A Linkset Quality Metric Measuring Multilingual Gain In SKOS Thesauri

Riccardo Albertoni, Monica De Martino, Paola Podestà

name.surname@ge.imati.cnr.it

IMATI - Istituto di Matematica Applicata e Tecnologie Informatiche "Enrico Magenes" – GENOVA Section, Consiglio Nazionale delle Ricerche, Italy
So many bubbles there, THAT’S SO COOL!!
BUT ....
Can I exploit that third party data for my OWN ANALYSES?

“QUALITY IS THE ISSUE WHEN REUSE”

Well-founded works on quality for DATASETS, ... but ...
MOTIVATION

“QUALITY IS THE ISSUE WHEN REUSE”

Well-founded works on quality for DATASETS, ...

but ...

NO GROUND CONCEPTs
About what makes a LINKSET suitable for A TARGET APPLICATION
"QUALITY IS THE ISSUE WHEN REUSE"

Well-founded works on quality for DATASETS, ...

but ...

NO GROUND CONCEPTs

About what makes a LINKSET suitable for A TARGET APPLICATION

Considering LOD’s promise: to make the web Evolving into a Global Data Space

LINKSET QUALITY should be AS IMPORTANT AS DATASET QUALITY

Riccardo Albertoni
“QUALITY IS THE ISSUE WHEN REUSE”
Well-founded works on quality for DATASETS, ...
but ...
What arrows in LOD are good for ?!?!

NO GROUND CONCEPTs
About what makes a LINKSET suitable for a TARGET APPLICATION
Considering LOD’s promise: to make the web Evolving into a Global Data Space
LINKSET QUALITY should be AS IMPORTANT AS DATASET QUALITY

Riccardo Albertoni
WHAT IS a LINKSET? (see VOID)
WHAT IS a LINKSET? (see VOID)

Subject X

Animal@en
Snake@en
Serpente@it
Dog@en
Puppy@en
Guard Dog@en

x1
x2
x3
x4
x5

y1
y2
y3
y4
y5
y6

skos:broader
skos:exactMatch
skos:prefLabel
skos:altLabel

Cane@it
Cagnolino@it
Cane da guardia@it
Trained Animal@en
Puppy@en
Dog@en

imati
WHAT IS a LINKSET? (see VOID)

**Subject X**
- Snake@en
- Serpente@it
- Dog@en
- Puppy@en
- Guard Dog@en

**Object Y**
- Animal@en
- Cane da guardia@it
- Cagnolino@it
- Dog@en
- Puppy@en

Relations:
- skos:broader
- skos:exactMatch
- skos:prefLabel
- skos:altLabel
WHAT IS a LINKSET? (see VOID)
WHAT IS a LINKSET? (see VOID)

Riccardo Albertoni, Asunción Gómez-Pérez: Assessing linkset quality for complementing third-party datasets. EDBT/ICDT Workshops 2013: 52-59

owl:sameAs linksets
WHAT IS a LINKSET? (see VOID)

Riccardo Albertoni, Asunción Gómez-Pérez: Assessing linkset quality for complementing third-party datasets. EDBT/ICDT Workshops 2013: 52-59

Inspired by LusTRE, a framework of interlinked Environmental Thesauri in the EU project
WHAT IS a LINKSET? (see VOID)

Riccardo Albertoni, Asunción Gómez-Pérez: Assessing linkset quality for complementing third-party datasets. EDBT/ICDT Workshops 2013: 52-59

owl:sameAs linksets

skos:exactMatch linksets

THIS WORK!!!

Inspired by LusTRE, a framework of interlinked Environmental Thesauri in the EU project
PROPOSAL: LINKSET IMPORTING

A quality scoring function on Linkset to answer

How good is the Linkset to import object dataset’s INFO into the subject SKOS Thesaurus?

The Overall Linkset Importing is defined as the average contribution of importing on single links of the Linkset

Importing on Links
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p, given a link, evaluates the percentage of values not present in the subject, but “gainable” from the Object, through the link
Accurate formal definition in the paper, here I am using a simplified version

\[ LkImp4p(l, p) = 100 \times \left( 1 - \frac{|Val4PropertyP_{in\_l's\_Subject}|}{|Val4PropertyP_{in\_l's\_Subject} \land Val4PropertyP_{in\_l's\_Object}|} \right) \]
Accurate formal definition in the paper, here I am using a simplified version

LkImp4p assumes the link is correct...

$LkImp4p(I, P) = 100 \times (1 - \frac{|\text{Val4PropertyP in I's Subject}|}{|\text{Val4PropertyP in I's Subject} \cap \text{Val4PropertyP in I's Object}|})$
Accurate formal definition in the paper, here I am using a simplified version

\[ L\text{imp}4p(l, P) = 100\times(1 - \frac{|\text{Val}_4\text{Property}_{P\text{ in } l\text{'s Subject}}|}{|\text{Val}_4\text{Property}_{P\text{ in } l\text{'s Subject} \cap \text{Val}_4\text{Property}_{P\text{ in } l\text{'s Object}}|}) \]
Accurate formal definition in the paper, here I am using a simplified version

$LkImp4p(l, P) = 100 * \left(1 - \frac{|Val4PropertyP_in_l's_Subject|}{|Val4PropertyP_in_l's_Subject \bigcap Val4PropertyP_in_l's_Object|} \right)$
Accurate formal definition in the paper, here I am using a simplified version

$LkImp4p(I, P) = 100 \times (1 - \frac{|Val4PropertyP_in_I's_Subject|}{|Val4PropertyP_in_I's_Subject| + |Val4PropertyP_in_I's_Object|})
Accurate formal definition in the paper, here I am using a simplified version

$LkImp4p(l, P) = 100 \times (1 - \frac{|Val4Prop_{P\_in\_l\_s\_Subject}|}{|Val4Prop_{P\_in\_l\_s\_Subject} \cup Val4Prop_{P\_in\_l\_s\_Object}|})$

$LkImp4p(...)$ = 100% iff there are values in the object and no values in the subject
Accurate formal definition in the paper, here I am using a simplified version

\[ L_{kimp4p}(I, P) = 100 \times \left(1 - \frac{|\text{Val4PropertyP_in_I's_Subject}|}{|\text{Val4PropertyP_in_I's_Subject} \land \neg \text{Val4PropertyP_in_I's_Object}|} \right) \]

\[ L_{kimp4p}(\ldots) = 100\% \text{ iff there are values in the object and no values in the subject} \]

if denominator = 0 then we define \( L_{kimp4p}(\ldots) = 0 \)
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p_L(x3, skos:prefLabel, l2, 'en')
LINK IMPORTING: Examples

LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p_{L}(x3, skos:prefLabel, l2, 'en')
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p_L(x3, skos:prefLabel, l2, 'en')
LINK IMPORTING: Examples

LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p_L(x3, skos:prefLabel, l2, 'en')
LINK IMPORTING: Examples

LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4pₜ(x3, skos:prefLabel, l2, 'en')
No Multilingual gain as “Dog@en” is already in the subject
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4pL(x3, skos:altLabel, l2, 'it')
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4pL(x3, skos:altLabel, l2, 'it')
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4pL(x3, skos:altLabel, l2, 'it')
LINK IMPORTING: Examples

LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4pL(x3, skos:altLabel, l2, 'it')

No altLabel in Italian
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p(x3, skos:altLabel, l2, ‘it’)
LkImp4p(ObjEntity, Property, Link, Language) --> [0 %, 100%]

LkImp4p_L(x3, skos:altLabel, l2, 'it') =100%
As Cagnolino@it is gained and it is the only altLabel in the complemented Subject
LkImp4p(ObjEntity, Property, L, Language) --> [0 %, 100%]
LINK IMPORTING: Examples

LkImp4p(ObjEntity, Property, L, Language) --> [0 %, 100%]

Disregarding the language, “_ _” means Unspecified Language

LkImp4p_L(x3, skos:altLabel, l2, _ )
LkImp4p(ObjEntity, Property, L, Language) --> [0 %, 100%]

Disregarding the language, "_" means Unspecified Language

LkImp4p_L(x3, skos:altLabel, l2, _)
LkImp4p(ObjEntity, Property, L, Language) --> [0 %, 100%]

Disregarding the language, "_" means Unspecified Language

LkImp4pL(x3, skos:altLabel, l2, _ )
LINK IMPORTING: Examples

LkImp4p(ObjEntity,Property,L,Language) --> [0 %, 100%]

Disregarding the language, "_ _" means Unspecified Language

LkImp4p_L(x3,skos:altLabel, l2, _)
LkImp4p(ObjEntity, Property, L, Language) \rightarrow [0 \%, 100\%]

Disregarding the language, "_" means Unspecified Language

LkImp4p_L(x3, skos:altLabel, l2, _ )
LkImp4p(ObjEntity,Property,L,Language) --> [0 %, 100%]

Disregarding the language, “_” means Unspecified Language

LkImp4p_L(x3, skos:altLabel, l2, _) = 50%

x3 has another altLabel (Puppy@en)
Then we gain one out two of the altlabels.
LINKSET IMPORTING: Examples
It can be applied to any properties, e.g., skos:broader, in this case, we gain entities instead of RDF literals.
It can be applied to any properties, e.g., skos:broader.
In this case, we gain entities instead of RDF literals

$LkImp4p_L(x5, \text{skos:broader}, l3, \_)$
It can be applied to any properties, e.g., skos:broader. In this case, we gain entities instead of RDF literals.

$LkImp4p_L(x5, \text{skos:broader}, l3, \_)$
It can be applied to any properties, e.g., `skos:broader`. In this case, we gain entities instead of RDF literals.

$Lk\text{Imp}p_L(x5, \text{skos:broader}, l3, _)$. 

**LINKSET IMPORTING: Examples**
It can be applied to any properties, e.g., skos:broader, In this case, we gain entities instead of RDF literals

$LkImp4p_L(x5, \text{skos:broader}, I3, \_)$
It can be applied to any properties, e.g., skos:broader, in this case, we gain entities instead of RDF literals.

\[ \text{LkImp4p}_L(x_5, \text{skos:broader, l3, _}) = 50\% \]

\( x_3 \) and \( y_3 \) are mapped so only \( y_6 \) is considered as an entity gained.
It can be applied to any properties, e.g., skos:broader, In this case, we gain entities instead of RDF literals

$LkImp_{P_L}(x5, \text{skos:broader}, l3, _) = 50\%$

$x3$ and $y3$ are mapped so only $y6$ is considered as a entity gained
It can be applied to any properties, e.g., `skos:broader`, in this case, we gain entities instead of RDF literals.

\[ \text{Lklmp4p}_L(x_5, \text{skos:broader}, l_3, _) = 50\% \]

\[ x_3 \text{ and } y_3 \text{ are mapped so only } y_6 \text{ is considered as a entity gained} \]
LINKSET IMPORTING can be exploited to check the complementation potential of any SKOS property.
LINKSET IMPORTING can be exploited to check the complementation potential of any SKOS property.

But in the application, we focus on skos:prefLabel and skos:altLabel in order to address INCOMPLETE LANGUAGE COVERAGE.
LINKSET IMPORTING can be exploited to check the complementation potential of any SKOS property. But in the application, we focus on skos:prefLabel and skos:altLabel in order to address INCOMPLETE LANGUAGE COVERAGE.

INCOMPLETE LANGUAGE COVERAGE arises when skos:prefLabel and skos:altLabel are provided in all the expected languages only for a subset of the thesaurus concepts.
EXAMPLE of APPLICATION (eENVPlus)
EXAMPLE of APPLICATION (eENVPlus)

EARTH LINKSETS to
• GEMET (E2GEM - 4365 links)
• AGROVOC (E2AGR - 1436 links)
EXAMPLE of APPLICATION (eENVPlus)

EARTH LINKSETS to
• GEMET (E2GEM - 4365 links)
• AGROVOC (E2AGR - 1436 links)

Is E2GEM better than E2AGR?!!
In term of # links sure!! (4365 links >> 1436 links)
EXAMPLE of APPLICATION (eENVPlus)

EARTH LINKSETS to
- GEMET (E2GEM - 4365 links)
- AGROVOC (E2AGR - 1436 links)

Is E2GEM better than E2AGR ??
In term of # links sure!! (4365 links >> 1436 links)

Multilingual Gain for skos:prefLabel

Average Linkset Importing

- E2AGR
- E2GEM
EXAMPLE of APPLICATION (eENVPlus)

EARTH LINKSETS to
- GEMET (E2GEM - 4365 links)
- AGROVOC (E2AGR - 1436 links)

Is E2GEM better than E2AGR??
In term of # links sure!! (4365 links >> 1436 links)
EXAMPLE of APPLICATION (eENVPlus)

EARTH LINKSETS to
• GEMET (E2GEM - 4365 links)
• AGROVOC (E2AGR - 1436 links)

Is E2GEM better than E2AGR ??
In term of # links sure!! (4365 links >> 1436 links)

Multilingual Gain for skos:prefLabel

Multilingual Gain for skos:altLabel

The best depends on which set of languages and properties we are focusing on...

But at least our Importing quality measure enables in a deeper analysis than #link or link coverage
CONCLUSION

Two-fold contribution

1. We draw the community attention to the critical issue of linkset quality
2. We propose LINKSET IMPORTING to measure the gain when complementing Thesauri
   - Applicable for estimating multilingual gain
   - Example in the context of EU project eENVplus

Future work

1. Importing as estimator for the “completeness of complemented thesauri” (experimental validation)
2. Further scoring functions to fully characterize linkset quality space and dimensions
THANK YOU

Questions ??

ALBERTONI@GE.IMATI.CNR.IT