an additional effort

```
### ShEx
%js{
    var ids = [];
    return true;
}
%
<Person> {
  :name xs:string ;
  :surname xs:string ;
  :id xs:integer
  %js{ if(ids.indexOf(_.o.lex) >= 0)
         return false;
       ids.push(_.o.lex);
       return true;
  }%
}
```



```
### XML Schema
<xs:element name="Person"
            maxOccurs="unbounded">
  <xs:complexType>
    <xs:all>
       <xs:element name="name"
                   type="xs:string" />
         <xs:element name="surname"
                     type="xs:string" />
           <xs:element name="id"
                       type="xs:integer" />
    </xs:all>
  </xs:complexType>
  <xs:unique name="onePersonPerID">
     <xs:selector xpath="."/>
     <xs:field xpath="id"/>
  </xs:unique>
</xs:element>
```

XMLSchema2ShEx

Results

- XML Schema elements conversion proposed (full list in the paper)
- Converted data can be validated against converted schema (see demonstration in the paper)
- Backwards conversion? Not always possible due to a loss of semantics
- What about non-deterministic schemata? Let’s see

XMLSchema2ShEx

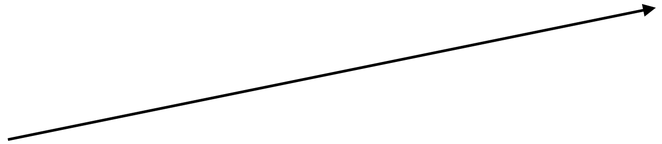
Non-deterministic schemata

- Ambiguity and Unique Particle Attribution
- Example: (ab | ac)
- On presence of a, the validator is not able to continue
- Why? To favor simple, streaming implementation

XMLSchema2ShEx

Non-deterministic schemata - ShEx conversion

```
### XML Schema
<xs:complexType name="nondeterministic">
    <xs:choice>
        <xs:sequence>
            <xs:element name="a"/>
            <xs:element name="b"/>
        </xs:sequence>
        <xs:sequence>
            <xs:element name="a"/>
            <xs:element name="c"/>
        </xs:sequence>
    </xs:choice>
</xs:complexType>
```



```
### ShEx
<nondeterministic> {
  a @<ab> OR @<ac> ;
}

<ab> {
  rdf:first @<a> ;
  rdf:rest @<ab1> ;
}

<ac> {
  rdf:first @<a> ;
  rdf:rest @<ac1> ;
}

<ab1> {
  rdf:first @<b> ;
  rdf:rest [rdf:nil]  ;
}

<ac1> {
  rdf:first @<c> ;
  rdf:rest [rdf:nil]  ;
}

<a> {
  :namea xs:string ;
}

<b> {
  :nameb xs:string ;
}

<c> {
  :namec xs:string ;
}
```

XMLSchema2ShEx

Non-deterministic schemata - ShEx result

- ShEx validates it correctly
- Why?
  - Structure of RDF Lists
  - Different validation form
  - Shape by shape (recursively)
- No ambiguity

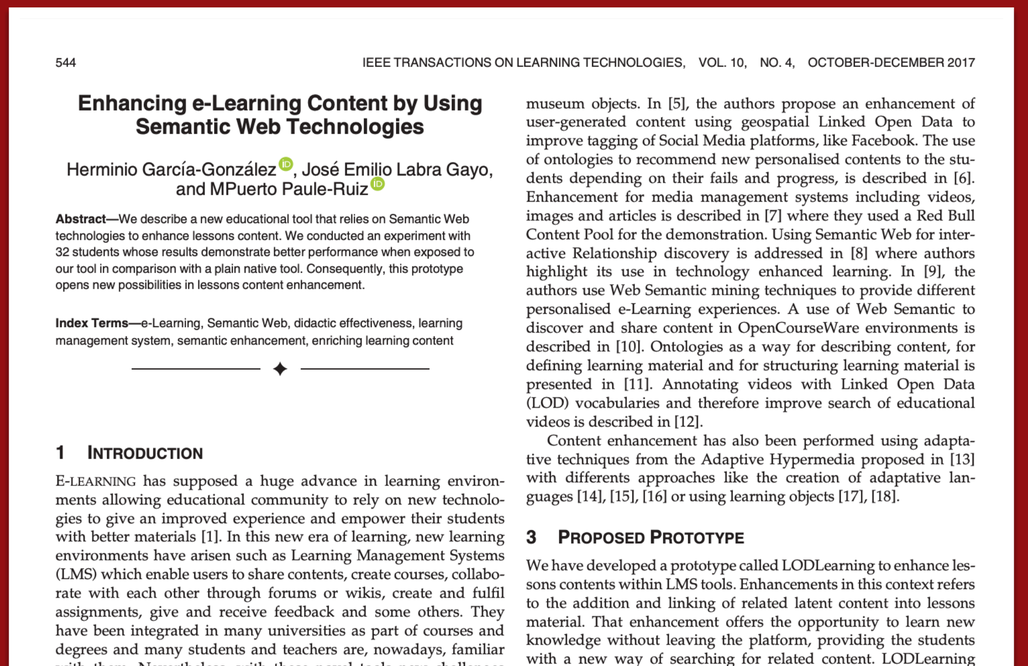
XMLSchema2ShEx

Conclusions & Future work

- Mapping proposed
- Working conversion
- Non-deterministic schemata solved
- Need to tackle loss of semantics
  - Difficult task due to difference in data models
- Other schema formats (Schematron and Relax NG)
- More data formats (e.g., JSON Schema)

Enhancing e-Learning content by using Semantic Web technologies

Herminio García-González, José Emilio Labra-Gayo and MPuerto Paule-Ruiz



Enhancing e-Learning Content by Using Semantic Web Technologies

Herminio García-González, José Emilio Labra Gayo, and MPuerto Paule-Ruiz

Abstract—We describe a new educational tool that relies on Semantic Web technologies to enhance lessons content. We conducted an experiment with 32 students whose results demonstrate better performance when exposed to our tool in comparison with a plain native tool. Consequently, this prototype opens new possibilities in lessons content enhancement.

Index Terms—e-Learning, Semantic Web, didactic effectiveness, learning management system, semantic enhancement, enriching learning content

1 Introduction

E-LEARNING has supposed a huge advance in learning environments allowing educational community to rely on new technologies to give an improved experience and empower their students with better materials [1]. In this new era of learning, new learning environments have arisen such as Learning Management Systems (LMS) which enable users to share contents, create courses, collaborate with each other through forums or wikis, create and fulfil assignments, give and receive feedback and some others. They have been integrated in many universities as part of courses and degrees and many students and teachers are, nowadays, familiar

museum objects. In [5], the authors propose an enhancement of user-generated content using geospatial Linked Open Data to improve tagging of Social Media platforms, like Facebook. The use of ontologies to recommend new personalised contents to the students depending on their fails and progress, is described in [6]. Enhancement for media management systems including videos, images and articles is described in [7] where they used a Red Bull Content Pool for the demonstration. Using Semantic Web for interactive Relationship discovery is addressed in [8] where authors highlight its use in technology enhanced learning. In [9], the authors use Web Semantic mining techniques to provide different personalised e-Learning experiences. A use of Web Semantic to discover and share content in OpenCourseWare environments is described in [10]. Ontologies as a way for describing content, for defining learning material and for structuring learning material is presented in [11]. Annotating videos with Linked Open Data (LOD) vocabularies and therefore improve search of educational videos is described in [12].

Content enhancement has also been performed using adaptative techniques from the Adaptive Hypermedia proposed in [13] with differents approaches like the creation of adaptative languages [14], [15], [16] or using learning objects [17], [18].

3 Proposed Prototype

We have developed a prototype called LODLearning to enhance lessons contents within LMS tools. Enhancements in this context refers to the addition and linking of related latent content into lessons material. That enhancement offers the opportunity to learn new knowledge without leaving the platform, providing the students with a new way of searching for related content. LODLearning

LODLearning

Introduction

- Semantic Web + e
- LMS content can
- Semantic Web as

After his spe
Obama excha
spent decade
killed when t

Barack Obama

Fecha de nacimiento:	1961-08-04
Fecha de la muerte:	Sin información
Lugar de nacimiento:	
Lugar de la muerte:	

Enlace a Wikipedia: Barack Obama

Barack Hussein Obama II (Acerca de este sonido [bəˈrɑːk huː ˈseɪn əˈbɑːmə] en inglés americano; Honolulu, 4 de agosto de 1961), conocido como Barack Obama, es el cuadragésimo cuarto y actual presidente de los Estados Unidos. Fue senador por el estado de Illinois desde el 3 de enero de 2005 hasta su renuncia el 16 de noviembre de 2008. Además, es el quinto legislador afroamericano en el Senado de los Estados Unidos, tercero desde la era de reconstrucción. También fue el primer candidato afroamericano nominado a la presidencia por el Partido Demócrata y es el primero en ejercer el cargo presidencial.Se graduó en la Universidad de Columbia y en la prestigiosa escuela de Derecho Harvard Law School, donde fue presidente de la revista Harvard Law Review. Posteriormente, trabajó como organizador comunitario y ejerció su carrera como abogado en derechos civiles, antes de ser elegido senador del estado de Illinois, desempeñando esa función desde 1997 a 2004. Fue profesor de Derecho constitucional en la facultad de Derecho de la Universidad de Chicago desde 1992 hasta 2004. En el año 2000 perdió la contienda electoral por un puesto en la Cámara de Representantes de los Estados Unidos, y tras su fracaso anterior, en enero de 2003 anunció su candidatura al Senado estadounidense. En marzo de 2004 venció en las elecciones primarias del partido demócrata, y en julio del mismo año

LODLearning

Related work

- Content enhancement in Adaptative Hypermedia
  - Learning objects
  - Adaptative languages
- Our approach
  - The teacher doesn’t need to have technical knowledge
  - We performed a numerical evaluation of didactic effectiveness

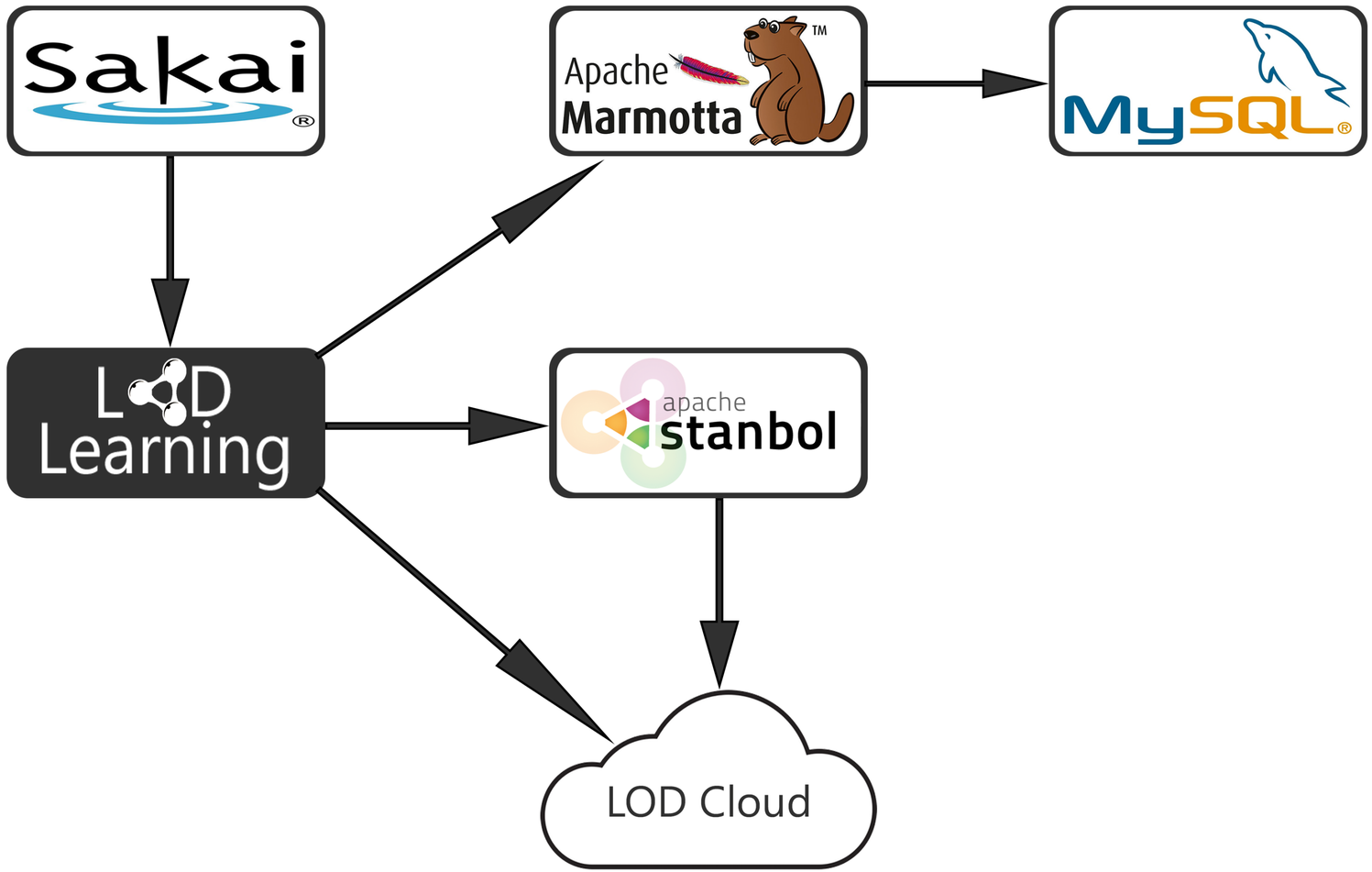
LODLearning

Research questions

- RQ1: Can the addition of Semantic Web content produce an improvement in didactic effectiveness?
- RQ2: What is the students' perception with these tools?
- RQ3: In which subjects would be more useful to include these tools?

LODLearning

The LODLearning prototype



LODLearning

Methodology - Sample

- 32 students from a State High School
  - Pursuing the mandatory education stage
  - 18 women and 14 men
  - From 13 to 14 ages old
- Divided in groups in a random manner
  - Control (Sakai LMS tool)
  - Experimental (LODLearning tool)

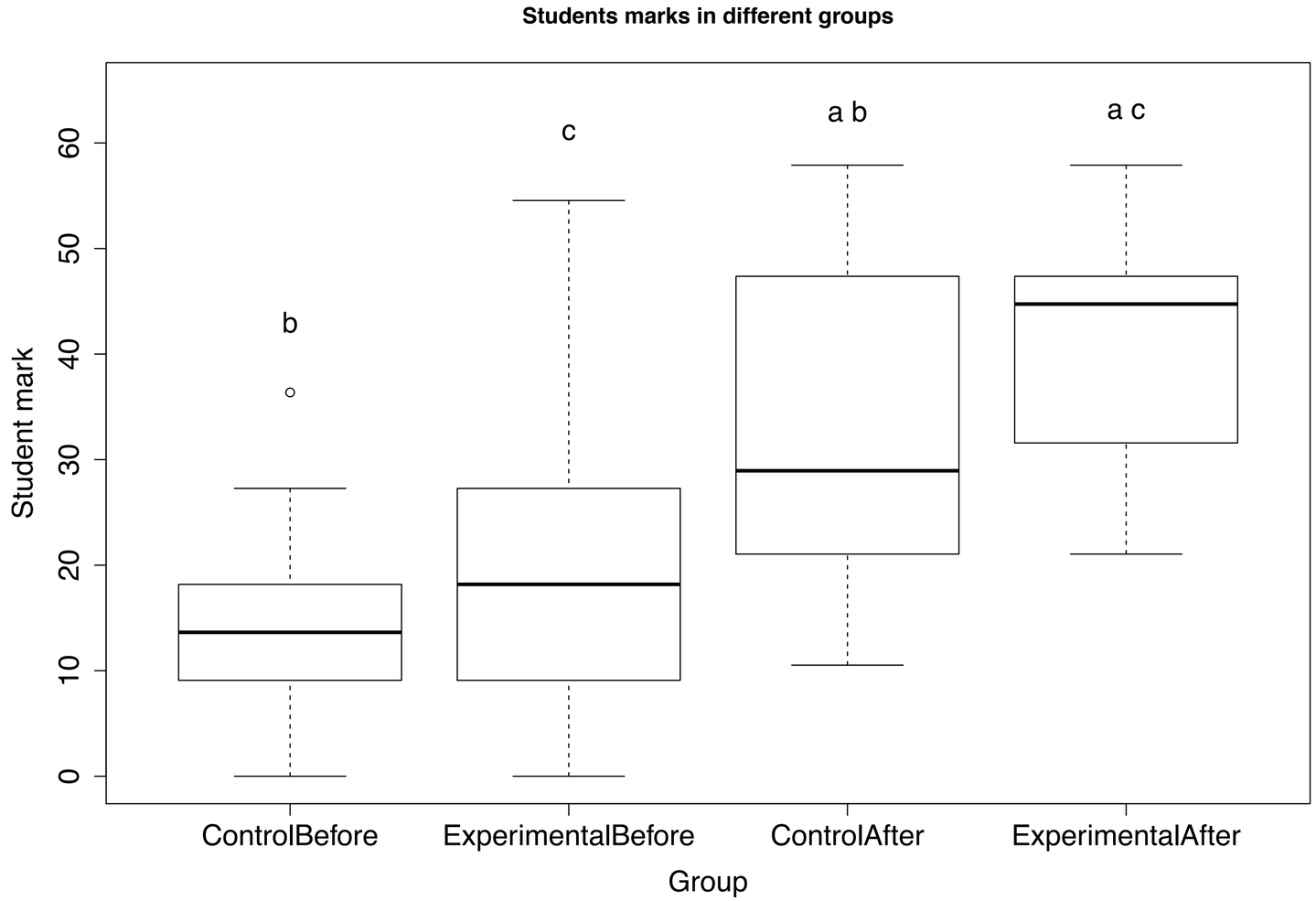
LODLearning

Methodology - Conduction

- Intra-subject study (between pretest and posttest)
- Inter-subject study (between control and experimental)
- Satisfaction questionnaire in the end

LODLearning

Results & Discussion



Student's t test

a
(p < .05)

b & c
(p < .001)

Cohen's d
.75

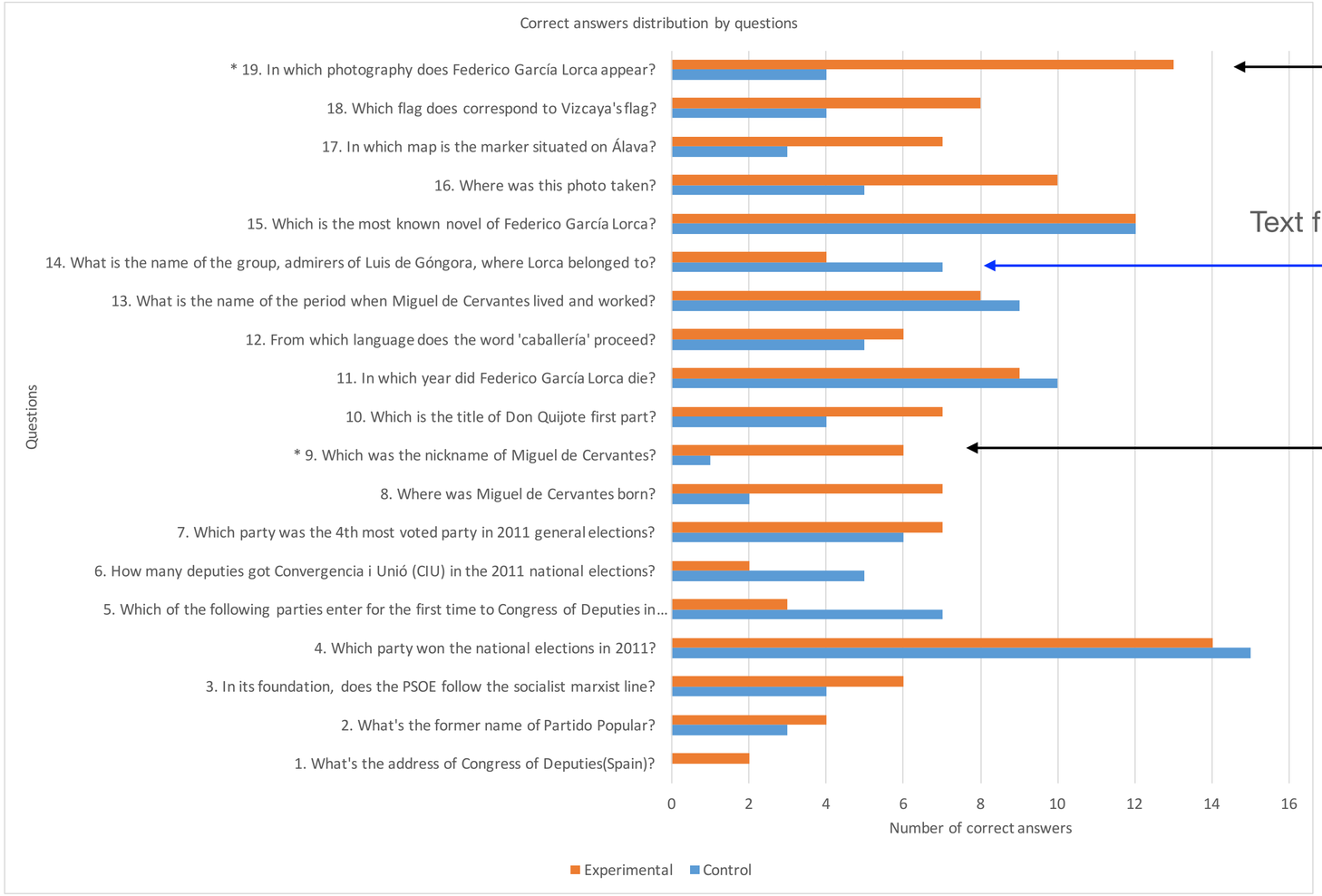
a
Inter-subject

b & c
Intra-subject

LODLearning

Results & Discussion - By questions

* ($p < .05$) Fisher's exact test



Use of images (as in 18,17 and 16)

Text from LMS, better for control group (as in 13, 11, 6 and 5)

Short text (as in 8, opposite to 2)

What to prioritise in cards?

- Short text
- Multimedia content

LODLearning

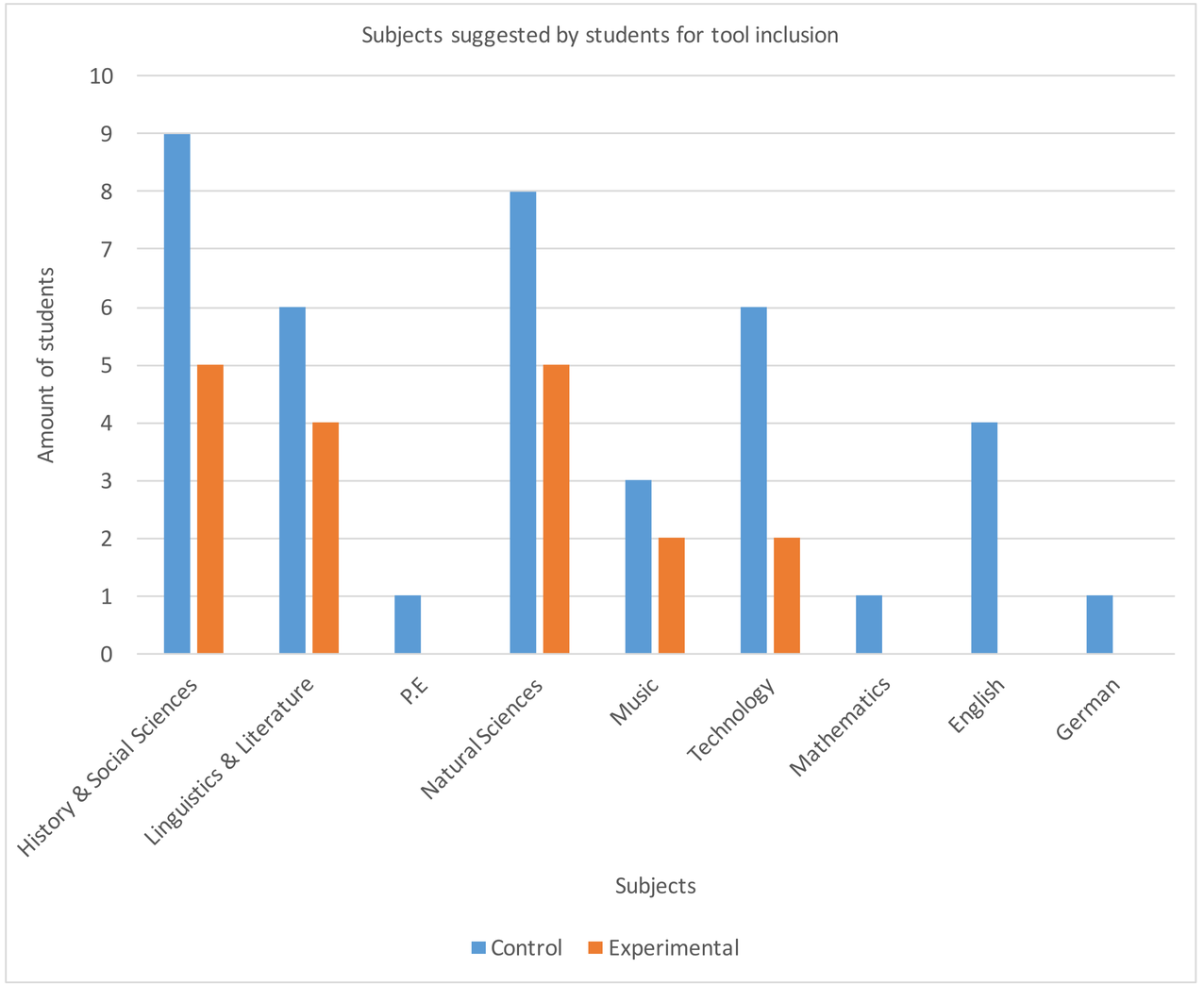
Results & Discussion

- No significant differences between groups in satisfaction questionnaire
  - They want both tools in their daily routine
  - No differences due to low familiarisation to virtual tools in a daily basis
- Adoption of LODLearning
  - Not affecting students’ workflow
  - No relevant issues using the prototype
  - Increasing students’ performance

LODLearning

Results & Discussion

- Most recommended subjects in line with experiment contents
- Also Natural Sciences and Technology
- Why not Fine Arts?



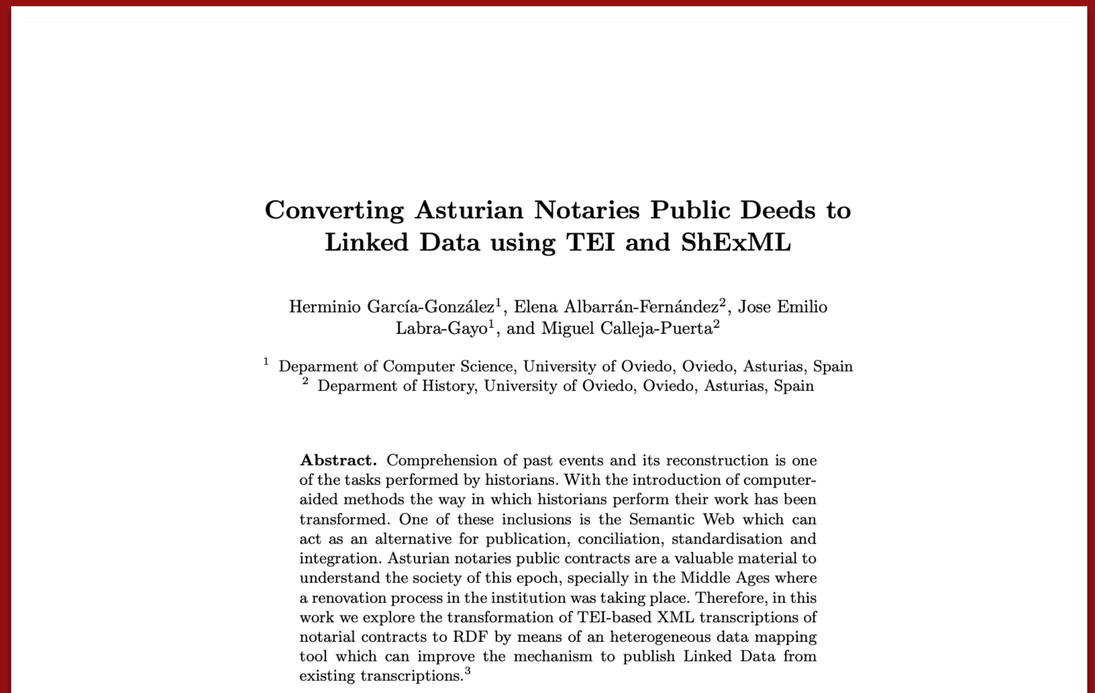
LODLearning

Conclusions & Future work

- LMS + Semantic Web -> It can improve courses didactic effectiveness
- More tool development
  - More cards types
- More experiments
  - More exhaustive and longer (long time) ones
- Pilot programme

Converting Asturian Notaries Public deeds to Linked Data using TEI and ShExML

Herminio García-González, Elena Albarrán-Fernández, José Emilio Labra-Gayo and Miguel Calleja-Puerta



Converting Asturian Notaries Public Deeds to Linked Data using TEI and ShExML

Herminio García-González[1], Elena Albarrán-Fernández[2], Jose Emilio Labra-Gayo[1], and Miguel Calleja-Puerta[2]

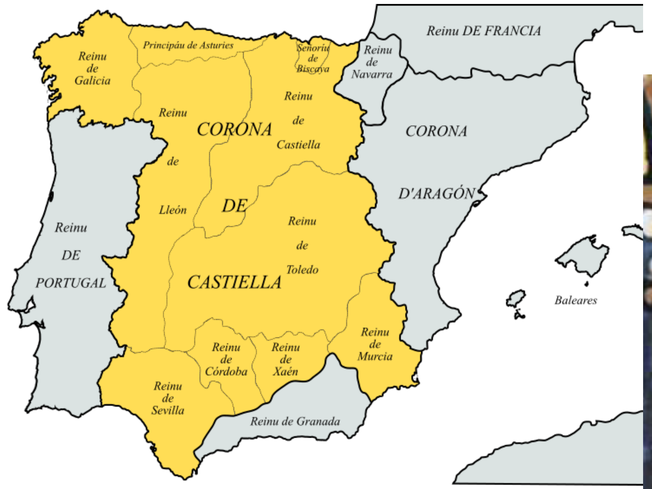
[1] Deparment of Computer Science, University of Oviedo, Oviedo, Asturias, Spain
[2] Deparment of History, University of Oviedo, Oviedo, Asturias, Spain

Abstract. Comprehension of past events and its reconstruction is one of the tasks performed by historians. With the introduction of computer-aided methods the way in which historians perform their work has been transformed. One of these inclusions is the Semantic Web which can act as an alternative for publication, conciliation, standardisation and integration. Asturian notaries public contracts are a valuable material to understand the society of this epoch, specially in the Middle Ages where a renovation process in the institution was taking place. Therefore, in this work we explore the transformation of TEI-based XML transcriptions of notarial contracts to RDF by means of an heterogeneous data mapping tool which can improve the mechanism to publish Linked Data from existing transcriptions.[3]

Converting Asturian Notaries deeds

Introduction - Historical context

- Asturias in the XIII century
  - Part of the Castilian crown
- King Alfonso X (1252-1284)
  - New policy for notaries
    - Written culture no church exclusivity
    - Services to everyone
    - Defined by the law



(Photos extracted from Wikimedia Commons)

Converting Asturian Notaries deeds

Introduction - Back to the present (some particularities)

- No notarial registers remain
- Lot of manuscripts guarded by ecclesiastical institutions
- Some of them in public-state archives

Converting Asturian Notaries deeds

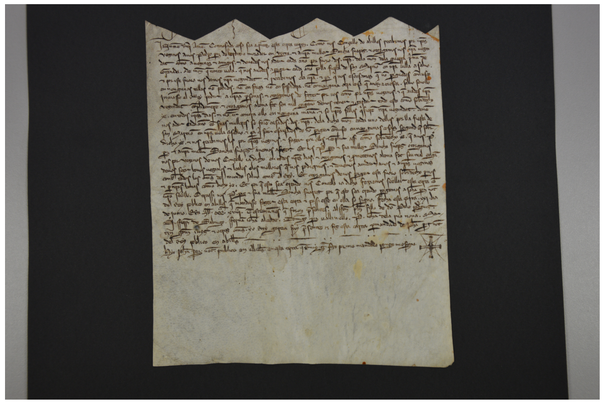
Introduction - Problem of this particularity

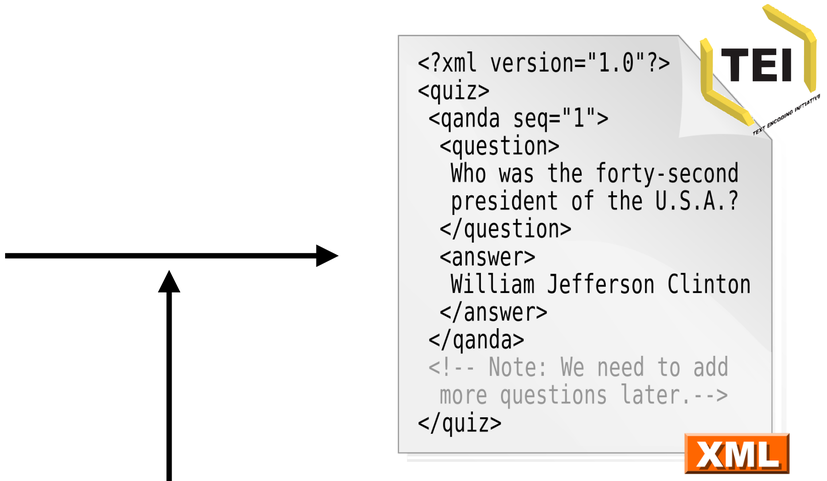
- Ecclesiastical institutions
  - No digitised versions
  - A transcriber must go there
  - Limited opening hours
- Public-state archives
  - Digitised versions in some cases
  - Low quality in many cases
  - OCR to be proven

A transcriber is a must for this material!

Converting Asturian Notaries deeds

Methodology





Data mapping tool

(XML icon extracted from Wikimedia Commons,
TEI logo from TEI website and
RDF logo from W3C)

Converting Asturian Notaries deeds

Methodology - Our TEI model

- Using TEI as the basis
- Need for diplomatic features
- Most similar extension CEI, but not complete
- Addition of diplomatic features
- More on [1]

[1] Albarrán-Fernández, E.: A TEI-Based model to encode notarial charters (Asturias, 1260-1350 ca .) (Sep 2019). https://doi.org/10.5281/zenodo.3447525

Converting Asturian Notaries deeds

Methodology - Translation to Linked Data

- Using ShExML
- Using schema.org vocabulary (schema:ArchiveComponent, schema:ArchiveOrganization, schema:Collection, schema:Place, etc.)
- Nesting subtypes (linked shapes)
- Linking with other LOD cloud entities

Converting Asturian Notaries deeds

Methodology - Translation to Linked Data

```
:Manuscript :[manuscriptId] {
    a schema:ArchiveComponent ;
    schema:holdingArchive @:Archive ;
    schema:editor @:Person ;
    schema:licence [manuscripts.licence] ;
    schema:material @:Material ;
    schema:abstract [manuscripts.abstract] @es ;
    schema:dataCreated [manuscripts.origDate] xs:date ;
    schema:locationCreated @:LocationCreated ;
    schema:producer @:Office ;
    schema:text [manuscripts.text] @ast ;
    schema:citation @:Biblio ;
}
```

Nesting

```
:Archive :[manuscripts.institution] {
    a schema:ArchiveOrganization ;
    schema:name [manuscripts.institution] ;
    schema:location @:ArchiveLocation ;
}
```

Nesting

```
:LocationCreated :[manuscripts.oficinaWhere] {
    a schema:Place ;
    schema:name [manuscripts.oficinaWhere] @es ;
    schema:sameAs wd:[manuscripts.oficinaWhere MATCHING city] ;
}
```

Linking

Converting Asturian Notaries deeds

Problems - The case of a diplomatic study

```
<oficina type="notaria" subtype="regia" where="Aviles">
    <persName xml:id="JuanPerez" role="notario">Juan Pérez, notario público
    del rey Avilés</persName>
    <persName xml:id="RuyGarcia" role="amanuense">Ruy García fuy presente e
    fiz esta carta</persName>
</oficina>
```

```
<seg function="suscriptio" type="notaria"><seg type="amanuense">Yo, Ruy
García, fuy presente e fiz esta carta per mandado de Iohan Périz, notario
del rey póblico en Abillés.</seg> <seg type="notario">Hyo Iohan Périz,
notario público en Abillés, en esta carta que Ruy García fizo por mio
mandado pongo mio signo</seg></seg>.
```

There is a specific relationship which is present in the text but not well reflected here!!!

```
:Office :[officeId] {
    a schema:Organization ;
    schema:name [manuscripts.oficina] ;
    schema:location @:LocationCreated ;
    :person @:Notario ;
    :person @:Excusador ;
    :person @:Amanuense ;
    :person @:Registrador ;
}

:Notario :[notarioId] {
    a :[manuscripts.notarioType] ;
    :text [manuscripts.notario] ;
}

:Amanuense :[amanuenseId] {
    a :[manuscripts.amanuenseType] ;
    :text [manuscripts.amanuense] ;
}
```

Not a worker but a special hierarchy!!!

Not a text but a diplomatic clause!!!

Converting Asturian Notaries deeds

Related ontologies - Charter Encoding Initiative

- Closest one
- Covers some of the notaries clauses (dispositio, datatio, invocatio, etc)
- Not an ontology (need for transitioning)
- Does not cover the previous issue
- Specific ontology required!!!

Converting Asturian Notaries deeds

Other limitations and challenges

```
<oficina type="notaria" subtype="regia" where="Aviles">
    <persName xml:id="JuanPerez" role="notario">Juan Pérez, notario público
    del rey Avilés</persName>
    <persName xml:id="RuyGarcia" role="amanuense">Ruy García fuy presente e
    fiz esta carta</persName>
</oficina>
```

Who is Juan Pérez?
Who is Ruy García?

These names and surnames can appear for other people!!!

Possibility: Include more variables

- Place (mobility?)
- Co-workers (change of role?)

Need for a personalised entity disambiguation mechanism

Converting Asturian Notaries deeds

Conclusions

- Notaries' manuscripts in Asturias
- Need for a transcriber
- Transcriptions in TEI
- Heterogeneous data mapping languages
- Limitations on existing vocabularies
- Further techniques (e.g., entity disambiguation)

Discussion, challenges and future work

Discussion

On data transformation

- Improved usability with ShExML
- More experiments and profiles
  - GUI for non-experts users? Really beneficial?
  - GUI for experts users? GUI vs text-based approaches?
  - Therefore, invest more time on graphical or text-based approaches?
  - Migration from graphical (after training) to text-based approaches?

Discussion

On data validation

- Possibility to also transform schemas
- Relation between data and schema transformation
  - Automatic data transformation
    - Schema alignments
  - Aided transformation
    - GUI to guide a semi-automatic transformation

Discussion

On use cases

- Heterogeneous data integration in Knowledge Graphs
  - Improving didactic effectiveness
  - A mean for historical material publication, conciliation, standardisation and integration
  - More use cases
    - FAIR principles
    - Reproducibility crisis

Conclusions

Conclusions

Benefits & thoughts

- Lower transformation costs
  - Easier migration to semantic technologies
- Knowledge integration, disambiguation, dissemination
  - Redounds in society
- Interdisciplinary research
  - To improve our tools and methods
  - To improve their tools and methods
  - Joint development!

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Semantic integration of heterogeneous big data sources

Herminio García González

Supervised by Prof. José Emilio Labra Gayo & Prof. Juan Manuel Cueva Lovelle