Perception and Relevance of Quality Issues in Web Vocabularies

Christian Mader
University of Vienna
Department of Computer Science
christian.mader@univie.ac.at

Bernhard Haslhofer
University of Vienna
Department of Computer Science
bernhard.haslhofer@univie.ac.at

ABSTRACT

Web vocabularies provide organization and orientation in information environments and can facilitate resource discovery and retrieval. Several tools have been developed that support quality assessment for the increasing amount of vocabularies expressed in SKOS and published as Linked Data. However, these tools do not yet take into account the users' perception of vocabulary quality. In this paper, we report the findings from an online survey conducted among experts in the field of vocabulary development to study the perception and relevance of vocabulary quality issues in the context of real-world application scenarios. Our results indicate that structural and labeling issues are the most relevant ones. We also derived design recommendations for vocabulary quality checking tools.

1. INTRODUCTION

Controlled vocabularies, such as taxonomies, thesauri, or categorization schemes, provide organization and orientation in information environments and are important instruments for document indexing, tagging, spelling suggestions, and many other automated information retrieval tasks. The AGROVOC thesaurus[1], for instance, contains 40,000 concepts in 22 languages and is used, e.g., for automatic document indexing in the United Nations Food and Agriculture Organization. The New York Times authoritative news vocabulary[2] has been maintained for more than 160 years and drives so-called *topic pages*, which provide access to all relevant articles the New York Times has ever written about a certain subject. Maintainers typically want to achieve high quality of their vocabularies because it has a direct impact on associated computation tasks.

Most vocabularies are still being curated manually by expert users making them prone to errors and mistakes. Recently, they are increasingly getting published as Linked Data on the Web [5], which means that they become available in structured formats and follow agreed-upon representation models such as the Simple Knowledge Organization System (SKOS) [11]. This makes them machine-processable and allows automated quality checkers to become part of vocabulary management systems such as the PoolParty Thesaurus Manager[3], just as spell-checkers are nowadays indispensable components of any text processing software. In the following we refer to such vocabularies as *Web vocabularies*. However, the notion of quality of Web vocabularies is to a great extent domain-specific and depends on the quality perception of the person(s) who curate(s) vocabularies. It might, for instance, be perfectly valid to include circular relations in one vocabulary but not in others.

Recently, a number of quality assessment tools have been published: the *Poolparty consistency checker*[4] implements tests for the six integrity conditions that are defined as part of the SKOS model and introduces custom checks such as URI syntax validation, identification of loose concepts, and missing labels. *Skosify*[5] follows a similar approach but adds identification of circular hierarchical structures and label syntax checks. *qSKOS*[6], which is being developed by the authors of this paper, goes beyond SKOS model integrity conditions. It implements checks for additional quality criteria that were derived from existing literature and guidelines such as identification of valueless associative relations or checks for presence of proper entry points into the concept structure. However, at the moment, none of these tools takes the user's perception of quality into account, where the *user* is meant to be the curator of some vocabulary (e.g., a taxonomist) who uses some vocabulary management system.

The work reported in this paper is a follow-up on our previous work [10], in which we identified 15 sources of possible vocabulary quality degradations, called *quality issues* and analyzed 15 existing vocabularies against these issues. We were able to identify potential quality problems in almost all of them. To find out about the impact of those quality issues, we performed a questionnaire-style survey on the perception of quality from the user's point of view. The goal of the survey was to (i) evaluate known issues from a user perspective and (ii) to explore possibilities to extend and improve quality checks in existing tools. Our contributions

[1] http://aims.fao.org/standards/agrovoc/about
[2] http://data.nytimes.com/

ISEM '13, September 04 - 06 2013, Graz, Austria

http://dx.doi.org/10.1145/2506182.2506184.

[3] http://poolparty.biz/products/poolparty-thesaurus-manager/
[4] http://demo.semantic-web.at:8080/SkosServices/check
[5] http://code.google.com/p/skosify/
[6] https://github.com/cmader/qSKOS

can be summarized as follows:

- We report the findings from an online survey conducted between September 20th and December 6th, 2012. During that time, 163 respondents from 28 countries and several different domains answered questions about the relevance and perception of given quality criteria.
- We introduced a set of usage scenarios for Web vocabularies and identified quality issues which are most important to support these scenarios.
- From these findings we derive design recommendations that can inform the development of vocabulary quality assessment tools.

The remainder of this paper is structured as follows: in the next section, we discuss related work in the area of quality evaluation and briefly describe potential quality issues. Section 3 provides details on the design of our online survey and Section 4 reports our findings from that survey. After presenting derived design recommendations in Section 5, we conclude this paper in Section 6 and offer directions for future research.

2. BACKGROUND

We first elaborate on related work that addresses existing approaches of assessing and assuring quality in Web vocabularies. We then briefly introduce the quality issues from our earlier work on which we rely in this paper.

2.1 Vocabulary Quality Assessment

We can distinguish between two main categories of quality assessment strategies: *intellectual assessment*, which is domain-specific and based on the individual user's quality perception, and *automated analysis*, which focuses on the formulation of metrics and rules for automatic quality assessment.

Intellectual assessment strategies are described in existing standards (e.g., [13]) and guidelines (e.g., [3, 9, 15]), which cover general criteria, such as inclusion of "all needed facets'". In practice, such criteria are also often defined adhoc within the context of a certain project or vocabulary (e.g., [4]).

Automated quality analysis procedures are usually defined as part of existing quality checking tools and bound to the formalism or model a vocabulary is expressed in: the Simple Knowledge Organization System (SKOS), for instance, defines in total six integrity conditions [11], each of which is a statement that specifies under which circumstances data are consistent with the SKOS data model. Suominen et al. [16] provide a comparison of quality criteria identified by three automated checking tools, the *PoolParty thesaurus consistency checker*, *qSKOS* and *Skosify*. To our knowledge, *Skosify* is the only tool that can also automatically correct certain quality issues. Manaf et al. [2] focus on statistical and structural properties of SKOS vocabularies such as number of concepts, maximum hierarchy depth, or SKOS property usage distribution among different vocabularies. However, the authors do not draw any conclusion about the implications of these properties on usability or quality of Web vocabularies.

Data quality has also been discussed more broadly in Semantic Web and Linked Data research. Hogan et al. [7] identify four categories of common errors and shortcomings in RDF documents and Heath and Bizer [5] summarize best practices for publishing data on the Web. In the field of ontology evaluation, catalogs of criteria for assessing ontology quality have been proposed ([14, 17]).

However, most existing quality standards and guidelines do not cover specific requirements of Web vocabularies. On the other hand, work on Linked Data quality mostly focuses on (semi-)formal constraints that can be automatically checked but are not specific for Web vocabularies. In our earlier work [10] we tried to bridge this gap by defining a number of potential quality issues, suitable for automated checking of Web vocabularies. In this paper we aim to evaluate these quality issues from a user perspective and find out about their practical applicability.

2.2 Web Vocabulary Quality Issues

Previously [10], we investigated how tools could better support taxonomists in improving SKOS vocabularies by pointing out quality issues that go beyond the integrity conditions defined in the SKOS specification. We reviewed intellectual quality assessment strategies, examined existing vocabularies, and identified potential quantifiable vocabulary quality issues, which we then formalized into computable quality checking functions that can be used to find affected resources in a given SKOS vocabulary. These issues fall into three main categories — *Labeling and Documentation Issues*, *Structural Issues*, and *Linked Data Specific Issues* — and can automatically be checked by our open-source *qSKOS*[7] vocabulary quality assessment tool.

Due to space limitations we focus on 8 different issues our survey participants identified as being most relevant in certain contexts (cf. Section 4.1). We furthermore provide only a short description of the quality issues and refer to our earlier work for semi-formal definitions.

2.2.1 Labeling and Documentation Issues

The intuition for this category is that SKOS vocabularies should be tagged and documented consistently with human-readable labels, across languages. We identified the following issues:

- *Omitted or Invalid Language Tags*: SKOS labeling or documentation (e.g., `skos:prefLabel`, `skos:note`) properties miss or apply invalid language tags.
- *Label Conflicts*: Concept pairs that have identical preferred, alternative or hidden labels (cf. Figure 1).
- *Undocumented Concepts*: Concepts that have none of the SKOS documentary properties (e.g., `skos:scopeNote`, `skos:example` or `skos:definition`).
- *Number of Synonyms and Non-descriptors*: Concept labels are important for, e.g., search recall and precision, so this metrics determines their number in a Web vocabulary.

[7] `https://github.com/cmader/qSKOS/`

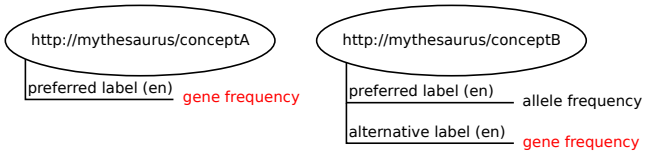

Figure 1: Example of a label conflict

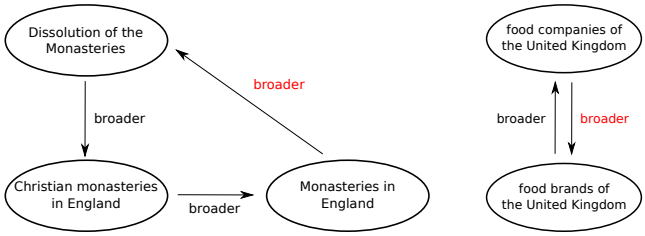

Figure 2: Exemplary hierarchical cycles

2.2.2 *Structural Issues*

SKOS Vocabularies are represented in RDF and therefore form a graph structure, which provides the basis for the following issues:

- *Cyclic Hierarchical Relations*: Concepts related to themselves by (chains of) hierarchical SKOS properties (e.g., `skos:broader`, `skos:narrower`) as shown in Figure 2.
- *Orphan Concepts*: Concepts without semantic relations (i.e. `skos:semanticRelation` or subclasses thereof) to other concepts.

2.2.3 *Linked Data Specific Issues*

Missing Out-Links are concepts that do not refer to third-party content on the Web. In the survey we distinguish further between two kinds of outgoing links:

- *Links to other vocabularies* published on the Web as Linked Data or browsable online.
- *Links to general resources* such as images or websites providing additional information.

3. METHODOLOGY

To learn about the perception and relevance of quality issues from a taxonomist's point of view, we conducted an online survey between Sep 20th and Dec 6th 2012.

3.1 Participants

Our survey targeted practitioners working with Web vocabularies: *vocabulary managers* who curate vocabularies, *contributors* who propose terms to be changed or included, or *users* who have no rights or intentions to change a vocabulary.

The survey was announced on several mailing lists[8] related to Web vocabulary development. We also contacted the Semantic Web Company's customer network and posted an invitation on its blog[9]. In the middle of the scheduled survey period on Oct 29th, we sent reminders via the same communication channels.

In total we received 163 responses with varying coverage because only a few questions were mandatory and some participants did not complete the survey. From 25 participants who indicated their role, 12 were vocabulary managers, one was a contributor and 4 identified themselves as users. Two of these 25 participants explicitly gave no answer and 6 stated other roles with 2 of them giving no exact role description. The maximum number of responses we received for a quality-relevant question was 56, decreasing towards the end of the survey with 28 being the minimum. The majority of the responses came from the US (39), followed by the UK (15) and Italy (10).

3.2 Survey Structure and Question Design

The questionnaire consists of four parts in which we (i) present introductory information, (ii) collect general domain and usage information, (iii) present open and closed-ended questions targeting vocabulary quality, and (iv) collect information about the participants. The analytic and explorative nature of the survey is reflected in the third part: To find out about the usefulness of existing quality issues, closed-ended questions that can be analyzed automatically were used. For exploring additional quality issues or improving existing ones we included open-ended questions that allow us to, e.g., infer rationales for rating decisions or information on the development processes.

We employed two different kinds of closed-ended questions: first, we use multiple choice checkboxes (including an "other" option) e.g., for selecting the domain or usage scenarios of Web vocabularies. Second, we formulated explicit *quality statements* (e.g., "Concepts should not be hierarchically related to themselves.") based on the issues identified in Section 2.2 and asked participants to express their level of agreement on a symmetric 6-point Likert scale, which included a neutral option and the possibility to give no answer. To learn about the participant's decision rationale, every closed-ended question was complemented by a free-text field for providing the decision's rationale. We used a similar symmetric 6-point Likert scale to find out about the relevance of the quality issues in relation to a vocabulary usage scenario (cf. Table 1). Participants were asked to select one of the categories *very important*, *important*, *neither*, *less important*, *not important*, and *no answer/don't know*.

We organized quality statements in three groups (*Labeling and Documentation Issues*, *Structural Issues* and *Linked Data Specific Issues*) and follow this structure when discussing our findings in the remainder of this paper.

3.3 Analysis of Survey Responses

We believe that due to our chosen survey distribution channels, we can trust in our participants expertise. We intentionally did not require the participants to have a background in SKOS so we could reach a wider target audience. Since a meaningful quantitative analysis and statistical interpretation would require a much larger, but hard to collect sample, we concentrated on a qualitative analysis based on the users' responses. To identify the usage scenarios that have been rated most important (average median value below 3) for the provided quality statements, for each statement we computed and sorted the agreement ratings by ascending median, mode, mean and standard deviation and

[8] `DC-VOCABULARY@JISCMAIL.AC.UK`, `NKOS-L@OCLC.ORG`, `public-esw-thes@w3.org`, `public-lod@w3.org`
[9] `http://tinyurl.com/d8wyntj`

Table 1: Vocabulary Usage Scenarios

Name	Description
Manual / Intellectual Indexing	Performed by domain experts who process a corpus of documents and extract relevant concepts
Automatic Indexing	Algorithmic extraction of common words in a text corpus based on statistical measures (frequency, co-occurrence,...)
Tagging	The vocabulary is used by end users to do subject indexing of a collection of items (text corpus, images,...)
Classification / Categorization	The vocabulary defines categories that can be assigned to items of a collection (text corpus, images,...)
Faceted Search	Facets describe content from multiple perspectives, by forming a mutually exclusive classification based on the indexed items
Multilingual Search	The vocabulary contains textual descriptions of concepts in multiple languages
Document Suggestion (Recommendation)	Based on the search query, similar documents are included in the search result
Spelling Suggestions and Corrections / Autocompletion	User input (at search and indexing time) is matched with the vocabulary terms and corrections are suggested
Term Suggestions	Based on the structural organization of the vocabulary and the user input, additional terms are suggested
Query expansion and refinement	Based on a controlled vocabulary structure, a user query is broadened or narrowed to adjust search recall
Navigation	Visual guidance for exploring information resources (e.g., websites, collections,...)
Search results grouping / ranking	Vocabulary-supported optimization of the visual representation of search results
Linking (Data Integration)	The controlled vocabulary is created as an intermediate step to provide compatibility with another data source
Publication	The controlled vocabulary is made available "as is" online for reuse by others to view or download

identified the three most important issues for each usage scenario. Furthermore, we computed the arithmetic mean of usage scenario importance over all quality statements and sorted them accordingly.

To find out the level of agreement for each quality statement, we calculated the relative number for each possible choice on the Likert scale (from *strongly agree* to *strongly disagree* and *no answer*) based on the total number of participants who answered the respective question. The rationales provided by the participants were analyzed qualitatively. We compared them to the agreement ratings, and collected those that overlap or contradict in meaning or are otherwise of interest for the findings section (Section 4).

For further studies we provide the anonymized data collected in the survey online[10].

4. FINDINGS

Since our survey focuses on the practical usefulness and implication of our defined quality issues, we first describe their relation to the identified vocabulary usage scenarios. Due to lack of space we focus on 8 issues our participants considered to be most important for the selected 6 usage scenarios. We then present the participants' agreement levels on quality issues, summarize their decision rationales, and discuss the findings we can derive from their answers.

4.1 Usage Scenarios

The closed-ended question on the vocabulary application scenario was answered by 76 respondents and "Classification/Categorization" was mentioned most often (58), followed by "Manual/Intellectual Indexing"(52) and "Faceted Search"(45). Multiple selections were allowed and only 5 participants selected "other" as a usage scenario. "Publication", "Navigation" and "Linking" were mentioned as most important for the provided quality statements. Table 2 lists these usage scenarios in alphabetic order and ranks the quality issues by importance (1 means most important). The quality issues are grouped by the three categories introduced in Section 3.2.

Table 2: Importance of Quality Issues for Usage Scenarios

Usage Scenario	Omitted or Invalid Language Tags	Label Conflicts	Undocumented Concepts	Number of Synonyms and Non-descriptors	Cyclic Hierarchical Relations	Orphan Concepts	Missing Out-Links to Other Vocabularies	Missing Out-Links to Other Resources
Classification / Categorization	2		3		1			
Faceted Search	1	3			2			
Linking	3						1	2
Manual Indexing		3		2	1			
Navigation	1				2	3		
Publication		1	2		3			

4.2 Labeling and Documentation Issues

This group of issues was considered most important for the selected vocabulary usage scenarios (cf. Table 2).

4.2.1 Omitted or Invalid Language Tags

In our previous study we observed that language tags in documentary concept properties (e.g., labels, notes) are either used consistently for all concepts or are omitted completely. This raised the question whether inclusion of language tags is a commonly desired feature in Web vocabularies. To learn about the participants' perception of omitted or invalid language tags, we included the statement *"Textual descriptions of concepts (e.g., labels) should make use of language tags"* in our survey. The majority of the participants (80.4% of the 56 respondents) agreed with that statement, 5.4% disagreed, the rest selected neutral or gave no answer.

Participants who provided a rationale for their decision stated that using language tags is highly useful in multilingual and/or multicultural environments. It supports lan-

[10] http://tinyurl.com/oc24r3o

guage independence and interoperability and enables the vocabulary to be utilized for translation use cases. Usability has also been pointed out as a benefit of making the used languages explicit. However, one contributor states that the user interface should inform the user about a vocabulary's language(s) instead of showing abbreviated codes used for language tags attached to the labels. Another argues that language tags might be superfluous for monolingual vocabularies.

4.2.2 Label Conflicts

In *qSKOS* we also defined a function to detect ambiguous labels (cf. Section 2.2) on a more general level than outlined in the SKOS primer [8]. This definition is expected to provide hints to duplicated concepts or misspelled labels. In our study we could observe that 8 of 15 reviewed vocabularies contain pairs of distinct concepts that have identical descriptors or non-descriptors. Thus, we included the statement *"Different concepts should not be labeled identically (i.e., their descriptors, non-descriptors or synonyms should not overlap)"*. From the total 39 answers approximately 67% of the respondents agreed, 10% disagreed and 23% gave no answer or voted neutral.

Respondents who disagreed with this statement pointed out that identical labels cannot always be avoided, e.g., in case of homographs or when the set of indexing terms must not be changed. One contributor claimed label ambiguity to be beneficial for exploring an information system because it would lead to new search questions. Others perceive unambiguous labels as important for human communication and automated processing, e.g., Natural Language Processing. Confusion (users select incorrect concepts) and decreased manageability have also been mentioned.

4.2.3 Undocumented Concepts

Documentation is often considered beneficial for human users who work with a Web vocabulary. However, documentation can be provided on various levels. Options are, for instance, documenting on the vocabulary level (e.g., content overview or intended usage), documentation of certain groups of concepts, or documenting at the concept level (e.g., scope or history notes, definitions). In our survey we focused on the last case by asking participants to rank their agreement with the statement *"Every concept should be documented (by, e.g., scope notes, definitions, history notes)"*. More than 77% agreed, 9% disagreed, the rest selected neutral.

Contributors who agree with this statement mention that labels alone are often insufficient for disambiguation and understandability. Concept-level documentation provides additional context which has been identified as essential for indexing and tagging usage scenario as well as for establishing mappings between terms and auto-categorization techniques. Three contributors point out the importance of providing history notes for documenting a vocabulary's evolution. Contributors who disagreed argued that not every concept needs documentation and that documentation causes maintenance overhead that could be avoided by providing scope by means of adequate labels and relationships. Also, for some usage scenarios like "large-scale indexing of general-interest content", providing documentation for every concept is perceived as impractical and unnecessary by one contributor.

4.2.4 Number of Synonyms and Non-descriptors

Web vocabularies differ widely in their support and quantity of synonyms and non-descriptors. Web vocabularies like DBpedia categories, for example, define only preferred labels and no alternative or hidden labels. Only 5,450 of over 170,000 concepts in GTAA[11] have alternative labels whereas AGROVOC, provides on average more than 4 alternative labels per concept[12]. To find out if and in what cases synonyms and non-descriptors (lexical variants) are important, we included the statement *"The more synonyms and non-descriptors are defined per concept, the more useful is the controlled vocabulary"* in the survey. More than 60% agreed, 10.5% disagreed and a relatively large number selected neutral (18.4%) or gave no answer (10,5%).

Again, the additional context given by a higher number of synonyms and non-descriptors has been pointed out as beneficial. One contributor stated that more synonyms improve usability whereas more non-descriptors (i.e. lexical variants) have the potential to improve interoperability with other sources. Similarly, it has been noted that synonyms enable more accurate searches and offer more choices in concept selection. Thus, the availability of synonyms and non-descriptors is seen as highly usage-scenario dependent. They may be more useful in text-focused applications but not for Linked Data applications. A rich number of synonyms has furthermore been mentioned as beneficial for manually mapping vocabulary terms. Contributors also stated that the quality of the included synonyms is crucial. They should be unambiguous and fit to the content, i.e., non-required synonyms should be excluded. Furthermore, the quantity may increase complexity and can increase recall and reduce precision. One contributor even argues that adding many synonyms is a waste of time because natural language dictionaries already exist for this task.

4.3 Structural Issues

Our participants rated structural issues as important for five out of the six usage scenarios we focus on.

4.3.1 Cyclic Hierarchical Relations

The negative aspects of cycles in hierarchical relations between concepts have been addressed in numerous tutorials and guidelines on vocabulary development ([3, 6]). Nevertheless, in our previous vocabulary study, we could find cycles in hierarchical relations in 3 out of 15 vocabularies. This led to the question on relevance of cycles for vocabulary quality and to inclusion of the statement *"Controlled vocabularies should not contain circular hierarchical dependencies between concepts"* in our questionnaire. 80% of the 30 respondents to this statement agreed (50% strongly agreed), 10% disagreed, 3.3% voted neutral and 6.7% provided no answer or did not know.

More specifically, we observed two kinds of cyclic relations: those that involve only one concept (reflexive cycles) and those that involve multiple concepts. Thus we included the two statements *"Concepts should not be hierarchically related to themselves"* and *"Controlled vocabularies should not contain circular hierarchical dependencies between concepts"* in our questionnaire.

[11] `http://datahub.io/en/dataset/gemeenschappelijke-thesaurus-audiovisuele-archieven`

[12] Numbers are taken from the dataset of our previous work, available at `https://github.com/cmader/qSKOS-data`

Concerning the first statement, more then 77% of a total 31 respondents agreed (more than 51% even strongly agreed) that concepts should not be hierarchically related to themselves. 6.4% disagreed, 9.7% were neutral and 6.5% gave and no answer or did not know.

Although cycles may not turn out as problems in some scenarios (e.g., if hierarchically related to others and not top concept), reflexive cycles are perceived as unintuitive and increase the complexity of a vocabulary because they do not add value. They may represent a degenerated cycle and contributors stated that they cannot imagine scenarios where reflexive cycles could be a requirement. One contributor stated that cycles can be a sign of "lack of care by the vocabulary publisher". Others point out possible technical problems due to these "loops".

Contributors have argued similarly for cycles involving multiple concepts. They are also perceived to decrease coherence, increase complexity, and are confusing and unintuitive. However, as one participant noted, cycles are only an issue if hierarchical relations are interpreted transitively. Others state that cycles might be caused by misuse of hierarchical relations and suggest the use of other constructs (e.g., alternative labels) to avoid cycles.

4.3.2 Orphan Concepts

Checking for orphan concepts, i.e., concepts that are not linked to other concepts, is a frequently employed quality assurance method. However, we experienced a high number of orphans in several Web vocabularies (e.g., GTAA, LCSH[13], DBpedia categories[14]). Thus we wanted to know how such structures are perceived in general in the Linked Data context and formulated the statement *"Every concept should be linked (e.g., associatively, hierarchically or equivalently) to at least one other concept of the controlled vocabulary"*. Approximately 65% of 37 total respondents agreed to the statement, a small number 2.7% provided no answer. A relatively large number of participants disagreed with the statement (22%) and 11% voted neutral.

From the provided rationale the main concern with orphan concepts was their lack of scope and context which impacts the user's understanding in a negative way. Furthermore they are of "little automated usage" and make it easier to navigate through the vocabulary. However, orphan concepts sometimes cannot be avoided because some usage scenarios do not require relations between concepts (e.g., glossaries). In these cases, unnecessary relations for the purpose of circumventing orphans should not be "invented".

4.4 Linked Data Specific Issues

Although the survey analysis indicates the importance of interlinking Web vocabularies for various usage scenarios, Linked Data specific issues have been considered most important only for the usage scenario *Linking*.

4.4.1 Links to Other Vocabularies

Establishing links to other vocabularies on the Web is a core Linked Data design principle and also suggested in controlled vocabulary development standards and guidelines (e.g., [1, 3]. However, it is currently unclear, how the value of links between online vocabularies of different provenance are perceived from a quality point of view. Thus, we included the statement *"Good-quality vocabularies reference (link) to other vocabularies on the web"* in our questionnaire. More than 64% of total 28 respondents who gave feedback on the statement agreed, 11% disagreed and 21% voted neutral. The rest selected no answer/don't know.

Additional scope and the ability to "share" resources are benefits of linking to other vocabularies on the Web. One participant meant that this is especially important for navigation, browsing and retrieval use cases. Other contributors noted that linking to other vocabularies "Allows better cross resource searching" and that it increases trust and understandability. However, contributors also mentioned that linked vocabularies must also have a high quality standard like, e.g., reasonably established vocabularies (LCSH, AGROVOC). Three contributors argued that vocabularies can be of very good quality on their own and that links to other vocabularies are not an indicator of quality.

4.4.2 Links to Other Resources

Linked Data allows for linking to any other kind of resource on the Web such as web pages that provide additional information about a concept. To find out the impact of such links on vocabulary quality, we included the statement *"Concepts should be linked to other resources on the Web (to, e.g., refer to additional information about the concept"*. More than 78% agreed with the statement, 3.6% disagreed (no participant strongly disagreed) and the rest voted neutral.

The decision rationales are very similar to those discussing linking to third-party vocabularies. Linking to resources on the Web provides additional context, rendering it "useful for end-users and automatic extraction methods" as one contributor stated. Context has also been mentioned important to assist users in choosing an appropriate term. However, link stability has been a concern of three participants. Linked resources should be permanently available and no broken links should be introduced. Those who do not agree mention that vocabularies should be complete on their own and that links to other resources provide additional values but are no substitute for good vocabulary-internal descriptions and definitions.

4.5 Summarized Findings

From the answers and results presented above we can infer the summarized findings listed below. They target the covered quality issues and their relevance according to the usage scenarios can be inferred from Table 2.

- Although not essential in a strictly monolingual context, language tags in RDF literals enhance understandability and usability of the vocabulary.
- It is generally desirable to have all concepts labeled in each supported language. However, this is not always possible due to missing equivalents in some languages.
- Presence of documentation on the concept-level is appreciated but costly and not always needed.
- Whenever possible, identical concept labels have to be avoided to maintain unambiguity and avoid confusion.
- If a vocabulary is intended to organize and contextualize concepts, orphans should generally be avoided.

[13] `http://id.loc.gov/authorities/subjects.html`
[14] downloadable at `http://wiki.dbpedia.org/Downloads38`

- Circular hierarchical dependencies are unintuitive and may indicate or lead to errors.
- When judging the quality of a published and linked Web vocabulary, also the quality of the linked resources has to be taken into account.
- Link stability (changing availability and semantics) is perceived a risk when interconnecting vocabularies on the Web.

5. RECOMMENDATIONS

In this section we combine the findings from the survey with our practical experience in working with controlled vocabularies and implementation of tools such as *qSKOS*. We now provide suggestions and guidelines for quality checking functions in Web vocabulary development tools.

5.1 Labeling and Documentation Issues

When a user creates concept labels or free-text literals, vocabulary development tools should (semi-)automatically add *language tags*. The appropriate tag can be determined from tool parameters (e.g., providing a default language as implemented by Skosify) or by employing existing language detection tools. Vocabulary development tools should also provide language information in a meaningful way on the user-interface level to, for instance, assist users in search term disambiguation. Label suffixes such as "`@de`" could confuse users who are not familiar with RDF-based technologies and should thus be hidden in favor of a clearer language presentation. In cases where identical labels cannot be avoided, we suggest to structure the vocabulary by making use of the SKOS extension for Labels (SKOS-XL) that allows modeling labels as resources instead of literals. As a consequence, additional information such as scope notes for disambiguation or context-specific usage information can be directly attached to labels when needed. When reporting ambiguous labels to the vocabulary creators, conflicts between alternative and preferred labels should be reported with higher priority than conflicts that occur between alternative labels of different concepts. When unique preferred labels cannot be avoided (e.g., in case of homographs), the vocabulary development software should prompt the user to add documentation (e.g., scope notes) for further disambiguation, especially when links between concepts are sparse. Furthermore, by monitoring search queries and user behavior, frequent mistakes can be (semi-) automatically added as hidden labels.

As translation use-cases are a common motivation for developing controlled vocabularies, development software must support creation of labels in multiple languages. If used in a multilingual setting, each concept should be labeled in every relevant language. However, this is often not possible because direct equivalents of concepts in different languages sometimes do not exist. Thus, at least one "default" language should be supported, i.e. one language for which each concepts *must* have a label. Vocabulary development software should tolerate these gaps in language support, but should prompt the user to provide at least a documentation property in the "missing" languages to provide orientation for human users. Also, developers should have the choice to handle similar language tags equally, e.g., concepts labeled in "en-GB" should not require a label for "en-US".

Concepts lacking context are problematic for using and adopting Web vocabularies. The best way to provide context is by establishing relations between concepts and linking to other (external) resources. Thus, vocabulary development software should encourage users to amend labels or documentation to concepts that still lack these interconnections.

Since detecting *conflicting labels* requires domain expertise and human input, vocabulary development and navigation interfaces should reveal a concept's surrounding (e.g., hierarchical) structure to support the user in manual disambiguation. This is also important for supporting resolution of conflicts which can be done by merging or renaming and provide hierarchical or associative links to other concepts. Tools could also apply predefined rules for automatic label rewriting, e.g., by including the broader terms' labels, and helping in resolving conflicts. If a greater number of label conflicts occur in a vocabulary, another possibility is to (automatically) split it into two, separately managed vocabularies with their own, clearly defined scope.

5.2 Structural Issues

As stated by respondents of our survey, *circular hierarchical relations* often root in misuse of hierarchical properties. Users might, for instance, interpret a concept hierarchy either as "has-a" or "is-a" relations. To some degree, tools could suggest replacement of such circularities. Suominen et al. [16] introduce strategies to remove different kinds of hierarchical cycles between concepts. This approach could possibly be extended by providing feedback to the vocabulary developer and suggesting replacement with an associative relation. Taking into account transitivity when checking for cycles is an important operation in this case because computation of cycles in the transitive closure can be omitted if the user perceives hierarchical links as not being transitive.

Orphan concepts decrease cohesiveness of vocabularies and lack context. Vocabulary development tools should suggest semantically related concepts (e.g., inferred from existing popular resources on the Web) that orphan concepts could reference by mapping relations. Tools that automatically identify orphan concepts could also order them by degree of documentation. Orphans without additional documentation properties are more likely to constitute an error or being misinterpreted than those with adequate documentation. Furthermore, context can also be provided by other orphans being members of the same concept schemes or collections.

Whether or not orphan concepts affect the quality of a vocabulary also depends on the vocabulary type and use case. As we observed in the survey results, orphan concepts are more critical for, e.g., navigation usage scenarios and less severe for glossaries. Thus, automatic quality assessment tools could use classification methods (e.g., based on structural properties as suggested in [12]) to infer these types and report orphan concepts only if necessary for the vocabulary type at hand.

5.3 Linked Data Specific Issues

Being able to assess the quality of a vocabulary's linked resources was another desire expressed by our survey participants. Therefore, tools that analyze vocabulary quality should offer the option to run this process also on vocabularies that are (i) *linked by* or (ii) *link to* the main vocabulary. To avoid undesired effects on the semantics of

third-party content, tools should recognize if the developer performs substantial changes to a concept that is linked by these resources. This is manageable in local but clearly more difficult in distributed settings, such as Linked Data. To find in-links in the latter, one has to rely on dataset registries[15] or metadata descriptors like VoID[16]. Given the changing nature of the Web, checks for broken links should be performed automatically on a regular basis and developers should be notified accordingly.

Outgoing links are generally perceived as a method to provide additional scope to concepts, even though they are not strictly necessary for most usage scenarios our participants want to support. Concepts that lack "internal" description and documentation should therefore be reported by vocabulary development tools with a higher priority. To effectively check the quality of linked vocabularies, they should be accessible via an SPARQL endpoint and described in a machine-friendly way, e.g., by a VoID dataset descriptor.

A common concern among our contributors was the increased responsibility when introducing changes to a vocabulary that is linked to others. Providing history notes (rationale of changes) and methods for tracking changes (e.g., keeping multiple versions of concepts and vocabularies) is thus an important feature of vocabulary development tools. Participants of our survey have stated the need for provenance (*who* changed *what* and *when*) of controlled vocabularies. This information should be automatically gathered and attached to the vocabulary. Keeping "historical" data is perceived essential by some of our participants because compatibility with existing systems should be maintained.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we reported the results of a survey we conducted to learn about how curators and users of Web vocabularies perceive quality. We asked the participants to express their opinion and experience on quality issues we identified in our previous work. Our findings clearly reflect the subjective dimension of data quality and point out contradictory approaches and opinions.

However, from the responses we can conclude that existing tools could support taxonomists in producing higher-quality vocabularies by providing semi-automated labeling, documentation, and relationship creation support. Based on the survey results and decision rationales from our participants, we gave recommendations on possible extensions and improvements of quality assessment tools. Thus, these tools could avoid and possibly fix quality problems in Web vocabularies.

Our immediate next step will be to implement these findings in our *qSKOS* vocabulary quality assessment library, which we later would like to integrate with existing vocabulary management platforms such as PoolParty. While our current work focuses on controlled vocabularies expressed in SKOS, we also would like to investigate which quality criteria could be applied for more general, not necessarily SKOS-based data on the Web.

Additionally, we plan to validate our recommendations by implementing them in the *qSKOS* tool and collect feedback from practitioners using the tool in an experimental setup.

[15]e.g., Sindice (`http://sindice.com/`), Data Hub (`http://datahub.io/`)

[16]http://www.w3.org/TR/void/

7. REFERENCES

[1] ISO 25964-2: Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies. Norm, International Organization for Standardization, 2011.

[2] N. Abdul Manaf, S. Bechhofer, and R. Stevens. The current state of SKOS vocabularies on the web. In *The Semantic Web: Research and Applications*, volume 7295 of *Lecture Notes in Computer Science*, pages 270–284. Springer Berlin Heidelberg, 2012.

[3] J. Aitchison, A. Gilchrist, and D. Bawden. *Thesaurus construction and use: a practical manual*. Aslib IMI, 2000.

[4] S. de Coronado, L. W. Wright, G. Fragoso, M. W. Haber, E. A. Hahn-Dantona, F. W. Hartel, S. L. Quan, T. Safran, N. Thomas, and L. Whiteman. The NCI thesaurus quality assurance life cycle. *Journal of Biomed. Inf.*, 42(3):530–539, 2009.

[5] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.

[6] H. Hedden. *The accidental taxonomist*. Information Today, 2010.

[7] A. Hogan, A. Harth, A. Passant, S. Decker, and A. Polleres. Weaving the Pedantic Web. In *Linked Data on the Web Workshop (LDOW2010) at WWW'2010*, 2010.

[8] A. Isaac and E. Summers. SKOS Simple Knowledge Organization System Primer. Working Group Note, W3C, 2009.

[9] D. Kless and S. Milton. Towards quality measures for evaluating thesauri. *Metadata and Semantic Research*, pages 312–319, 2010.

[10] C. Mader, B. Haslhofer, and A. Isaac. Finding quality issues in SKOS vocabularies. In *TPDL 2012*, Germany, May 2012.

[11] A. Miles and S. Bechhofer. SKOS Simple Knowledge Organization System Reference. Recommendation, W3C, 2009.

[12] H. Nagy, T. Pellegrini, and C. Mader. Exploring structural differences in thesauri for SKOS-based applications. I-Semantics '11, pages 187–190. ACM, 2011.

[13] NISO. *ANSI/NISO Z39.19 - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, 2005.

[14] M. Poveda, M. del Carmen Suárez-Figueroa, and A. Gómez-Pérez. Common pitfalls in ontology development. In *CAEPIA*, pages 91–100, 2009.

[15] D. Soergel. Thesauri and ontologies in digital libraries: tutorial. In *Proc. 2nd Joint Conf. on Digital libraries (JCDL)*, 2002.

[16] O. Suominen and E. Hyvönen. Improving the quality of SKOS vocabularies with Skosify. In *Proc. of the 18th int. conf. on Knowledge Engineering and Knowledge Management*, EKAW'12, pages 383–397, Berlin, Heidelberg, 2012. Springer-Verlag.

[17] D. Vrandecic. Ontology evaluation for the web - PhD proposal. In J. Diederich, E. Motta, and E. P. Bontas, editors, *KWEPSY 2006*, Budva, Montenegro, Juni 2006.