Universität Bielefeld

The need for Lexicalization of Linked Data

John McCrae

Cognitive Interaction Technology Excellence Center – Universität Bielefeld

CITEC

Universität Bielefeld

Linked Data

- Linked data is growing rapidly...
- ... but mostly it looks like this:

```
24016     <Calculate_lengths rdf:about="#Calculate_side_length_of_right_triangle_from_one_side_and_one_angle">
24017         <creator>1</creator>
24018         <created>2009-21-17 04:15:34</created>
24019         <modified>2009-25-17 17:06:19</modified>
24020         <commonName xml:lang="de">die Seitenlänge in einem rechtwinkligen Dreieck über eine Seite und einen Winkel bere
24021         <commonName xml:lang="fr">calculer un côté d'un triangle rectangle connaissant un côté et un angle</commonName>
24022         <type rdf:resource="#Calculate_lengths" />
24023         <hasTopic rdf:resource="#Formula_for_length" />
24024         <hasTopic rdf:resource="#Right-Triangle" />
24025         <hasTopic rdf:resource="#Trigonometric_ratio" />
24026     </Calculate_lengths>
24027     <Calculate_results_of_operations rdf:about="#Calculate_simple_calculation">
24028         <creator>1</creator>
24029         <modified>2009-26-22 09:11:46</modified>
24030         <created>2009-21-22 04:15:37</created>
24031         <commonName xml:lang="de">einfache Berechnungen durchführen</commonName>
24032         <commonName xml:lang="es">hacer cálculos simples</commonName>
24033         <commonName xml:lang="fr">effectuer des calculs simples</commonName>
24034         <type rdf:resource="#Calculate_results_of_operations" />
24035         <hasTopic rdf:resource="#Simplecalculationoflength" />
24036     </Calculate_results_of_operations>
```

CITEC

Universität Bielefeld

Linked Data

- We need:
  - Natural Language Generation/Interface
    - Description in text
  - Question Answering
    - Mapping natural language description to (SPARQL) queries
  - Machine Translation
    - Adapting linked data vocabularies to new languages

CITEC

New to Holland *Dutch government immigration website*

Home | Checklist | Subjects

Welcome

If you come to the Netherlands to live, work or study, you are likely to have some questions about the arrangements you need to make. This site will guide you to the government organizations you may have to deal with.

What can you do?

If you fill in your profile, you will see a list of subjects that are relevant for you, with links to information from the organizations concerned. You can also choose a subject directly.

Your profile

What is your age?
Make your choice

What is your nationality?
Make your choice

Length of your stay?
Make your choice

The purpose of your stay?
(You may give more than one answer)

- [ ] work in Holland
- [ ] study in Holland
- [ ] live in Holland

Create my checklist

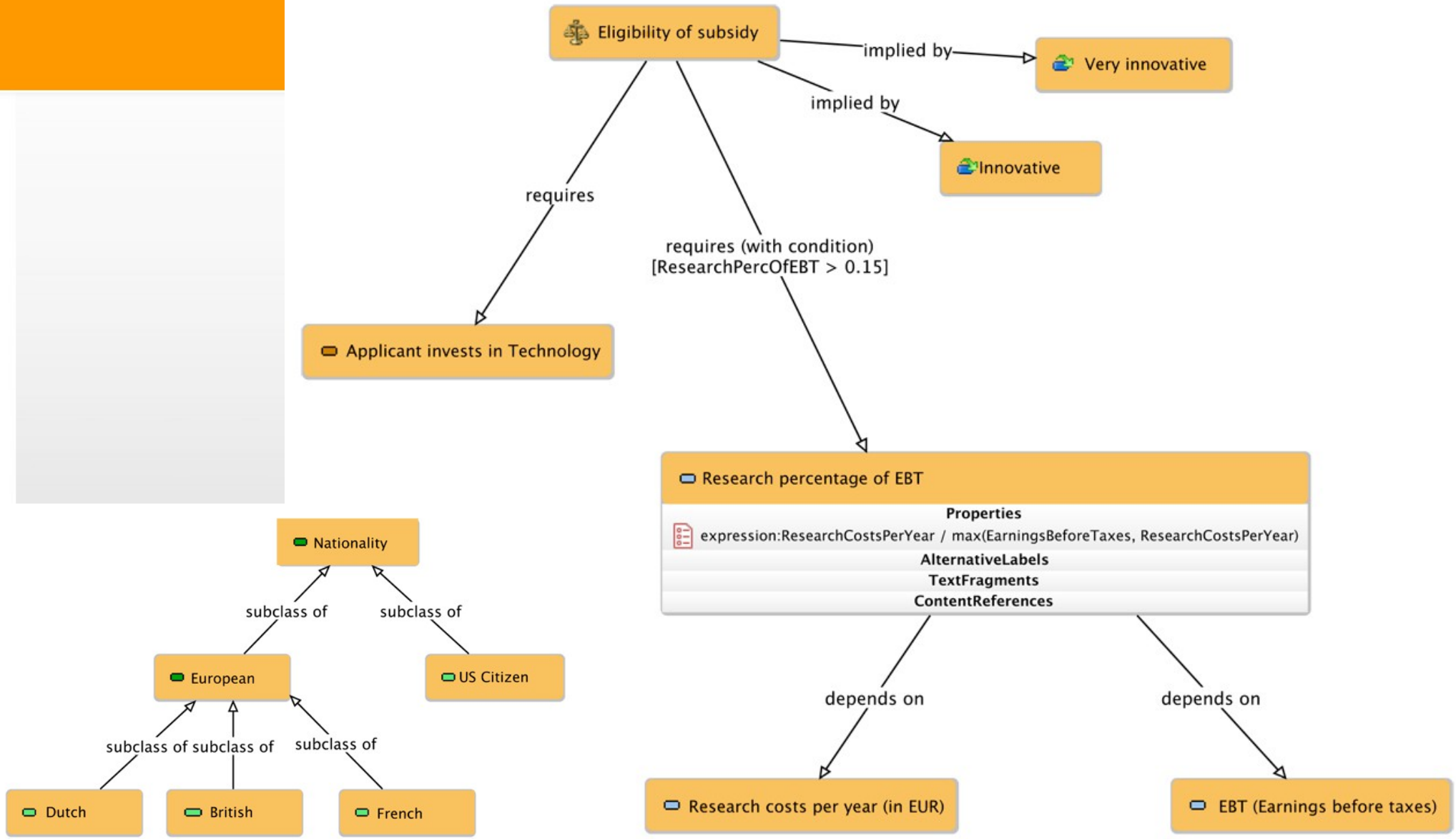
Subjects

- Education
- Employment
- Permits and visa
- Social security
- Taxes
- Vehicles

See also

- Employers (dutch)
- The forum (dutch)

be informed


Eligibility of subsidy
implied by
Very innovative
implied by
Innovative
requires
requires (with condition)
[ResearchPercOfEBT > 0.15]
Applicant invests in Technology
Research percentage of EBT
Properties
expression:ResearchCostsPerYear / max(EarningsBeforeTaxes, ResearchCostsPerYear)
AlternativeLabels
TextFragments
ContentReferences
depends on
depends on
Research costs per year (in EUR)
EBT (Earnings before taxes)
Nationality
subclass of
subclass of
European
US Citizen
subclass of subclass of
subclass of
Dutch
British
French

Universität Bielefeld

INFORMATION

PYTHIA

ONTOLOGY-BASED QUESTION ANSWERING

geobase

how many rivers run through texas?

Ask Pythia

I found 1 answer.

ANSWER: 44

Show analysis

Are you satisfied with this answer? Yes No

I found 1 answer.

ANSWER: 44

Show analysis

Are you satisfied with this answer? Yes No

WHAT?

PYTHIA is an ontology-based question answering system. This demo is made available for research purposes - feel free to play around with it.

HOW?

PYTHIA translates natural language into SPARQL by means of a deep linguistic analysis. Learn more »

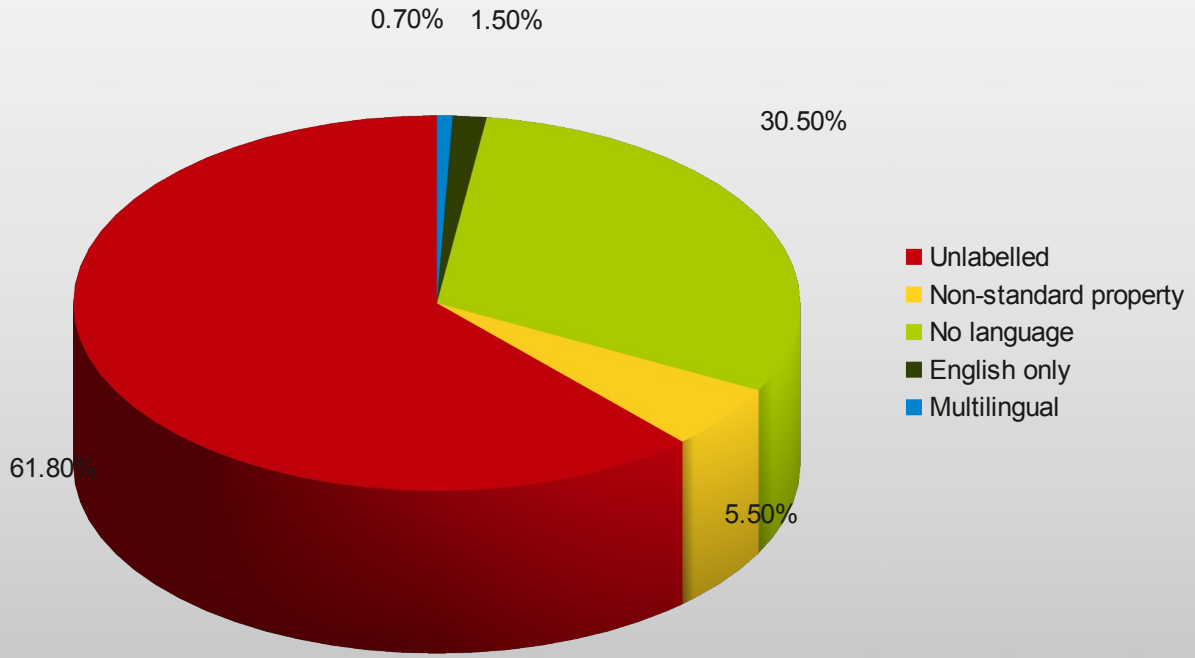
WHO?

PYTHIA is developed by the Semantic Computing Group @ CITEC, Bielefeld University.

CITEC

Labels

- Linguistic description of linked data terms by `rdfs:label`
- Usage statistics:



Source: Ell, B., Vrandecic, D. & Simperl, E. Labels in the Web of Data. In Proc. of ISWC-2011.

Universität Bielefeld

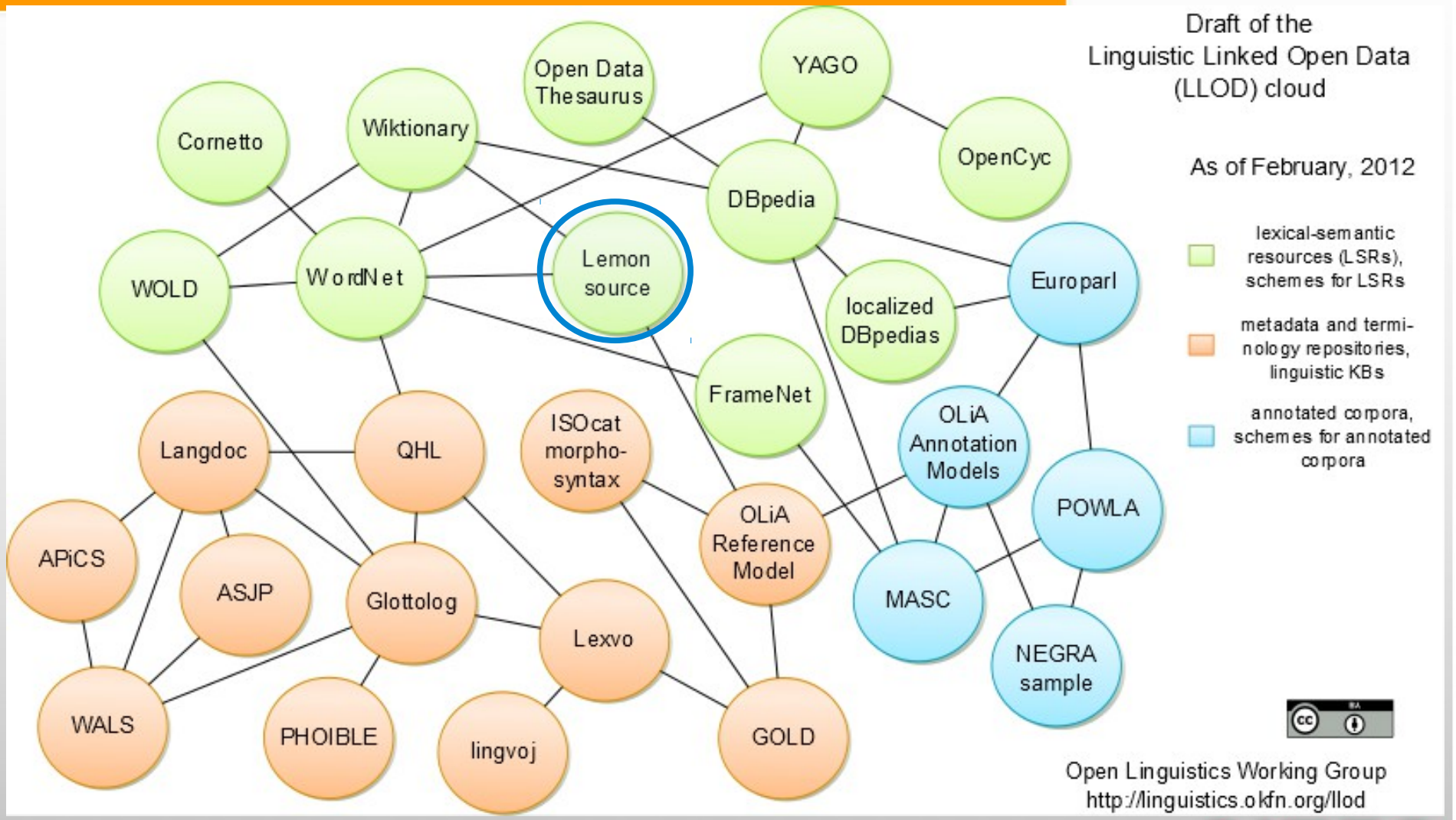
Labels are not enough!

- Simple labels are very ambiguous, e.g.,
  - "addresses" (from openEHR Demographic)
    - The "addresses" of an organization?
    - Someone "addresses" an audience?
    - A set of web "addresses"??
- Use URIs for labels not/as well as strings!

CITEC

Universität Bielefeld

Linguistic Linked Data



CITEC

Universität Bielefeld

Lexicon model for ontologies

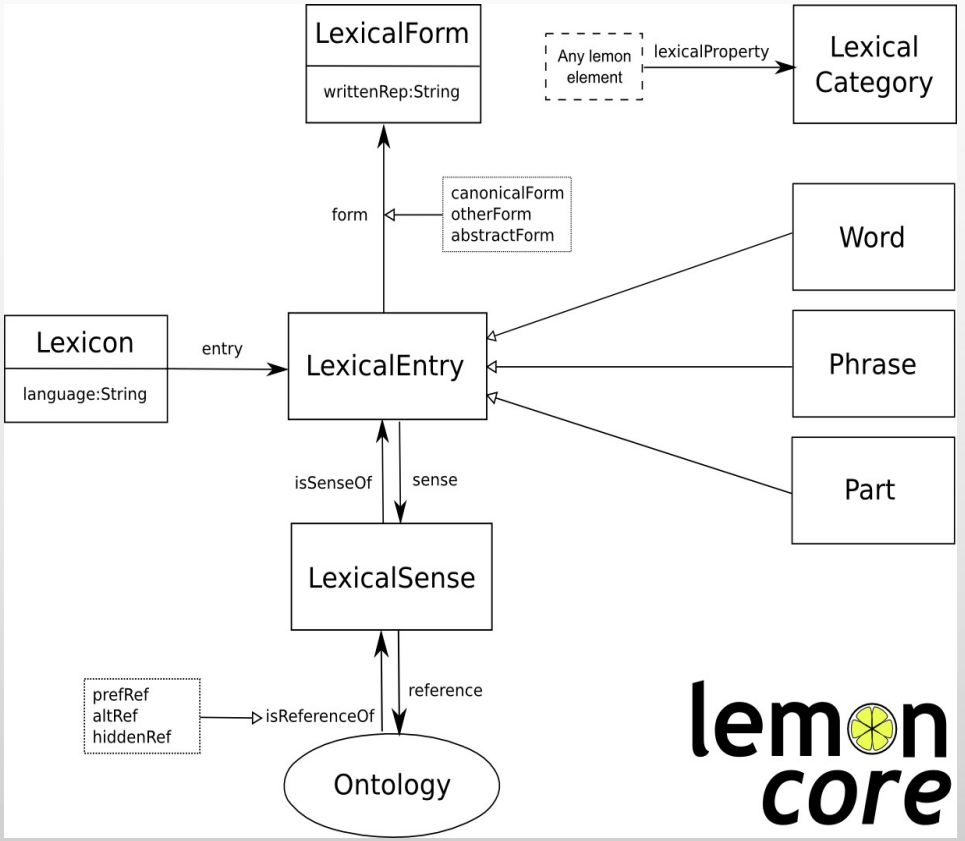
- Common format for describing lexical information relative to 'ontologies' (OWL, RDF(S))
- Built on existing models
  - Lexical Markup Framework (ISO 24613)
  - SKOS
- Design:
  - Modular
  - Concise
  - RDF-native
  - Not prescriptive

CITEC

Universität Bielefeld

Lexicon model for ontologies

- Allows full linguistic description
- Further development under W3C OntoLex community group
- Described in cookbook



CITEC

Universität Bielefeld

Using Lemon

- People will not create a *lemon* model for each vocabularies
- Instead refer to repositories on *lemon* data
  - Such as *lemon source*
- Before *lemon*
  - `openehr:addresses rdfs:label "Addresses"@en`
- With *lemon*
  - `openehr:addresses lemon:lexicalization lemonsource:address__noun__sense1__en`
- Full linguistic description available by dereferencing URI

CITEC

Universität Bielefeld

Thank you!

- Ontolex Community group
  - http://www.w3.org/community/ontolex
- Lemon cookbook
  - http://lexinfo.net/lemon-cookbook.pdf
- Monnet project
  - http://www.monnet-project.eu/

CITEC