Undefined 1 (2009) 1–5
IOS Press

Assessing completeness when complementing SKOS thesauri: two quality measures on skos:exactMatch linksets

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Riccardo Albertoni [a,*], Monica De Martino [a] and Paola Podestà [a]
[a] *Istituto di Matematica Applicata e Tecnologie Informatiche,*
Consiglio Nazionale delle Ricerche,
Via De Marini, 6, 16149 Genova, Italy
E-mail: {albertoni,demartino,podesta}@ge.imati.cnr.it

Abstract. Quality is one of the big challenges when consuming Linked Data. Measures for quality of linked data datasets have been proposed, mainly by adapting concepts defined in the research field of information systems. However, very limited attention has been dedicated to the quality of linksets, that might be as important as dataset's quality when consuming data coming from distinct sources. In this paper, we address linkset quality proposing two measures, the *reachability* and the *importing*, to assess the completeness of linkset-complemented SKOS thesauri. In particular, the reachability and importing estimate the ability of a linkset to enrich a thesaurus, respectively, with new concepts and their properties. We validate the proposed measures with an in-house developed synthetic benchmark and we show an example of their exploitation on real linksets in the context of the EU project eENVplus.

Keywords: Quality measures, linkset quality, SKOS, dataset complementation, completeness

1. Introduction

Linked Data is largely adopted by data producers such as European Environment Agency, US and some EU Governs, whose first ambition is to share (meta)data making their processes more effective and transparent. The increasing interest and involvement of data providers surely represents a genuine witness of the Linked Data success, but in a longer perspective, the quality of the exposed data will be one of the most critical issues in the data consumption process. After all, as discussed in [23], data is only worth its quality.

The research pertaining to Linked Data quality is especially focused on datasets [23]. However, one of the most interesting promises that Linked Data makes is "Linked Data will evolve the current web data into a Global Data Space", which implicitly assumes the exploitation of data items coming from different sources as a whole. In the Linked Data context, this is possible by connecting information belonging to different sources by way of linksets. Through linksets a Linked Data consumers can reach new information to complete and enrich data at hand, so, in order to keep the Linked Data promise, the quality of connections (hereinafter linkset quality) is as important as the quality of data.

This paper proposes a method to shed light on this. It presents two measures, the *reachability* and the *importing*, which estimate the ability of a linkset to enrich a thesaurus with new concepts and their properties re-

*Corresponding author. E-mail: albertoni@ge.imati.cnr.it

spectively. The proposed metrics do not aim at assessing the completeness and the correctness of a linkset or of the thesauri the linkset is connecting. They rather estimate the completeness of the linkset-complemented thesaurus, which is the thesaurus enriched with information that is obtained by cross-walking one of its linkset. This paper extends the notion of linkset quality introduced in [2] focusing on `skos:exactMatch` linksets, a kind of linkset among thesauri exposed as Simple Knowledge Organization System (SKOS) Ontology in the Linked Data. `skos:exactMatch` linksets have been chosen considering the application scenarios we are facing in the EU funded project eENVplus (CIP-ICT-PSP grant No. 325232), where we deal with a remarkable number of environmental thesauri exposed as Linked Data [4]. Considerable efforts have been spent to interlink thesauri such as GEMET, EARTh [3], AGROVOC [7], EUROVOC, UNESCO, RAMEAU, TheSoz [22], but, currently, there is no way to assess the *value* of these interlinks in terms of usefulness and information gain.

To this purpose this paper specifically contributes with

- the formalization and experimental validation of *reachability*, a metric which values the number of new concepts reachable passing through a linkset. In particular, *reachability* can be deployed to estimate the ability of a linkset to enrich the set of concepts that are browsable through a thesaurus (aka., the thesaurus' browsing space);
- the experimental validation of *importing*, a measure we formalized in [5], which checks the linkset complementation potential for any SKOS property. In particular, when applied on the properties `skos:prefLabel` and `skos:altLabel`, *importing* can help in addressing the incomplete language coverage[1] issue, which affects many popular SKOS thesauri [21];
- an application of these two measures in the context of eENVplus (CIP-ICT-PSP grant No. 325232), in which the potential of the aforementioned measures is shown considering two examples of linksets in the Linking Open Data cloud diagram 2014.

The paper is organized as follows. Section 2 introduces concepts on which the paper relies on (i.e., dataset, linkset and complementation of a dataset via its linkset). Section 3 formalizes the *importing* and *reachability* providing related quality indicators and score functions. Section 4 introduces the goal of our experimentation and explains the methodological and architectural setting adopted to validate the proposed quality measures. Section 5 discusses the results of experimentation showing the proposed measures as an effective predictor for multilingual and concept gain which may be obtained by complementing thesauri via their `skos:exactMatch` linksets. Section 6 applies the *importing* and *reachability* on some examples of linksets in the LOD cloud. Finally, we discuss related work in Section 7 and the conclusions and future work in Section 8.

2. Basic Concepts

This paper considers resources on the Web represented by the Resource Description Framework (RDF)[2]. A RDF triple *(s,p,o)*, whose elements s, p and o are generically indicated as RDF terms (hereafter, the set **RDFTerms**), is a tuple $(s, p, o) \in$ (RDFRiri $\cup$ BNode) $\times$ RDFProp $\times$ (RDFRiri $\cup$ BNode $\cup$ RDFLit) where **RDFRiri** is the set of RDF resources denoted by an IRI (e.g., *http://dbpedia.org/resource/Earth*), **BNode** is the set of blank nodes, and **RDFLit** is the set of RDF literals (e.g., Dog). **RDFProp** represents the set of RDF properties, which might be divided in Object Property (**OBJProp**), namely properties connecting resources/blacknodes, and Data Properties (**DTProp**),which are properties connecting resource to RDFLit. Each literal in RDFLit has a datatype $\mathcal{D}$, belonging to the set of possible datatypes $\mathfrak{D}$. $\mathfrak{D}$ includes xsd datatypes and in particular, strings with ISO language tags[3] (**RDFLitLtag**) (e.g., Dog@en).

In particular, we use the notion of dataset and linkset provided in the Vocabulary of Interlinked Datasets (VoID) [8], an RDF vocabulary commonly adopted for expressing metadata about RDF datasets exposed as Linked Data. A **dataset (D)** is a set of RDF triples published, maintained or aggregated by a single provider. Formally, let X be a dataset, the predicate $t_X(s, p, o)$ holds if and only if the RDF triple *(s,p,o)* $\in$ X.

A **linkset (L)** is a special kind of dataset containing only RDF links between the *void:subjectsTarget* and the *void:objectsTarget* respectively representing

[1] 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.

[2] http://www.w3.org/TR/rdf11-primer/

[3] http://tools.ietf.org/html/bcp47#section-2.2.9

the **object** and the **subject** of the linkset. Each RDF link is RDF triple *(s,p,o)*, where s and o belong respectively to the *subject* and *object* dataset of the linkset, *p* is a object property (e.g., *skos:exactMatch*, *owl:sameAs*) that indicates the type of the link. RDF links in a linkset should all have the same type, otherwise, the linkset should be split in distinct linksets. This paper considers ***skos:exactMatch* linksets**, that are made only by *skos:exactMatch* links among SKOS thesauri. Such type of linksets binds SKOS concepts with equivalent meaning.

The notions of multi-relational network, multidimensional adjacency matrix, power matrix presented in [18] are exploited and extended to mathematically captures the connectivity among RDF resources.

Definition 1 ***(Multi-relational network from a RDF triple set).*** *A multi-relational network is defined for each set of RDF triples X as* $M_X = (V_X, \mathbb{E}_X)$ *, where*

- $V_X = \{s|t_X(s,p,o)\} \cup \{o|t_X(s,p,o) \wedge o \in$ *RDFRiri*$\cup$*BNode*$\}$ *is the set of vertices in the network, which represent RDF resources and blank nodes in X;*
- $\mathbb{E}_X = \{E_1, E_2, ..., E_m\}$ *is the family of edge sets in the network, such that, there is an edge set* $E_q \subseteq (V \times V) : 1 \geq q \geq m$ *for every object property in X, namely,* $\forall p.t_X(s,p,o), \exists q.q =$ *indexOf*(p).

Assuming $n = |V_X|$, $m = |\mathbb{E}_X|$, $x, y \in V_X$, $z \in \mathbb{E}_X$, *indexOf* : $V_X \cup RDFProp \rightarrow [1, max(n,m)]$ *is a surjective function mapping vertices and object properties in their correspondent indexes.*

Given a set of RDF triple X, let $W \in \{0,1\}^{1\times m}$ be an array of weights for object property, such that, $W[q]$ is equal to 1 if the object property q occurring in the set of RDF triple X is considered relevant for a given analysis.

Definition 2 ***(Multi-dimensional weighted adjacency matrix for a RDF triple set).*** *The multi-relational weighted adjacency matrix is the result of the function* $\mathcal{A} : (2^{RDFTriple} \times \{0,1\}^{1\times m}) \rightarrow \{0,1\}^{m\times n\times n}$ *defined as*

$$\mathcal{A}(X,W)_{q,i,j} = \begin{cases} W[q] & \text{if } t_X(x,z,y) \\ 0 & \text{otherwise.} \end{cases}$$

assuming $x, y \in V_X$, $i =$ *indexOf*(x), $j =$ *indexOf*(y) *and* $z \in E_q$, q = *indexOf*(z).

For each object properties z, such that q = indexOf(z) and $W[q] > 0$, $\mathcal{A}(X,W)_q$ corresponds to the adjacency matrix built considering only edges of the multi-relational network $M_X = (V_X, \mathbb{E}_X)$ which correspond to the property z.

Definition 3 ***(Flattening function).*** *Let* A *be a multi-dimensional weighed adjacency matrix and i,j= 1, . . ., n. The flattening function* $F : \{0,1\}^{m\times n\times n} \rightarrow \mathbb{N}_0^{n\times n}$ *is defined as follows:*

$$F(A)_{i,j} = \sum_{h=1}^{m} A_{h,i,j}$$

The above flattening function projects a multi-labeled graph to a single-labeled graph. $F(A)_{i,j}$ is equal to n if there are n distinct object properties connecting the resources represented by vertices i and j.

Definition 4 ***(Powers of Adjacency matrix).*** *Let* A *be the adjacency matrix* $\mathbb{N}_0^{n\times n}$. *The power k of an adjacency matrix denoted with* A^k *is defined as*

$$A^k = \begin{cases} A^{(k-1)}A & \text{if } k > 1 \\ A & \text{if } k = 1 \end{cases}$$

where the product between two matrices A *and* B *is* $(AB)_{i,j} = \sum_n a_{i,n} b_{n,j}$.

The power k of adjacency matrix A represents the number of paths of length k among edges of the corresponding graph. For example, $A^k_{i,j}$ represents the number of paths of length k from i to j. Assuming $A = F(\mathcal{A}(X,W))$, $A^k_{i,j} = n$ if and only if in the set of triple X, there are exactly n paths of length k connecting i with j with relevant object properties.

Definition 5 ***(Matrix of Paths).*** *Let* A *be an adjacency matrix,* S^k *is the matrix denoting the number of paths of length* $h \leq k$ *defined as*

$$S^k = \sum_{h=1}^{k} A^h, k \geq 1$$

$\{j|S^k_{i,j} \geq 1\}$ corresponds to the set of vertices that can be visited starting from a vertex i with paths of length minor or equal to k.

This paper adapts the notion of **complementation via a linkset** introduced in [2] to SKOS thesauri. Given two thesauri T_s, T_o and a linkset L linking some concepts in T_s with some concepts in

T_o, T_s can be complemented with T_o via L for k hops, resulting in a third thesaurus identified with $T_s^{L_k}$. Informally, $T_s^{L_k}$ contains all RDF triples of T_s and the SKOS/RDF triples reachable in T_o with path of length less than k via L. Formally, let p be a SKOS property, and S^k the power k of the adjacency matrix associated to T_o we define: $T_s^{L_k} = \{(s,p,o) \mid [t_{T_s}(s,p,o)] \vee [t_L(s, \textit{skos:exactMatch}, y) \wedge t_{T_o}(y,p,o)] \vee [k \geq 2, t_L(s', \textit{skos:exactMatch}, o') \wedge S^{k-1}_{o',s} \geq 1 \wedge t_{T_o}(s,p,o)]\}$. Notice that, $T_s^{L_k}$ and $T_s^{L_k} \cup T_o$ usually differ. The former corresponds to T_s in which triples induced by the *skos:exactMatch* reachable in k hops have been materialized, while the latter also include all the triples from T_o which are not reachable in k hops.

3. Linkset Quality

This section formalizes the *importing* and *reachability* as quality measures which assess linksets as good as they improve a thesaurus with its interlinked concepts and concepts' properties. Thesauri are special kind of datasets, and, in the following we will refer only to thesauri. *Importing* and *reachability* are structured coherently with the well-known quality terminology presented in [6] including **quality indicators**, **scoring functions** and **aggregate metrics**. **Quality indicators** are characteristics in datasets and linksets (e.g., pieces of dataset content, pieces of dataset meta-information, human ratings) which give indication about the suitability of a dataset/linkset for some intended use. **Scoring functions** are functions evaluating quality indicators to measure the suitability of the data for some intended use. **Aggregate metrics** are user-specified metric built upon scoring functions. These aggregations produce new assessment values through the average, sum, max, min or threshold functions applied to the set of scoring functions. In the following subsections, we formalize three indicators and the importing and reachability scoring functions. Moreover, we assume the correctness of thesauri and since it is not the focus of our measures, in fact, our objective is to evaluate the additional information collected by the subject SKOS thesaurus from an object SKOS thesauri through different linksets. We assume also completeness for `skos:exactMatch` linkset, that is, any concept in the subject thesaurus having an exact equivalent concept in the object thesaurus must be involved in a `skos:exactMatch` link and of course in the linkset. Otherwise, our measures might take into account duplicated information and the final evaluation might differ too much from the real one, leading to misleading conclusions. Aggregated metrics on linksets can be defined, for example, combining applications of our scoring functions on different set of parameters.

3.1. Indicators

We present the indicator **val4P** (abbreviation of "values for property") that given an RDF Resource IRI e in a dataset X returns all the values associated to e for a specific RDF property, with the possibility to specify or not (using _) a language tag.

Definition 6 ***(Value for property).*** *Let e be a RDFRiri of dataset X, p be a RDFProp, v and $v@ln$ be respectively in RDFLit and RDFLitLtag, and ln be an ISO language tag or _. We define:*

$$\textit{val4P}_X(e,p,ln) = \begin{cases} \{v \mid t_X(e,p,v)\} & \textit{if } ln = \_ \\ \{v@ln \mid t_X(e,p,v@ln)\} & \textit{otherwise.} \end{cases}$$

Example 1 *Considering thesaurus T_s in Figure 1, val4P$_{T_s}$(x_2, skos:prefLabel, en)={Snake@en}, whilst val4P$_{T_s}$(x_2, skos:prefLabel, _)={Snake@en, Serpente@it}, since in the latter there is no constraint on the language tag.*

Then, given a set of RDF Terms E and a linkset L, we define an operator $[\ ]_L$ that returns the elements of E if they are not involved in any *skos:exactMatch* link, or their linked RDF Terms, otherwise.

Definition 7 ***(Mapping operator).*** *Let L be a linkset, Z be a set of RDFTerms not including blank nodes. The operator $[\ ]_L$ is defined as follows:*
$[Z]_L = \{y \mid z \in Z \wedge (t_L(z, \textit{skos:exactMatch}, y) \vee ((\neg t_L(z, \textit{skos:exactMatch}, y) \vee z \in RDFLit) \wedge y = z))\}$. [4]

Example 2 *Considering thesauri T_s and T_o and linkset L in Figure 1, [{Dog@en}]$_L$={Dog@en}, since {Dog@en} $\subset$ RDFLitLtag$\subset$ RDFLit, and [{x_2,x_5}]$_L$ ={x_2,y_5}, since x_2 has no skos:exactMatch link, and y_5 is the skos:exactMatch-linked RDFTerm for x_5.*

[4]The L's RDF dump or SPARQL endpoint is specified in L's VOID description. Thus $\neg t_L(z, \textit{skos:exactMatch}, y)$ can be verified under the close-world assumption.

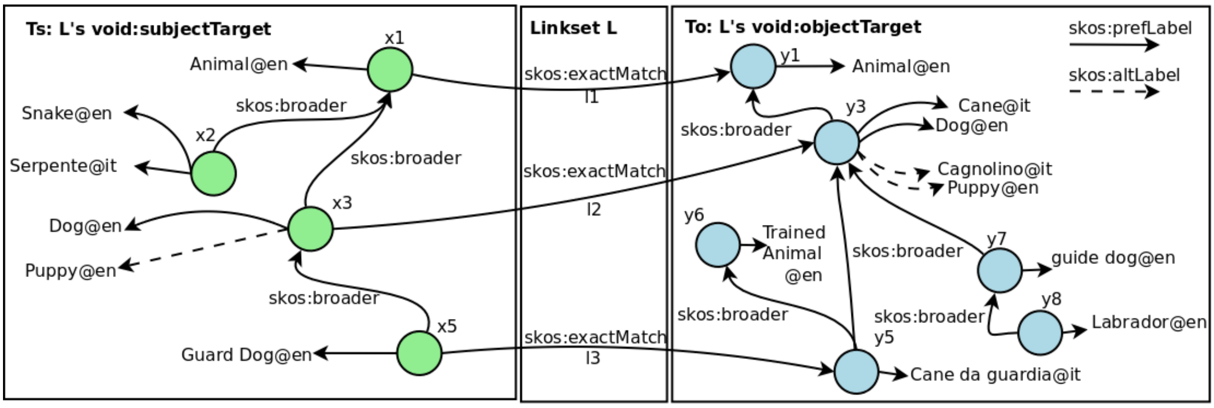


Fig. 1. Example of RDF/SKOS thesauri and *skos:exactMatch* Linkset.

Definition 8 ***(Visited verteces).*** *Let X be a set of RDF triples, W be the array denoting the weights for object properties in X, S be the matrix denoting the number of paths of length $l \leq k$ starting from adjacency matrix $A(X, W)$, $o \in V_X$, VV, the set of verteces visited in k hops, is defined as follows:*

$$VV^k(o) = \begin{cases} \{o\} \cup \{j | j \in V \wedge S^k_{o,j} \geq 1\} & \textit{for } k \geq 1 \\ \{o\} & \textit{for k=0} \end{cases}$$

Example 3 *Considering the thesaurus T_o in Figure 1, and considering relevant the properties skos:narrower and skos:broader, $VV^1(y_3) = \{y_1, y_5, y_7\}$, since y_1, y_5, y_7 are reached in one hop starting from y_3, whilst $VV^2(y_3) = \{y_1, y_5, y_7, y_6, y_8\}$ since considering a further hop also y_6 and y_8 are reached.*

3.2. Scoring functions

Using the indicators presented in the previous section, we define now, the scoring functions characterizing our linkset quality measure.

The *Importing scoring function* evaluates the percentage of "gained values" for a RDF property p. "Gained values" are values not already present in the subject dataset X, but reachable through the linkset L in object dataset Y. *Importing assumes that the linkset correctness has been previously validated.* In the following, we present the importing scoring function for a single link, then, we generalize defining the average importing scoring function for the whole linkset L. The values of all scoring functions in this section are percentages normalized between 0 and 1.

Definition 9 ***(Link importing for property).*** *Let e be in RDFRiri, l be in L and ln be an ISO language tag or _. The link importing for e considering property p through l is defined as follows:*

$$\textit{LinkImp4p}_L\textit{(e,p,l,ln)} = \begin{cases} \textit{0} & \textit{if den} = 0 \\ LI4p_L(e,p,l,ln) & \textit{otherwise.} \end{cases}$$

where

$$LI4p_L\textit{(e,p,l,ln)} = 1 -$$

$$\frac{|\textit{val4P}_X\textit{(e,p, ln)}|}{\underbrace{|[\textit{val4P}_X\textit{(e,p,ln)}]_L \cup \textit{val4P}_{X^L}([\{e\}]_{\{l\}},p,ln)|}_{den}}.$$

Example 4 *Considering the vertex x_3 and its link l_2, the vertex x_5 and its link l_3, and the properties pl = skos:prefLabel, al = skos:altLabel and br = skos:broader showed in Figure 1.*
$LinkImp4p_L(x_3, pl, l_2, \_) =$
$1 - \frac{|\{Dog@en\}|}{|[\{Dog@en\}]_L \cup val4P_{X^L}([\{x_3\}]_{\{l_2\}}, pl, \_)|} =$
$1 - \frac{|\{Dog@en\}|}{|\{Dog@en\} \cup \{Dog@en, Cane@it\}|})$ = 0.5 *and*
$LinkImp4p_L(x_3, al, l_2, en)$= 0 *are, respectively, the normalized percentage of new skos:prefLabel in any language and new skos:prefLabel in English gained by x_3 via l_2* $LinkImp4p_L(x_5, br, l_3, \_) =$
$1 - \frac{|\{x_3\}|}{|[\{x_3\}]_L \cup val4P_{X^L}([\{x_5\}]_{\{l_3\}}, br, \_)|} =$
$1 - \frac{1}{|\{y_3\} \cup \{y_3, y_6\}|} = 0.5$ *is the normalized percentage of broader concepts gained by x_5 via l_3. Only y_6 is gained, since y_3 is considered a duplication of x_3 ($[\{x_3\}]_L$= {y_3}).*

The importing aims at measuring the gain in completeness when complementing via a linkset, thus, it returns 1 if and only if new values from the linked object are imported for an empty subject. The importing is defined below generalizing the link importing on a property and it represent the average importing of all links included in the considered linkset L.

Definition 10 ***(Importing).*** *Let be ln be an ISO language tag or _, the importing of L with respect to p is defined as :*

$$\textit{importing(L,p,ln)} = \frac{1}{|L|} \sum_{e \in \{x | t_L(x,*,*)\}, l \in L} \textit{LinkImp4p}_L\textit{(e,p,l,ln)}$$

The *Reachability* is defined to measure the percentage of vertexes in the object dataset which are reachable cross-walking the links but are not direct object of the links. *The reachability assumes that the linkset is correct and complete*, namely it assumes no link connects unrelated concepts and all subject concepts which have a corresponding object concept are linked. To formally define reachability, we consider the set of verteces in object dataset Y with at least one of the relevant properties, defined as $V_Y|r$= $\{o|(t_Y(s,r,o) \vee t_Y(o,r,s)) \wedge W[\text{indexOf}(r)] > 0\}$), and the set of vertexes involved as objects in links defined as $O = \{o|t_L(s, \text{skos:exactMatch}, o)\}$.

Definition 11 ***(Reachability).*** *Given a linkset L, O the set of vertexes involved as objects in L, W the vector of weight indicating the relevant object property to be considered, and $V_Y|r$ the set of verteces in object dataset with at least one of the relevant properties, the reachability for a path of length minor or equal to k is defined as:*

$$reachability_k(L) = \frac{|(\bigcup_{n \in O} VV^k(n)) \setminus O|}{|V_Y|r \setminus O|}$$

Example 5 *Considering the example shown in Figure 1, and assuming that the skos:broader and skos:narrower are the only object properties weighted greater than zero in the weight matrix, the set of the objects of the link is $O = \{y_1, y_3, y_5\}$, the set of vertexes in the object dataset having at least one of the relevant property is $V_Y|r$= $\{y_1, y_3, y_5, y_6, y_7, y_8\}$ since we assume that the skos inverse relations have been materialized, and as a consequence we have that for each skos:broader in Figure 1 there is the proper inverse skos:narrower and viceversa. The set of vertexes reachable from the linkset in one hop is $\bigcup_{n \in O} VV^1(n)$=$\{y_1, y_3, y_5, y_6, y_7\}$, and then, $r_1(L) = \frac{|\{y_6,y_7\}|}{|\{y_6,y_7,y_8\}|}$=0.66 because y_8 is not reachable in one hops from the links' objects. Moreover, if we consider two hops, $\bigcup_{n \in O} VV^2(n)$=$\{y_1, y_3, y_5, y_6, y_7, y_8\}$ and $r_2(L) = \frac{|\{y_6,y_7,y_8\}|}{|\{y_6,y_7,y_8\}|}$ =1 since all the object vertexes that are not directly linked via the linkset are reachable in two hops.*

4. Validation Framework

The validation aims at demonstrating the ability of importing and reachability, respectively proposed in Definitions 10 and 11 to evaluate the improvement in completeness of a thesaurus when this is complemented via linkset related information. We want to demonstrate, the importing as a good predictor for multilingual gain and the reachability as an estimator of the percentage of new concepts gained cross-walking a linkset. Thus, in the experimentation we focus on the SKOS thesauri and we consider the these two set of SKOS properties for importing $T_{SKOSlabel}$= {`skos:prefLabel`,`skos:altLabel`} and for reachability $T_{SKOSrel}$={`skos:narrower`, `skos:related`, `skos:broader`}.

We address the following research questions:

RQ 1 Do our measures detect linksets that do not bring advantages in term of completeness of the complemented thesaurus? Do our measures detect linksets that bring advantages in term of completeness of the complemented thesaurus?

RQ 2 What about the reliability of the our measures results? When do our measures provide reliable information for completeness? When are they not reliable?

In the following, we introduce the basic concepts of the methodology adopted to validate the importing and the reachability. Then, we present the modular validation architecture discussing in detail each component and the choices made.

4.1. Methodology Motivation and Principles

Due to the novelty of the research field of Linked Data quality, there are only few benchmarks to validate aggregated quality measures and quality score functions (e.g., lodqa[5] and the LACT link specification[6]).

[5] `http://lodqa.wbsg.de/`

[6] `https://github.com/LATC/24-7-platform/tree/master/link-specifications`

Unfortunately, none of them focuses on completeness of complemented thesauri and can be exploited to answer our research questions. For these reasons, our validation methodology relies on a synthetic benchmark built from scratch. The synthetic benchmark is made of a well-defined set of tests. Each test is composed of two SKOS thesauri and one `skos:exactMatch` linkset between them. Thesauri and linksets in the test sets are created altering a seed SKOS thesaurus including a varied kind of situations our metrics should be able to discern (a.k.a., the problem space). Our research questions investigate the relation between importing / reachability and the completeness of the complemented thesauri. Thus, our *ground truth*, is based on the completeness gain reached in the complemented thesaurus. In order to estimate the completeness, we consider the seed thesaurus, which is the most *complete* thesaurus at hand, as the *gold standard*. The design of our validation explicitly addresses the drawbacks identified in [9] affecting synthetic benchmark: (i) **lack of realism:** tests are artificially created to cover a problem space thus they are not necessary representative of all the issues that can be encountered in the reality; (ii) **lack of variability:** it is not possible to vary the seed and the applied transformations; (iii) **lack of discriminability:** tests are not able to really discriminate between working and non working measures, since they are not enough difficult. Concerning the lack of realism, the goal of our system, as already discussed, is to validate the importing and reachability assuming: (i) correctness for thesauri; (ii) correctness and completness for linkset. This assumptions seem reasonable, since, there are various tools we can use to reach them. For example, SILK [13] and LIMES [17] adopt parametric similarity to discover links and provide specific mechanism to restrict or enlarge the similarity criteria enabling in a fine grained tuning of discovery. qSKOS [21] is specifically suited to detect some of the most common issues affecting SKOS thesauri. Moreover, crowdsourcing methods have been deployed to check and to improve the dataset quality [1] and can be deployed to double-checking linkset correctness and completeness.

We delineate the problem space identifying the following critical and interesting situations: (i) the linkset does not provides any importings for the considered SKOS properties or does not reach any new SKOS concepts; (ii) the linkset imports a significant number of new values for the considered SKOS properties or it reaches a significant number of new SKOS concepts; (iii) the linkset covers a very limited number of SKOS concepts exposed in the thesauri.

After that, we define a set of modifiers, specific for importing and reachability, that apply the following macro-alterations: different percentage of deletions in the subject thesaurus, in the object thesaurus and in the linkset. Then, this modifiers are also combined, to create a range of alterations in the input thesauri and in the linkset covering the aforementioned critical situations.

Lack of variability and discriminability are addressed, as suggested in [9], developing a flexible, extensible, open architecture[7]. Our architecture provides a *Test Sets Generator* module that, through a fine tuning of parameters, ensures the possibility to vary the seed thesaurus and to perform random alteration on the seed thesaurus with different precision, with the aim of fully-cover the problem space. In the future, the framework can be extended by third parties to provide further alterators or seed thesauri to enlarge the problem space.

4.2. A Modular Validation Architecture

In this section, we present the general validation architecture, shown in Figure 2, used for both importing and reachability. It has been implemented using Java, and in particular the technology provided by Jena API to manage RDF datasets, and it is made by two main modules: the *Test Sets Generator* and the *Completeness Gain and Measure Assessment* modules.

The *Test Sets Generator* module performs two essential tasks. First, the creation from the gold standard/seed thesaurus T_G of the subject and object thesauri, and of the `skos:exactMatch` linkset between them. Second, the alteration of the subject, object thesauri and of the linkset to generate test sets.

The *Completeness Gain and Measure Assessment* module provides the building blocks for working out the proposed measures (Definitions 11, 10), the complemented thesauri (defined at the end of Section 2), and the completeness gain for complemented thesauri considering the gold standard.

4.2.1. Test Sets Generation.

The goal of this component is to provide an extensive collection of test sets representative enough to possibly fully-cover the problem space. This component takes in input a seed thesaurus T_G, that is elaborated by the *Synthetic Thesauri and Linkset Genera-*

[7]The framework will be available at `http://purl.org/net/linksetq`

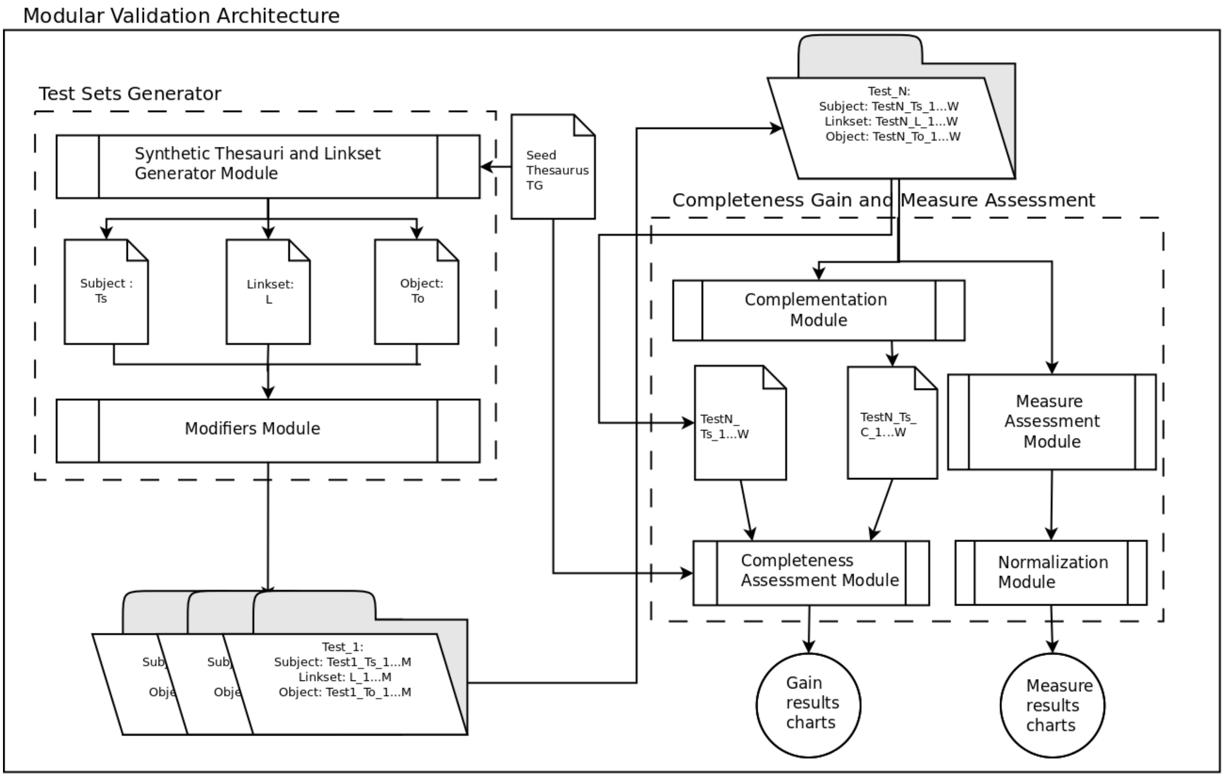


Fig. 2. Modular Validation Architecture

tor module. Such a component duplicates T_G in two thesauri T_s and T_o changing the original namespace in two different namespaces, in order to have the same concepts with the same properties in both thesauri. A linkset *L* is then generated between them. The thesauri T_s, T_o and *L* are taken in input by the *Modifiers* module that creates different test sets applying some alteration on each of the three sets.

We define one modifier for importing: the deletions of certain percentage of `skos:prefLabel` and `skos:altLabel` in a thesaurus. The whole spectrum of possibilities are covered applying different percentages of deletions for `skos:prefLabel` and `skos:altLabel` on subject thesaurus (test set 1), on object thesaurus (test set 2), and for links on linksets (test set 3). Then, the same deletions are combined, altering subject and object thesauri jointly (test set 4), and subject, object and linkset (test set 5), see Table 1 for details. All the importing modifiers preserve correctness of the linkset.

The modifiers for reachability focus on the following SKOS semantic relations `skos:narrower`, `skos:related`, `skos:broader`, considering only the SKOS concepts linked by them. The modifiers for reachability are explained in the following.

– Creation of paths with length K (`mk_k`). The objective is to create randomly a certain numbers of paths with length K in T_o. Each path starts in a `skos:Concept` that is an object of a links in *L*. In order to create paths in T_o, it is necessary to delete some concepts and their associated links in T_s. The procedure starts identifying all the possible paths of length great or equal to K existing in T_s. Then we really construct only a percentage of all the possible paths, in particular, we consider two percentages: 10% and 40%. After this first step, we consider a SKOS concept c_s in T_s that is a subject of a link l in *L* with object c_o, and the root of at least one path of length $\geq K$. Starting from c_s we delete all the SKOS concepts c_{s_i}, belonging to the considered path, related each others by a `skos:narrower` or a `skos:related` or maybe a `skos:broader` and their associated triples. We consider two cases: paths of length 0 (i.e., we do not create apply the modificator `mk_k`) and paths of length 4, since it seems a reasonable number of steps that a user should performs, for example, during the search of a specialization/generalization of a SKOS concept in a thesaurus. We need this modifier in order to be sure that in T_o there exist some paths of length K, and consequently some `skos:Concept` can be reached from c_o only after K steps.

- Deletion of concepts in a thesaurus (`dc`). It randomly deletes a certain percentage of `skos:Concept c` belonging to T_s or to T_o and all the triples involving `c`. It does not consider concepts involved in *L*. We consider two percentages of deletion: 30 and 90.
- Deletion of links in the linkset (`dl`) and of its related concepts the subject/object thesaurus. It randomly deletes a certain percentage of `skos:exactMatch` links in *L* and all the triples in the considered thesaurus (T_s or T_o), involving the `skos:Concept` c_x that can be respectively the subject/object of the link l. We consider two percentages of deletion: 30 and 90. All the modifiers for reachability are designed to preserve both completeness and correctness of the linkset (*L*).

We show, in Figure 3, how modifiers are combined to possibly cover the problem space. The basic idea under the creation of the test sets for reachability is the following: first of all we must be sure that there exists `skos:Concept` reachable only through a path of length K in T_o starting from the object concept of the linkset *L*. This task is performed by the module `mk_k` that consider only T_s and *L*. On the thesaurus and linkset resulting from `mk_k` (T_s_mk_k and L_mk_k) and on T_o we apply the two sequence: (i) `dc` 30% and 90% and then `dl` 30% and 90% and inversely (ii) `dl` 30% and 90% and `dc` 30% and 90%. In Figure 3, T_z can be substituted with T_s or T_o. At the end of a sequence a test, with a specific subject, object and linkset is created. For example, considering T_s, *L*, and T_o using the sequence of modifiers: `mk_k` with k=4, `dc` applied to T_s with 30% and `dl` applied to T_o with 90%, we create a test having the subject thesaurus indicated with *T_s_mk_4_dcs_30_dlo_90*, the object thesaurus with *T_o_mk_4_dlo_90*, and the linkset with *T_o_mk_4_dlo_90*. Considering all the possible combinations of `dl` and `dc` we create 40 different tests.

4.2.2. Completeness Gain and Measure Assessment.

This component evaluates the importing and the reachability wrt the linksets in the generated test sets and the completeness gain of the complemented thesaurus. It relies on the notions of (i) *Thesaurus values restricted to p*, (ii) *Property/Concepts Completeness wrt Gold Standard*, (iii) *Average Completeness Gain wrt Gold Standard*, and (iv) *Normalization factor wrt Gold Standard.*

In the following, we consider the SKOS thesaurus T_h and its associated gold standard T_G. We adapt the concepts of restriction introduced in Definition 11 to SKOS thesauri. Thus, considering the SKOS thesaurus T_h, a specific SKOS property p, the restriction is the set of `skos:Concept` in T_h having the property p.

Definition 12 ***(T_h restricted to p).*** *The restriction of T_h wrt $p \in T_{SKOSlabel} \cup T_{SKOSrel}$ is defined as:* $T_h|_p$= {c | c is a `skos:Concept` $\wedge$ (t_{T_h}(c,p,*) $\vee$ t_{T_h}(*,p,c)})}.

The notion of completeness of a SKOS thesaurus with respect to the gold standard, is derived by property completeness for thesauri [23][10]. More specifically, for importing we refer to the property completeness corresponds to the comparison between the number of values for the SKOS property p in the considered thesaurus and the number of values for p in the gold standard.

Definition 13 ***(Property Completeness wrt Gold Standard).*** *The completeness of T_h wrt T_G for $p \in T_{SKOSlabel}$ is defined as follows:*
$Q_p(T_h, T_G) = \frac{1}{|T_G|_p|} \sum_{t \in T_h|_p} \frac{|val4Prop_{T_h}(t,p,*)|}{|val4Prop_{T_G(t,p,*)}|}$

For reachability we refer to population completeness, see [23], that represent all the necessary objects for a given task, that is, the comparison between the SKOS concepts reachable in a SKOS thesaurus using a SKOS relation r and the SKOS concepts reachable, using r, in the gold standard.

Definition 14 ***(Concepts Completeness wrt Gold Standard).*** *The completeness of T_h wrt T_G for r$\in T_{SKOSrel}$ is defined as follows:*
$Q_r(T_h, T_G) = \frac{|V_{T_h|_r} \cap V_{T_G|_r}|}{|V_{T_G|_r}|}$.

Using the notion of completeness wrt the gold standard we define the average completeness gain wrt gold standard. We refer to a subject thesaurus T_s, an object thesaurus T_o, a linkset L and a gold standard T_G. Then, the average completeness gain calculates the increment of completeness of $T_s^{L_k}$, the result of the complementation of T_s in k hops, with the information contained in T_o, using a specific linkset L. We present a unique definition for importing and reachability.

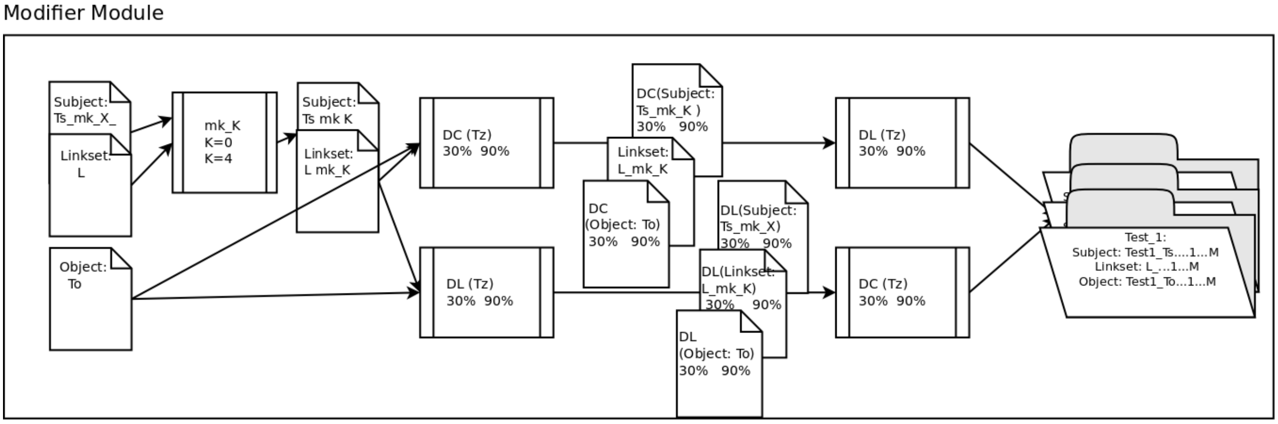


Fig. 3. Modifier module for reachability.

<table>
<tr><td>Test set 1. Alteration of T_s. T_o and L do not change. Deletion in subject thesauri:
Test 1.1: 10% of <code>skos:prefLabel</code> and <code>skos:altLabel</code>;
Test 1.2: 30% of <code>skos:prefLabel</code> and 10% <code>skos:altLabel</code>;
Test 1.3: 60% of prefLabel and 50% <code>skos:altLabel</code>;
Test 1.4: 100% of <code>skos:prefLabel</code> and 0% <code>skos:altLabel</code>;</td><td>Test set 2. Alteration of T_o. T_s and L do not change. Deletion in object thesauri:
Test 2.1: 10% of <code>skos:prefLabel</code> and <code>skos:altLabel</code>;
Test 2.2: 30% of <code>skos:prefLabel</code> and 10% <code>skos:altLabel</code>;
Test 2.3: 60% of prefLabel and 50% <code>skos:altLabel</code>;
Test 2.4: 100% of <code>skos:prefLabel</code> and 0% <code>skos:altLabel</code>;</td></tr>
<tr><td colspan="2">Test set 3. Alteration of L, while, T_s and T_o do not change. Creation of different linkset deleting the 10% (Test 3.1), 30% (Test 3.2), 50% (Test 3.3), 90% (Test 3.4), 99% (Test 3.5) and 99.9% (Test 3.6) of <code>skos:exactMatch</code></td></tr>
<tr><td colspan="2">Test set 4. Alteration of T_s and T_o; L does not change. 8 different combinations of 2 subjects, 4 objects and one linkset. Deletions in T_s:
T_s_1: 90% of <code>skos:prefLabel</code> and 50% <code>skos:altLabel</code>;
T_s_2: 100% of <code>skos:prefLabel</code> and 90% <code>skos:altLabel</code>;
and for T_o:
10% of <code>skos:prefLabel</code> and 10%<code>skos:altLabel</code>;
30% of <code>skos:prefLabel</code> and 10% <code>skos:altLabel</code>;
60% of <code>skos:prefLabel</code> and 50% <code>skos:altLabel</code>;
100% of <code>skos:prefLabel</code> and 90% <code>skos:altLabel</code>;
The test sets: (i) Test 4.1/4.2/4.3/4.4 has T_s_1 as fixed subject thesaurus and change the object thesauri; (ii) Test 4.5/4.6/4.7/4.8 has T_s_2 as fixed subject thesaurus and change the object thesauri; the linkset L is the same for all the tests.</td></tr>
<tr><td colspan="2">Test set 5. Alteration: T_s and T_o and L, 48 different combinations of 2 subjects, 4 objects 6 and linksets. Modification for the subject T_s: (i) T_s_1: 90% of <code>skos:prefLabel</code> and 50% <code>skos:altLabel</code>; (ii) T_s_2 100% of <code>skos:prefLabel</code> and 90% <code>skos:altLabel</code>.
Modification for the object T_o: (i) T_o_1: 10% of <code>skos:prefLabel</code> and 10%<code>skos:altLabel</code>; (ii) T_o_2: 30% of <code>skos:prefLabel</code> and 10% <code>skos:altLabel</code>; (iii) T_o_3: 60% of <code>skos:prefLabel</code> and 50% <code>skos:altLabel</code>; (iv) T_o_4: 100% of <code>skos:prefLabel</code> and 90% <code>skos:altLabel</code>.
Modification for the linkset$L$ deleting: (i) L_1: 10% of <code>skos:exactMatch</code> ; (ii) L_2: 30% of <code>skos:exactMatch</code> ; (iii) L_3: 50% of <code>skos:exactMatch</code> ; (iv) L_4: 90% of <code>skos:exactMatch</code> ; (v) L_5: 99% of <code>skos:exactMatch</code>; (vi) L_6: 99.9% of <code>skos:exactMatch</code>.
Test sets organized in groups for X=1, . . ., 5 as follows $Test_5.X$ groups a subset of tests where only the linkset L_X is fixed, while from $Test_5.X_5.1$ to $Test_5.X_5.4$ the subject is fixed as T_s_1 and the object change from L_1 to L_6; and, from $Test_5.X_5.5$ to $Test_5.X_5.8$ the subject is fixed as T_s_2 and the object change from L_1 to L_6.</td></tr>
</table>

Table 1

Description of the Test sets generated.

Definition 15 ***(Average Completeness Gain wrt Gold standard).*** *Let $T_s^{L_k}$ the complemented thesaurus of T_s, p in $T_{SKOSprop}$ and r in $T_{SKOSrel}$. Thus, considering $Q_\sigma \in \{Q_p, Q_r\}$, the average completeness gain of T_s wrt T_G is defined as follows:*
$Gain(Q_\sigma, T_s, T_s^{L_k})= Q_\sigma(T_s^{L_k}, T_G) - Q_\sigma(T_s, T_G)$.

As highlighted in [23], in order to deal with completeness it is necessary to consider the "close-world" assumption where a gold standard is available and it is used as reference thesaurus. In the real situations considered for assessing linkset quality, an explicit gold standard is usually not available, thus, our metrics are defined considering only the thesaurus at hand, even if the dimensions of the thesaurus at hand and of the gold standard might be very different. Due to this possible difference, some sort of dimensional harmonization, defining a proper normalization factor, is needed to compare the $Gain$ and our measures. The normalization factor for importing is derived from linkset completeness [11] and linkset coverage [2]. It represents the *coverage* of the linkset wrt the entities in the gold standard with property values for p, it is based on the cardinality of the linkset (i.e., $|Links(L)|$) and on the cardinality of the gold standard restricted to property p (i.e., $|T_G|_p|$).

Definition 16 ***(Linkset Property Coverage wrt Gold standard).*** *The the linkset coverage of property p wrt T_G is defined as follows:*
$C_p^L(T)= |Links(L)|/|T_G|_p|$.

The basic idea for reachability is to consider all the SKOS concepts reachable in the object thesaurus through the linkset. Thus, the normalization factor for reachability is the comparison between the number of concepts involved in at least in one of the considered SKOS relations belonging to the object thesaurus T_o wrt those belonging to the gold standard T_G. More formally, the definition is based on the $V_{T_o|_r}$ and $V_{T_G|_r}$ that represent, the verteces (concepts) of, respectively, the RDF triples sets T_o and T_G, reachable through the relation r. Moreover, since the purpose is to consider the concepts only reachable through certain SKOS relations, we discard the concept involved in the linkset that are surely reachable through a `skos:exactMatch` relation, thus we need also the set $O = \{o | t_L(s, \text{skos:exactMatch}, o)\}$ (already defined in 11).

Definition 17 ***(Concepts Coverage wrt Gold standard).*** *The concept coverage of relation r in T_o wrt T_G is defined as:*
$C_r^L(T_o, T_G)= \frac{|V_{T_o|_r} \setminus O|}{|V_{T_G|_r}|}$.

The *Completeness Gain and Measure Assessment* component takes in input all the tests created by the *Test Sets Generator*. Thus, let consider a test set *TestN*, let $TestN\_i$=< *TestN_T_s_i, TestN_L_i, TestN_T_o_i*> (i.e., <subject_thesaurus, linkset, object_thesaurus>), where each of its triples is generated as described in Table 1 and in Figure 3. The measure (importing or reachability) is evaluated directly on the linksets, considering the triple $TestN\_i$. On the other side, the completeness evaluation requires a further step, the generation of the complementation of the subject thesaurus *TestN_T_s_i* with the object thesaurus *TestN_T_o_i* via *TestN_L_i*. We identify with *TestN_T_s_C* the result of the *Complementing Module*. Then, the *Completeness Gain Assessment Module* takes in input the subject thesaurus *TestN_T_s_i* and its complemented *TestN_T_s_C* and calculates the average completeness gain.

5. Experimental Results

The goal of the experimental evaluation is to investigate the effectiveness of normalized importing and normalized reachability (in the following simply importing and reachability) in the evaluation respectively of the property and concepts completeness gain (in the following, simply property gain and concepts gain) of a thesaurus complemented via a specific linkset. We analyse the behaviour of these two functions on the synthetic test sets presented in Section 4. We have considered the GEneral Multilingual Environment Thesaurus (GEMET) as seed thesaurus. GEMET is a `cc-by` licensed thesaurus which includes 5220 `skos:Concept` with `skos:prefLabel` and `skos:altLabel` in more than 30 languages; and several SKOS semantic relations among SKOS concepts.

In the experiments for the importing, as already discussed, we focus on the properties `skos:prefLabel` and `skos:altLabel`. However, considering that GEMET has a quite limited number of `skos:altLabel` and that the analysis of results for the two properties brings to similar conclusions, in the rest of the section, we focus on the result for p =`skos:prefLabel`. Figure 4 shows on the x axis all the tests considered and on the y axis the values for normalized importing and property gain.

The two curves have pretty the same behaviour and the normalized importing results being an upper bound for the property gain. Moreover, since the normalization factor we adopt is considered a quality measure in [11] and [2], we compare and analyse the behaviour of all the three distinct curves (gain, importing and normalization factor), showing the result in Figure 5.

In the experiments on the reachability, we have considered `skos:broader`, `skos:narrower` and `skos:related` as the relevant properties when building the multi-relational weighted matrix. The Figures 6, 7, 8 show on the y axis the values for the concepts gain and the normalized reachability, and on the x axis all the tests considered, respectively tests with the 40% and 10% of the paths of length 4 created (modifier mk_4), and tests without path creation (modifier mk_0). The figures show that the normalized reachability behaves similarly to the concepts gain, and in particular, it is a lower bound for the concepts gain.

Now we discuss importing and reachability behaviour with respect to the research questions we want to investigate.

RQ1: Do our measures detect linksets that do not bring advantages in term of completeness of the complemented thesaurus? Do our measures detect linksets that bring advantages in term of completeness of the complemented thesaurus? *RQ1-1: Considering the case in which the linkset does not provide any importing for the considered properties or any new SKOS concepts.* As general observation both measures have a good behaviour in particular when the value of gain is very low or equal to zero. So, we can say the measures correctly detect when a linkset is not bringing advantages in term of gain.

Importing. Considering Figure 4, for test sets 2 and 3, both importing and property gain are zero, in fact, the subject is complete, consequently we do not import any information. This is an interesting example showing that an high number of links, 5220 in this case, does not necessarily mean a good linkset quality.

Reachability. We distinguish between tests created by `mk_4` from tests created by `mk_0`. In the former tests (Figure 7 and Figure 6), the subject thesaurus is never complete, since we have deleted some concepts to create paths of length 4. As a consequence, the concepts gain should never be zero. Considering both test sets generated with mk_4, substantially concepts gain and reachability are quite zero ($\leq$0.1) for tests number 2, 5, 6, 24, 26, 34, 36 in Figures 6 and 7, plus tests 9 and 10 in Figure 6. In all these tests the object thesaurus has been heavily reduced through the modifier `dco_90`, that deletes the higher number of concepts. Moreover, in some tests also the linkset is reduced, thus, a very low value for concepts gain and reachability is reasonable. For tests created by `mk_0` the subject of some tests is complete so the concepts gain is exactly zero, as shown in Figure 8. It is useful to highlight that the thesauri and linkset originated by `mk_0` have such characteristic: each concept in the subject is connected with the correspondent concept in the object, thus the linkset is made by 5220 links, one for each concept. The modifiers `dcs_x` and `dco_x` when directly applied to these sets do not perform any modifications since they do not delete concepts involved in the linkset. Only tests in which the modifiers `dls_x` or `dlo_x` are applied first can give some gain. This is the reason why many tests in Figure 8 have concepts gain and reachability equal to zero, in all these cases the subject is complete, thus, the linkset does not reach any new information.

RQ1-2: Considering the case in which the linkset provide importing for the considered properties or new SKOS concepts. Both the measures are greater than zero when the gain is greater than zero. So, we can say the measures correctly detect when a linkset is bringing advantages in term of gain.

Importing. As shown in Figure 4, the importing is an upper bound of the property gain. It correctly detects the linkset which brings new values, in fact, as it can be observed in Figure 4, for each peak in the curves of the property gain there is a correspondent peak in the curve of the importing.

Reachability. The reachability (see Figures 7, 6) is a lower bound of the concepts gain. Thus, reachability identifies a minimum level of quality of the linkset considered. This is a useful and safety information for producers and consumer that can be sure to reach at least a certain level of quality through a specific linkset. This group contain all the tests not included in the *RQ1-1*.

RQ2: What about the reliability of the our measure results? When does our measure provide reliable information for completeness? When is it not reliable? In general, we can observe that for low val-

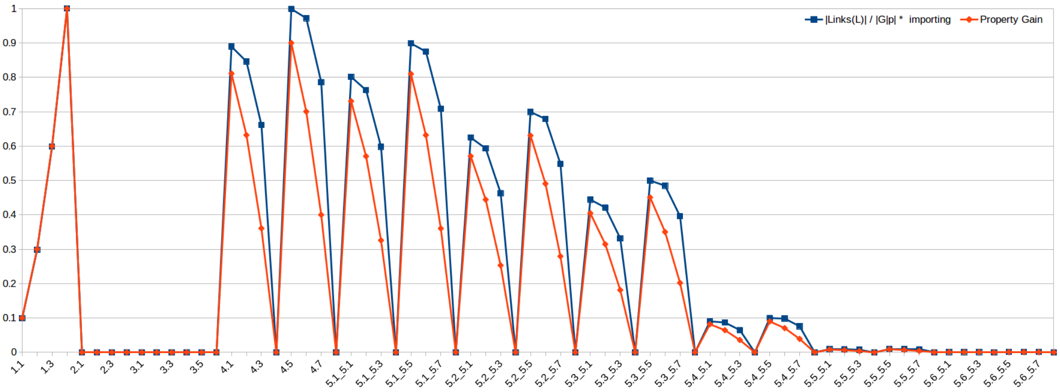


Fig. 4. Normalized importing and Property Gain considering `skos:prefLabel`

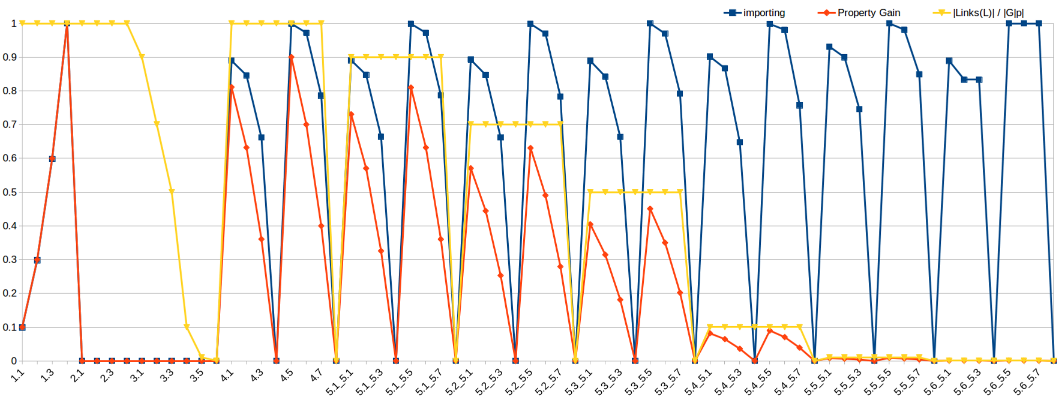


Fig. 5. Importing, Property Gain and C_r^L considering `skos:prefLabel`

ues of gain both measures are very reliable, while, for high values of gain the error between the gain and the measures increase. We guess that this situation probably depend on the closeness between the complemented thesaurus and the gold standard, but, for reachability this needs some further investigation to be better explained.

Importing. Normalized importing is an upper bound of property gain; it distinguish the situations in which we do not have any gain from those in which the gain is greater than zero. Moreover, in the situations in which the property gain is greater than zero the dimension of the error of the normalized importing depends on the closeness between the complemented thesaurus and the gold standard property values. We can observe that:

1. when the complemented thesaurus is equal to the gold standard the values of importing and property gain are the same; as shown in Figure 4 for test sets 1.1, 1.2, 1.3. In these tests, we delete information only in the subject, thus, through the interlinking we capture in the object thesaurus all the information necessary to complete the subject; in this way, the complemented thesaurus become exactly equal the gold standard;
2. the error of importing increases proportionally to the difference between the complemented thesaurus and the gold standard. In fact, considering the importing curve, each peak correspond roughly to a different test set. Analysing each peak (see Figure 4), we observe that the difference between property gain and importing is high when the number of test inside a test set is high (e.g., for test set 4, the difference for test 4.3 is higher than for test 4.1). Considering Table 1, we discover that for high test number, inside the same test sets, we increase the percentage of deletions we apply to subject, object and linkset. These deletions raise the difference between complemented thesaurus and gold standard.

Reachability. The normalized reachability is in general a lower bound for the concepts gain. In particular, the measure is more precise when the concepts gain is low and in particular is zero when the concepts gain is zero. For tests in which the concepts gain is higher than 0.4, the maximum error between the value of the two functions is 0.2. Thus, reachability is able to correctly discriminate when two linksets have very different values of reachability, while when the two values are similar the function should not be totally reliable. As a consequence, a refinement of the error evaluation is needed to better discriminate all the situations. We guess that also for reachability there is a correspondence between error and the closeness of complemented thesaurus and gold standard, that should be further investigated.

As clearly shown in Figure 5, the property gain and the importing without normalization are affected by the linkset coverage, that is our normalization factor. A low coverage means few links which import `skos:prefLabel` values. It implies a low property gain, since the majority of the considered concepts are not involved in the linkset and can not be improved by complementation. But, these few links might import a significant percentage of `skos:prefLabel` values for single link. In this case, the not normalized importing is high, but, it is representative only of a small subset of the concepts involved in the linkset, so it might be misleading for the property gain of the overall complemented thesaurus. Comparing the normalized importing in Figure 4 with the linkset coverage in Figure 5, we can observe that the normalized importing is a better property gain estimator than the interlinking completeness/ linkset coverage.

6. Application

A prototype of the importing and reachability scoring functions has been implemented in JAVA/JENA, and applied to evaluate the quality of linksets among environmental SKOS thesauri. We focus on two linksets, the first is a linkset from EARTh to GEMET (E2GEM with 4365 links) and the second is a linkset from EARTh to Agrovoc (E2AGR with 1436 links). Both linksets have EARTh as linkset object, which is a thesaurus of 14351 concepts, and they have respectively GEMET (5220 concepts) and AGROVOC (over 32000 concepts) [7] as object datasets. E2GEM and E2AGR have been created and validated in the context of eENVplus project [3]. Our purpose is to investigate (i) which of the two linksets imports in EARTh the greater number of `skos:prefLabel` / `skos:altLabel` in different languages and (ii) which of the considered linksets is better in terms of new concepts that can be reached through thesauri cross-walking.

In real applications, there is no a gold standard against which the completeness of complemented thesaurus can be evaluated. Thus, the part of normalization coefficient which refers to gold standard has to be estimated according to the need of specific applications. In order to assess the importing, we can assume that a thesaurus having one `skos:prefLabel` and one `skos:altLabel` for each language and concept is multilingually complete. Thus, to measure the multilingual importing capability of a linkset we can consider the $|T_G|_p|$ (Definition 16) equal to the number of concepts of the subject thesaurus. In our application, we estimate the coverage of the linkset wrt gold standard with the function $|Links(L)|/|ConceptInEARTh|$. The results are shown in Figure 9, where radial axes include: (i) one axis for each considered language, and (ii) an axis "total" representing the average gain independently from the language. Focusing on `skos:prefLabel`, see Figure 9.a, the linkset E2GEM has an higher multilingual gain than E2AGR, in fact, it imports in EARTh more values for a greater number of languages. On the other hand, considering importing for `skos:altLabel`, see Figure 9.b, E2AGR has a very modest multilingual gain, but it imports in EARTh `skos:altLabel` for languages that can't be obtained with E2GEM. So, the choice of the best linkset is not trivial, it depends on the specific languages in which we are interested. For example, considering the importing for `skos:prefLabel` (Figure 9.a) the set of languages imported from E2GEM largely differs from those importable via E2AGR. In fact, only 10 out of the 41 considered languages are importable from both linksets (i.e., ar, ru, es, tr, pt, pl, de, fr, hu, cs), about 19 out of 41 (e.g., bg, ga, fi, sl, eu, ro) can be imported only considering E2GEM, and 8 out of 41 only from E2AGR. While, for `skos:altLabel` (Figure 9), we import about 20 out of 41 languages from E2AGR and about 4 of 41 from E2GEM. The maximum multilingual gain obtained with E2AGR and E2GEM is around 0.1 and 0.3 respectively, which implies that complementing EARTh via the aforementioned linkset we might earn

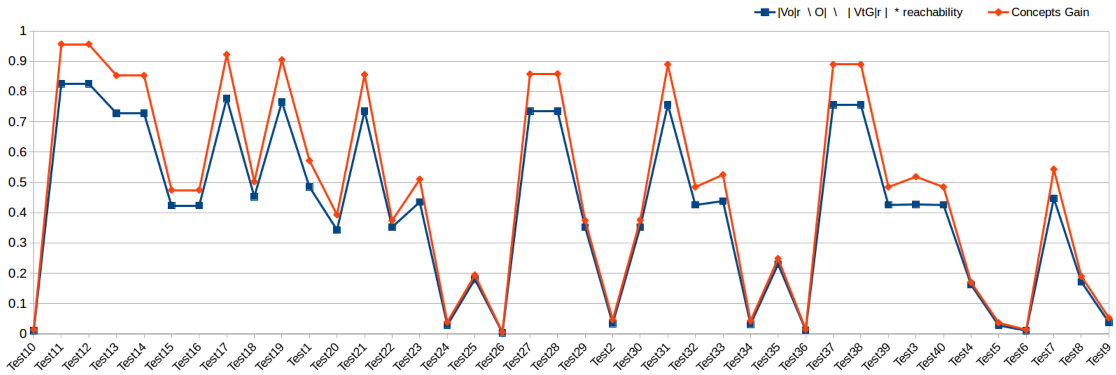


Fig. 6. Normalized reachability vs Concepts Gain, results on test sets created with 40% of 4-hops-length paths.

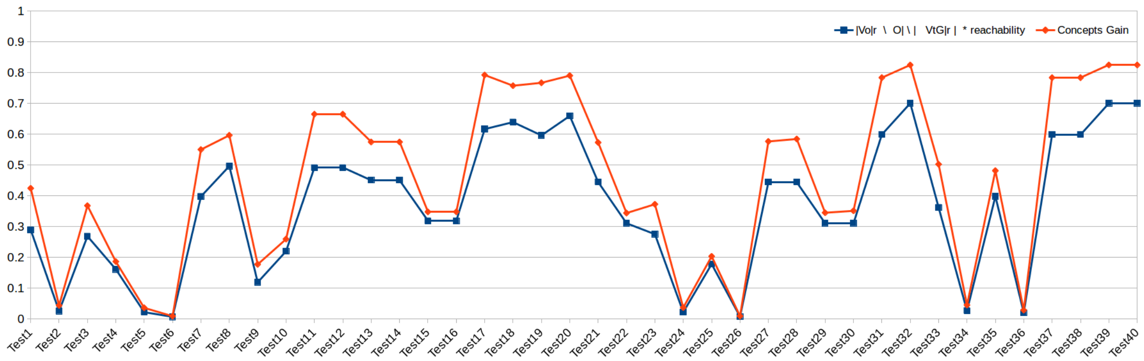


Fig. 7. Normalized reachability vs Concepts Gain, results on test sets created with 10% of 4-hops-length paths.

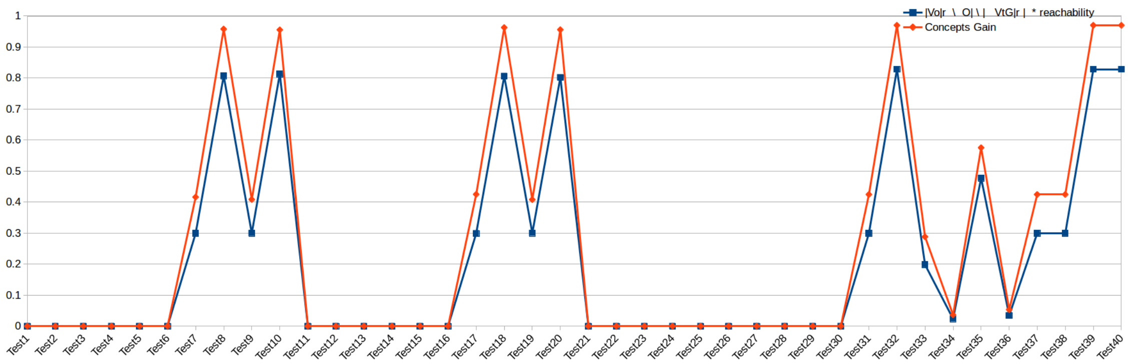
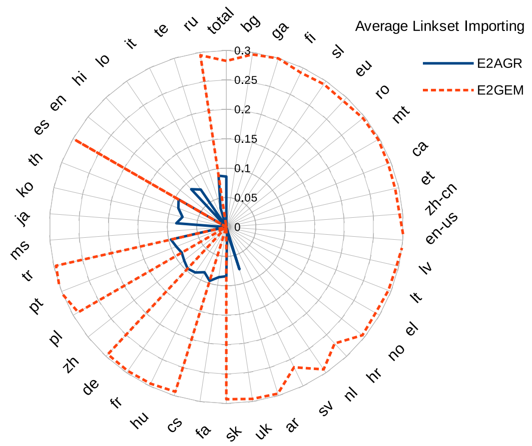


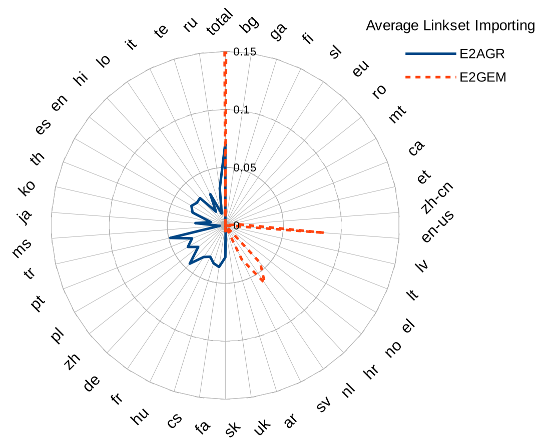
Fig. 8. Normalized reachability vs Concepts Gain, results on test sets created with 0% of 4-hops-length paths.

at maximum the 10% and 30% of the overall translations in the complemented dataset.

Concerning reachability, we have different possibilities to set the normalization coefficient. For example, if we want to estimate the percentage of concepts gained in the subject's browsing space through the linkset cross-walking, we can set the coefficient $V_{T_G|r}$ in Definition 17 equal to the number of concepts available in the subject dataset. In this way, a reachability equals to 1 implies we have doubled the number of concepts in the subject thesaurus browsing space; a reachability equals to 1.5 implies we have gained one and half the number of concepts in the subject dataset, and so on. Alternatively, we can set $V_{T_G|r}$ as the goal number of concepts we wish to add in the subject thesaurus browsing space. In this case, the reachability represent the closeness to the prefixed goal: a reachability value equal to 1 implies the complete achievement of the ob-



(a) skos:prefLabel



(b) skos:altLabel

Fig. 9. Importing for E2AGR and E2GEM normalized by linkset coverage (i.e., $|Links(L)|/|ConceptInEARTh|$)

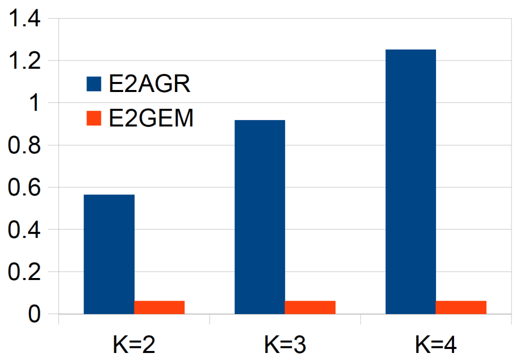


Fig. 10. Reachability for E2AGR and E2GEM normalized with $(NumOfConceptsInObject - |Links(L)|)/|ConceptInEARTh|)$

jective. The reachability has been applied on the linksets E2GEM and E2AGR for K = 2, 3, 4 and the results are shown in Figure 10. It is evident that E2AGR outperforms E2GEM: independently from the K considered, E2GEM gains less than 0.1 (namely, 10% of the EARTh concept cardinality), while E2AGR gains a minimum of 0.5 (50%) when $K \geq 2$ up to the 1.2 (120%) when $K = 4$, which means that the EARTh browsing space is more than doubled.

The limited contribution of E2GEM in enriching the EARTh browsing space is not very surprising, since EARTh can be considered as an extension and a revision of GEMET [3]. In fact, more than 4000 concepts of the GEMET concepts have been natively included in EARTh, thus only few concepts of GEMET can enrich the EARTh browsing space.

In conclusion, it is important to remark that even if the number of links in the linkset has been often considered as a rough measure of linkset quality and our two measures offer a far better insight in linkset quality. In fact, just considering the number of links, E2GEM outperforms E2AGR, whereas, as shown in Figure 9 and in Figure 10, the linkset quality largely vary depending on the considered complementation purpose (e.g., improving a thesaurus multilingualism or enriching its browsing space). Although E2GEM has more links than E2AGR, E2AGR drastically outperforms E2GEM for enriching the browsing space, see Figure 10, and it imports much more `skos:altLabel` for languages such as polish (pl) , japanese (ja), portuguese (pt), see Figure 9.b.

7. Related work

A recent systematic review of quality assessment for linked data can be found in the SWJ submission [23] and in the deliverable produced by EU funded project PlanetData [15]. They reviews quality dimensions which are traditionally considered in data and information quality (e.g., availability, timeliness, completeness, relevancy, availability, consistency), as well as more Linked Data specific dimensions such as licensing and interlinking. Among the measures reviewed in the aforementioned works, we discuss the measures for completeness and interlinking closely related to the contribution of this paper. In [23], completeness is measured in terms of (i) *schema completeness*, the degree to which the classes and the properties of an ontology are represented [10] [16]; (ii) *property completeness*, the measure of the missing values for a specific property [10]; (iii) *population completeness*, the percentage of all real-world objects of a particular type that are represented in the datasets [10,16]. These measures basically correspond to the notions of intensional, extensional and LDS completeness discussed in [15] that are defined for datasets, not for linkset quality. In particular, we have considered the *property completeness* and the *population completeness* to define, respectively, property and concepts gain of the complemented thesaurus.

Framework LINK-QA [11], considered also in [23] for the interlinking quality dimension, defines two network measures specifically designed for Linked Data (Open SameAs chains, and Descriptive Richness) and three classic network measures (degree, centrality, clustering coefficient) for determining whether a set of links improves the overall quality of Linked Data. Our importing and reachability substantially differ from the scoring functions proposed in LINK-QA for these reasons: (i) LINK-QA works on links independently from the fact that they are part or not of the same linksets; (ii) LINK-QA addresses correctness of links, it does not deal with gain in completeness of the complemented dataset. [8] In particular, the Descriptive Richness measures how much is added to the description of a resource through the use of a `owl:sameAs`, which is very close to the principle behind our importing and reachability. However, Descriptive Richness is somehow a coarse aggregation of our measures, which does not distinguish among the different kinds of information that can be added, the number of hops to consider, thus it does not seem useful in targeting the scenario considering the enrichment of thesaurus multilingualism and the enrichment of browsing space in k hops. As pointed out in [23], LINK-QA also proposes an interlinking completeness which determines the degree to which entities in the dataset are interlinked. This completeness measure is closely related to our previous work [2], that proposes a set of scoring functions for assessing `owl:sameAs` linkset quality. It strongly relies on the notion of types, namely classes of the entities exposed in a dataset and its related linksets. In particular, linkset entity coverage for type specializes the interlinking completeness proposed in [11], grouping entities according to their type. The measure $|Links(L)|/|T_G|_p|$ used to normalized the proposed *importing measure* is derived by these previous coverage functions. While, the *importing measure* goes beyond the linkset completeness and linkset entity coverage for type, in fact, it measures the link contribution in the completness of the complementated thesaurus and not only the average number of links for entities.

More recent works are LiQuate [19] and [12] which aim at investigating the quality of data and links in life science Linked Data from Bio2RDF project. LiQuate [19] relies on Bayesian Networks to answer to quality validation requests including the probability of incomplete and inconsistent links among a set of given resources. [12] conducts an empirical link analysis deploying network metrics such as degree, average distance and clustering coefficient to analyze three types of graphs: the graph of datasets, entities and terms. Neither LiQuate nor [12] attempt to characterize the quality of linkset in terms of the added value a linkset can bring when complementing a dataset.

A set of quality measures specific for SKOS thesauri relevant for this paper have been proposed in [21]. The paper summarizes a set of 26 quality issues for SKOS thesauri and shows how these can be detected and improved by deploying qSKOS [14], PoolParty checker, and Skosify [20]. Among the mentioned issues, incomplete language coverage is particularly worth for our work. Incomplete language coverage arises when the set of language tags used by the literal values linked with a concept are not the same for all concepts. Our measure assesses the goodness of a linkset when complementing a thesauri to import further `skos:altLabel` and `skos:prefLabel`, and might represent a shortcut to address incomplete language coverage. Unfortunately, an analysis on linksets among thesauri is not included in [21]: missing out-links and in-links are adopted as indicators of SKOS thesaurus quality, but, their potential for complementation, in terms of reachability of new concepts or importing of `skos:prefLabel` and `skos:altLabel` values, when dealing with incomplete language coverage, is not considered.

8. Conclusions and Future Work

In this paper, we address an aspect of linkset quality, while the majority of existing works focus on dataset quality. We state that, in the evolution of the Web of Data into the Global Data Space, linksets should have the same importance of datasets, thus, linkset quality should be considered as an independent branch of

[8] The quality dimensions addressed by LINK-QA are not explicitly stated. We exclude that LINK-QA considers completeness, since, it tries to correlate network measures and bad link detection.

Linked Data quality, and not simply as one of the quality dimensions.

We propose two linkset quality measures, the *importing* and the *reachability*, which evaluate `skos:exactMatch`-linkset potential when complementing datasets with their interlinked information. The reachability and importing estimate, respectively, the ability of a linkset to enrich a thesaurus with new concepts and with new property values. Our measures assumes SKOS core and inverse properties materialization as well as the correctness of the thesauri and linksets. The reachability also requires the completeness of the linksets.

We experimentally relate our measures to the completeness gain: the gain of new concepts and new property values obtained by cross-walking a linkset. Unfortunately, as far as we know, there not exist benchmarks suited for testing the completeness of complemented thesauri. Thus, in order to validate our measures, we have built a benchmark from scratch addressing as much as possible the drawbacks usually affect synthetic benchmarks.

The validation shows that the normalized measures, namely the *importing* and *reachability* multiplied for linkset coverage and concept coverage respectively, are good predictors for completeness gain in the complemented thesaurus. Both the normalized measures discern correctly whether or not a linkset brings some concepts or property completness gain. The normalized importing is an upper bound for the completeness gain, and its error increases as it increases the differences between the complemented thesaurus and desired complete thesaurus. The normalized reachability is a lower bound for concept completeness gain in the complemented thesaurus. The error of normalized reachability, even if not fully characterized, is in average rather small, its maximum results be lower than 20%. Thus, the measure is reliable when comparing linksets with quite distant reachability values, but, it is less reliable when reachability value are very close.

We provide an example of real application of *importing* and *reachability* to evaluate the gain in term of multilingual labels and browsing space on real linksets developed in the EU project eENVplus. The example demonstrates that our metrics provide an insight into the linkset potential.

As future work, we plan to further and better analyze and characterize the estimation errors for both importing and reachability in order to improve their reliability. We also plan to apply our measure for evaluate the quality of other `skos:exactMatch` linksets existing in the context of LOD cloud, in order to start a sort of quality analysis of interlinking in the LOD cloud. Moreover, issues concerning the application of our measures to other SKOS properties have also to be investigated. We plan to investigate the application of our measures on `owl:sameAs` linksets and to possibly extend the characterization of linkset quality considering further dimensions and defining new scoring functions.

Acknowledgements. This research activity has been partially carried out within the EU funded project eENVplus (CIP-ICT-PSP grant No. 325232).

References

[1] M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, S. Auer, and J. Lehmann. Crowdsourcing linked data quality assessment. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II*, volume 8219 of *Lecture Notes in Computer Science*, pages 260–276. Springer, 2013.

[2] R. Albertoni and A. Gómez-Pérez. Assessing linkset quality for complementing third-party datasets. In *EDBT/ICDT Workshops*, pages 52–59. ACM, 2013.

[3] R. Albertoni, M. D. Martino, S. D. Franco, V. D. Santis, and P. Plini. Earth: An environmental application reference thesaurus in the linked open data cloud. *Semantic Web*, 5(2):165–171, 2014.

[4] R. Albertoni, M. D. Martino, and P. Podestà. Environmental thesauri under the lens of reusability. In *EGOVIS 2014*, volume 8465 of *Lecture Notes in Computer Science*, pages 222–236. Springer, 2014.

[5] R. Albertoni, M. D. Martino, and P. Podestà. A linkset quality metric measuring multilingual gain in SKOS thesauri. In A. Rula, A. Zaveri, M. Knuth, and D. Kontokostas, editors, *Proceedings of the 2nd Workshop on Linked Data Quality co-located with 12th Extended Semantic Web Conference (ESWC 2015), Portorož, Slovenia, June 1, 2015.*, volume 1376 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2015.

[6] C. Bizer and R. Cyganiak. Quality-driven information filtering using the WIQA policy framework. *J. Web Sem.*, 7(1):1–10, 2009.

[7] C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques, and J. Keizer. The AGROVOC linked dataset. *Semantic Web*, 4(3):341–348, 2013.

[8] R. Cyganiak, J. Zhao, M. Hausenblas, and K. Alexander. Describing linked datasets with the VoID vocabulary. W3C note, W3C, Mar. 2011. http://www.w3.org/TR/2011/NOTE-void-20110303/.

[9] J. Euzenat, M.-E. Rosoiu, and C. T. dos Santos. Ontology matching benchmarks: Generation, stability, and discriminability. *J. Web Sem.*, 21:30–48, 2013.

[10] C. Fürber and M. Hepp. Swiqa - a semantic web information quality assessment framework. In V. K. Tuunainen, M. Rossi, and J. Nandhakumar, editors, *ECIS*, 2011.

[11] C. Guéret, P. T. Groth, C. Stadler, and J. Lehmann. Assessing linked data mappings using network measures. In *ESWC*, volume 7295 of *Lecture Notes in Computer Science*, pages 87–102. Springer, 2012.

[12] W. Hu, H. Qiu, and M. Dumontier. Link analysis of life science linked data. In M. Arenas, Ó. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. T. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, and S. Staab, editors, *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II*, volume 9367 of *Lecture Notes in Computer Science*, pages 446–462. Springer, 2015.

[13] R. Isele and C. Bizer. Active learning of expressive linkage rules using genetic programming. *Web Semantics: Science, Services and Agents on the World Wide Web*, 23:2 – 15, 2013. Data Linking.

[14] C. Mader, B. Haslhofer, and A. Isaac. Finding quality issues in skos vocabularies. In P. Zaphiris, G. Buchanan, E. Rasmussen, and F. Loizides, editors, *TPDL*, volume 7489 of *Lecture Notes in Computer Science*, pages 222–233. Springer, 2012.

[15] P. N. Mendes, C. Bizer, J. H. Young, Z. Miklos, J.-P. Calbimonte, and A. Moraru. Conceptual model and best practices for high-quality metadata publishing. Technical report, PlanetData, Deliverable 2.1, 2012.

[16] P. N. Mendes, H. Mühleisen, and C. Bizer. Sieve: linked data quality assessment and fusion. In D. Srivastava and I. Ari, editors, *EDBT/ICDT Workshops*, pages 116–123. ACM, 2012.

[17] A. N. Ngomo and S. Auer. LIMES - A time-efficient approach for large-scale link discovery on the web of data. In T. Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 2312–2317. IJCAI/AAAI, 2011.

[18] M. A. Rodriguez and J. Shinavier. Exposing multi-relational networks to single-relational network analysis algorithms. *J. Informetrics*, 4(1):29–41, 2010.

[19] E. Ruckhaus, M. Vidal, S. Castillo, O. Burguillos, and O. Baldizan. Analyzing linked data quality with liquate. In V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, and A. Tordai, editors, *The Semantic Web: ESWC 2014 Satellite Events - ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers*, volume 8798 of *Lecture Notes in Computer Science*, pages 488–493. Springer, 2014.

[20] O. Suominen and E. Hyvönen. Improving the quality of skos vocabularies with skosify. In *EKAW*, volume 7603 of *Lecture Notes in Computer Science*, pages 383–397. Springer, 2012.

[21] O. Suominen and C. Mader. Assessing and improving the quality of skos vocabularies. *J. Data Semantics*, 3(1):47–73, 2014.

[22] B. Zapilko, J. Schaible, P. Mayr, and B. Mathiak. Thesoz: A SKOS representation of the thesaurus for the social sciences. *Semantic Web*, 4(3):257–263, 2013.

[23] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer. Quality assessment for linked open data: A survey. *Semantic Web*, 7(1):63–93, 2016.