Linked Vocabulary Recommendation Tools for Internet of Things: A Survey

127

NIKLAS KOLBE, University of Luxembourg, Luxembourg
SYLVAIN KUBLER, Université de Lorraine, France and CRAN, France
JÉRÉMY ROBERT and YVES LE TRAON, University of Luxembourg, Luxembourg
ARKADY ZASLAVSKY, CSIRO, Australia and Deakin University, Australia

The Semantic Web emerged with the vision of eased integration of heterogeneous, distributed data on the Web. The approach fundamentally relies on the linkage between and reuse of previously published vocabularies to facilitate semantic interoperability. In recent years, the Semantic Web has been perceived as a potential enabling technology to overcome interoperability issues in the Internet of Things (IoT), especially for service discovery and composition. Despite the importance of making vocabulary terms discoverable and selecting the most suitable ones in forthcoming IoT applications, no state-of-the-art survey of tools achieving such recommendation tasks exists to date. This survey covers this gap by specifying an extensive evaluation framework and assessing linked vocabulary recommendation tools. Furthermore, we discuss challenges and opportunities of vocabulary recommendation and related tools in the context of emerging IoT ecosystems. Overall, 40 recommendation tools for linked vocabularies were evaluated, both empirically and experimentally. Some of the key findings include that (i) many tools neglect to thoroughly address both the curation of a vocabulary collection and effective selection mechanisms, (ii) modern information retrieval techniques are underrepresented, and (iii) the reviewed tools that emerged from Semantic Web use cases are not yet sufficiently extended to fit today's IoT projects.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Software and its engineering → Interoperability**; • **Information systems →** *Service discovery and interfaces*;

Additional Key Words and Phrases: Linked vocabularies, ontologies, semantic web, Internet of things, open ecosystems, linked open data

The research leading to this publication is supported by the European Union's Horizon 2020 Research and Innovation programme under grant agreement № 688203 – bIoTope.

Authors' addresses: N. Kolbe, University of Luxembourg, Interdisciplinary Center for Security, Reliability and Trust, 29 Avenue J.F. Kennedy, Luxembourg, Luxembourg, L-1855; email: niklas.kolbe@uni.lu; S. Kubler, Université de Lorraine, CRAN, UMR 7039, Campus Sciences, BP 70239, Vandœuvre-lès-Nancy, France, F-54506, CRAN, UMR 7039, Campus Sciences, BP 70239, Vandœuvre-lès-Nancy, France, F-54506; email: s.kubler@univ-lorraine.fr; J. Robert, University of Luxembourg, Interdisciplinary Center for Security, Reliability and Trust, 29 Avenue J.F. Kennedy, Luxembourg, Luxembourg, L-1855; email: jeremy.robert@uni.lu; Y. Le Traon, University of Luxembourg, Interdisciplinary Center for Security, Reliability and Trust, 29 Avenue J.F. Kennedy, Luxembourg, Luxembourg, L-1855; email: yves.letraon@uni.lu; A. Zaslavsky, CSIRO, Data61, Research Way, Clayton, Australia, VIC 3168, Deakin University, School of Information Technology, Faculty of Science Engineering & Built Environment, 221, Burwood Highway, Burwood, Australia, VIC 3125; email: Arkady.Zaslavsky@data61.csiro.au.

0360-0300/2019/01-ART127 $15.00
https://doi.org/10.1145/3284316

ACM Reference format:
Niklas Kolbe, Sylvain Kubler, Jérémy Robert, Yves Le traon, and Arkady Zaslavsky. 2019. Linked Vocabulary Recommendation Tools for Internet of Things: A Survey. *ACM Comput. Surv.* 51, 6, Article 127 (January 2019), 31 pages.
https://doi.org/10.1145/3284316

1 INTRODUCTION

The Internet of Things (IoT) envisions a world in which *Things* with sensing and actuating functions are connected through the Internet, bringing boundless societal and economic opportunities [164]. Unfortunately, building a single global ecosystem of Things that communicate with each other seamlessly is virtually impossible today, the main reason being that the IoT is essentially a collection of isolated "Intranets of Things" (also referred to as "vertical silos"), where data are siloed in a unique system, cloud, domain, and stay there. This situation imposes significant limitations on the IoT vision, in which *"people and things [are] connected anytime, anyplace, with anything and anyone"* [142]. Current research trends, such as 5G-IoT, indicate further growth and a need for convergence of these heterogeneous data and IoT middleware solutions, as well as IoT data analytics [78].

One part of the interoperability problem in the IoT relates to the semantic layer as there is no unique way of annotating IoT data when publishing it to the Web (also known as *Web of Things* (WoT) [46]). The wide range of employed data modeling approaches as well as available data models hinder the efficient development of disruptive cross-platform and cross-domain applications [106, 155] because it makes it difficult to efficiently (and on demand) discover, access, and integrate heterogeneous IoT data sources. To tackle this issue, increased research efforts investigate the integration of Semantic Web technologies to move toward a truly open and connected IoT ecosystem [11], along with difficult standardization efforts for semantic interoperability in the IoT [31]. Indeed, the Semantic Web [15] provides a machine-understandable knowledge infrastructure on the Web that can be easily integrated into existing software environments [133]. It is inherent to the Semantic Web that vocabularies can be shared, reused, extended, and integrated through the Web. Despite its advantages, the adoption of Semantic Web technologies adds further challenges. For example, modeling data with linked vocabularies is not trivial as the fundamental principle to achieve semantic interoperability between distributed systems is to reuse existing vocabulary terms and establish interconnectivity between them [56, 133]. This requirement led to the need of recommendation tools that help various Semantic Web users (e.g., vocabulary creators, data modelers, Linked Data consumers) to find, select, and apply appropriate vocabularies and terms.

According to one author [133], the reuse of vocabularies is divided into three aspects: discovery, selection, and integration. Furthermore, vocabulary recommendation is performed for a specific purpose or scenario [122]; that is, a recommendation could differ based on the user's intent of usage. Such Semantic Web scenarios and tools include, for example, ontology-based query answering and semantic browsing [122], data mapping and publishing Linked Open Data (LOD) [129], and vocabulary and knowledge engineering [133], as well as semantically annotating IoT data streams [49]. The following example illustrates this issue.

> *Example:* Alice and Bob are both looking for a recommendation about collected *observations*. While Alice would like to publish it as a statistical dataset in the LOD cloud, Bob intends to annotate a data stream generated by a sensor network. One reasonable recommendation for Alice could be to reuse the term <*http://purl.org/linked-data/cube#Observation*> because of its wide adoption in

existing LOD datasets whereas for Bob one reasonable recommendation could be the term <*http://www.w3.org/ns/sosa/Observation*> as it provides a way to further model the sensor setup.

Within this context, this survey aims at reviewing and assessing relevant tools with regard to existing state-of-the-art theories, techniques, and approaches. This evaluation is subsequently used as basis for identifying and discussing challenges of the integration of vocabulary recommendation in IoT ecosystems. Vocabulary recommendation is a composition of several processes, which themselves inherit various challenges. As of the time of writing, one may find related surveys of recommendation tools and related work on architectural design considerations with respect to vocabulary discovery and/or selection [34, 56, 122, 156], as well as integration of semantics in the IoT [132, 146]. However, to the best of our knowledge, no previous work has proposed a joint conceptualization nor an extensive framework to compare existing recommendation tools of various types with similar purpose, nor reviewed the feasibility of such tools for IoT ecosystems. In this article, the term *Vocabulary Recommendation Tool (VRT)* is used as an umbrella term for tools that provide means for the discovery and/or selection of linked vocabularies.

The rest of the article is structured as follows: Section 2 discusses the main concepts, theories, and techniques underlying the Semantic Web, along with its importance (from an interoperability perspective) considering emerging IoT ecosystems. Section 3 presents the evaluation methodology of this survey. The associated vocabulary recommendation evaluation framework is developed in Section 4. The specified framework thereby serves as a basis for comparing existing linked vocabulary recommendation tools in Section 5, which further presents the findings. Section 6 discusses the integration challenges of vocabulary recommendation in today's IoT projects, and Section 7 summarizes the identified research challenges and directions; the conclusion follows. An overview of all acronyms used in this article is given in Appendix A.

2 SEMANTIC WEB AND IOT ECOSYSTEMS

This section aims at introducing in Section 2.1 the main concepts, theories, and techniques underlying the Semantic Web, along with details about the vocabulary recommendation process. Section 2.2 discusses the important role of vocabulary recommendation in the context of open IoT ecosystems. Section 2.3 concludes related semantic challenges and the contribution of this survey.

2.1 Semantic Web: Concept and Terminology

The Semantic Web offers a technology stack that makes it possible to (i) fundamentally represent a web-embedded graph structure (a schema and corresponding instances) with clear referencing to entities through Universal Resource Identifiers (URIs) (i.e., RDF [90]), (ii) define concept taxonomies and relationships (i.e., RDFS [89]), (iii) define logical constraints and rules between concepts, relations, and instances (e.g., based on Description Logics with OWL [4] and SWRL [62]), (iv) reason over the defined models to automatically infer new relations (e.g., with reasoners like Pellet [134]), (v) enrich vocabularies and datasets with metadata, and (vi) use query languages to retrieve information (e.g., SPARQL [111]).

The Semantic Web approach comes with various characteristics that prevail in the way to work with these technologies. First, data modeling is separated from the syntax, meaning that RDF-based models can be serialized in various formats. Second, vocabularies (i.e., classes, relations, constraints, etc.) and data (i.e., instances of classes and properties, including metadata) are represented with the same formalism so that the model and instance level are not clearly separated. Vocabularies themselves are expressed as *Web Data*, and thus Semantic Web tools often do not clearly distinguish these levels. In contrast, the underlying knowledge representation formalisms

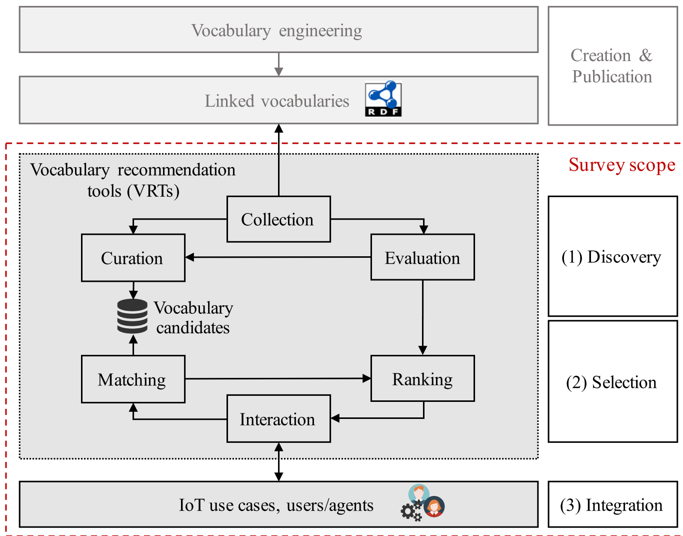


Fig. 1. Dimensions of vocabulary recommendation with regard to discovery and selection.

make a clear distinction between these two, which are referred to as Terminological Box (TBox) for the schema and Assertional Box (ABox) for the data [40]. In the Semantic Web community, the terms *vocabulary*, *ontology*, and *knowledge base* are commonly used. A widely accepted definition of *ontology* is given by Gruber, who defines it as *"an explicit specification of a conceptualization"* [44]. In the Semantic Web context, an ontology corresponds to the schema definition (TBox). The term *vocabulary* is often used interchangeably with the term *ontology*, as argued by the World Wide Web Consortium (W3C) *"there is no clear division between what is referred to as vocabularies and ontologies."*[1] However, it is further noted that the term "ontology" is often used for *"more complex"* schema definitions, whereas the term "vocabulary" does not focus on a *"strict formalism."* It can be perceived that recent popular schemas that could be described as lightweight ontologies (e.g., schema.org[2] and the redesign of the Semantic Sensor Network (SSN) ontology,[3] particularly its new core SOSA [64]) often use the term "vocabulary." Semantic interoperability in the IoT focuses mainly on a common collection of terms to describe relations among concepts, and—as a first step—does not necessarily require complex semantic constraints. Thus, the term *linked vocabulary* or *short vocabulary*, is used in this article to refer to models defined with Semantic Web technologies. The *knowledge base* is commonly referred to as a populated vocabulary/ontology, or instantiations of classes that represent data (i.e., TBox and ABox) [35].

The development of linked vocabularies is part of the vocabulary engineering process (Figure 1) and considered a complex task. Various methods, methodologies, and tools (such as ontology editors) have been proposed to support and guide the engineering process with regard to design considerations and vocabulary evolution aspects to accurately capture the domain of discourse. The vocabulary engineering process is out of the scope of this survey, and the reader is referred to the literature for further reading [42, 82, 140]. VRTs are concerned with finding and choosing the most appropriate published vocabulary/term based on a query. Figure 1 provides a greater insight into the vocabulary recommendation process, which can be divided into two fundamental tasks:

[1] https://www.w3.org/standards/semanticweb/ontology – accessed 09/2018.
[2] http://schema.org/ – accessed 09/2018.
[3] https://www.w3.org/TR/vocab-ssn/–accessed 09/2018.

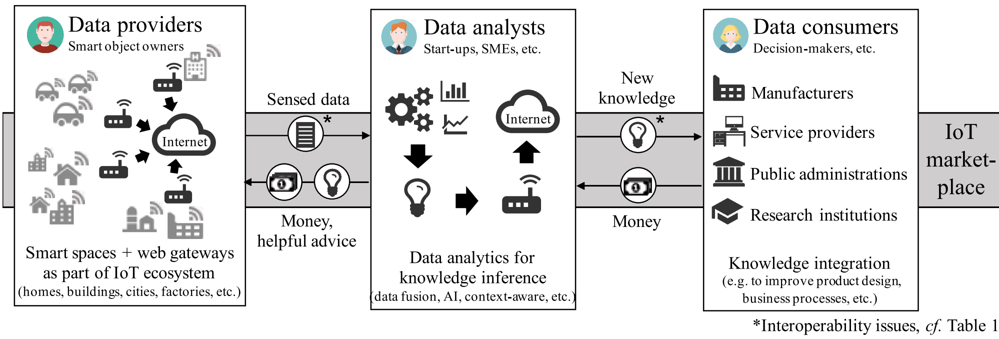


Fig. 2. IoT ecosystem vision (based on [103]).

discovery and selection. Both tasks consist of various steps, which will be referred to as *dimensions* in the following discussion of the evaluation framework. The figure further illustrates how these steps are interconnected. The discovery process is comprised of *collection*, *evaluation*, and *curation*, which are respectively concerned with finding/gathering existing vocabularies on the Web, assessing their quality, and maintaining the repository of suitable candidates. The selection process, on the other hand, requires *interaction*, *query matching*, and *ranking*, which are respectively concerned with providing intuitive interfaces for users/agents, finding a match of suitable candidates in the repository based on a query, and ranking these candidates for vocabulary/term recommendation purposes. The dimensions of vocabulary recommendation are discussed in detail in Section 4 to identify key features and specify an evaluation framework for related tools. The last step, integration of the vocabulary recommendation in IoT use cases (Figure 1), is the subject in Section 6.

2.2 Toward Emerging IoT Ecosystems: IoT Data Trading

Several organizations and standardization fora have started to build up consortia and initiatives with the aim of creating IoT ecosystems that are fundamentally based on openness [137], including identification, discovery, and interoperation of services across platforms [115]; these include the Alliance for Internet of Things Innovation (AIOTI) launched by the EU [1], the Open Platform 3.0™ at The Open Group, the OneM2M global standards initiative [145], the IEEE Internet of Things (IoT) initiative [93], and the International Technical Working Group on IoT-Enabled Smart City Framework developed at NIST [117].

The IoT ecosystem vision that is followed in these projects aims for the breakdown of vertical silos and achievement of horizontal integration [2] and the emergence of open innovation ecosystems with co-creation capabilities [63], as well as the creation of a new value chain through establishing an environment for data trading, as depicted in Figure 2. Three ecosystem stakeholders are illustrated, including end-users who own smart objects (e.g., a smart fridge); data analysts (startups, SMEs, etc.) who may be interested in accessing smart object-related data to deliver new services that fulfill untapped needs of data consumers, whether end-user needs (e.g., offering a new service that propose recipes with food items that are going to exceed the best-before date) and/or business needs (e.g., generating some knowledge such as usage patterns, failure prediction, etc., which could benefit the fridge manufacturer to improve the fridge design). As emphasized in Figure 2, various types of incentives between these stakeholders can be imagined that could be supported by a digital marketplace acting as an IoT search engine and thus enabling multimodal registration, discovery, and trading of data and services (see e.g., [18, 103, 119]). One key challenge to realize this vision is to enable interoperability between the IoT data published from heterogeneous sources and the data consumed by analysts. Achieving such interoperability is not

Table 1. Interoperability Issues Based on LCIM [149]

Interoperability level		Description
6	Conceptual	Refers to a fully specified model to be shared among all stakeholders.
5	Dynamic	Refers to means to track the evolution of and the ability to discover services.
4	Pragmatic	Refers to description of the service to access relevant data; e.g., RESTful, WSDL, Swagger.
3	Semantic	Refers to understanding of the data model, the meaning of terms, relations, language, etc.
2	Syntactic	Refers to agreement about the data format; e.g., XML, JSON(-LD), CSV.
1	Technical	Refers to OSI-layers 1-6.

only a technical matter [48], as summarized in Table 1, from a conceptual perspective. This survey addresses the semantic interoperability problem, which is a prerequisite for the upper levels of the interoperability scheme and thereby also for IoT ecosystem building blocks such as IoT marketplace-like components.

2.3 Semantic Challenges and Survey Contribution

This state-of-the-art survey is motivated by the presented IoT ecosystem vision in which VRTs could provide the building blocks essential to converge to semantic interoperability. Several vocabularies have been proposed to model data with respect to IoT aspects like sensor setups, observations, actuators, services, and the like which have previously been reviewed in the literature [30, 73, 131, 157]. However, when it comes to modeling the information that smart objects (or *Things*) provide, domain-specific ontologies are required to annotate the data [131], which emphasizes the need for appropriate recommendation tools. However, existing IoT platforms often rely on a predefined data model that the published data must comply with in order to be incorporated into the platform. These data models have inherited characteristics (e.g., different formats, units, languages) that make them incompatible with one another. From an IoT ecosystem perspective, no single data model should be imposed at the data provider level. Indeed, it is neither feasible nor manageable to create a single data model/ontology that describes all aspects of the IoT and related domains (i.e., one that would satisfy all stakeholders) [84]; moreover, it is not possible to develop a single approach to semantically annotate sensor data for gateways [104]. Nonetheless, semantic annotations are a requirement to discover and integrate available IoT data in intelligent and autonomous systems (e.g., for WoT search engines [150]). This is a key motivational aspect that convinced us to survey and evaluate existing VRTs for the IoT. Overall, recommendation is meant to guide providers and consumers in finding and reusing the *most suitable* vocabularies/terms for their specific intent and circumstance.

Furthermore, the semantic-oriented vision of the IoT [7], with its related challenges and the benefits of linked vocabularies in the IoT, goes beyond the interoperability issue [79, 158]. Ontologies have been excessively used in IoT settings for intelligent systems such as context- and situation-awareness approaches [72, 105], activity recognition [68], and other analytics (Figure 2). In these systems, linked vocabularies are often combined with logical constraints and rules to apply semantic reasoning. VRTs can also support the development of such applications since the recommendation can similarly be applied to select vocabularies for knowledge specification of the respective domain. Moreover, VRTs could not only support the discovery of related IoT data streams, but further ease the integration of data in application knowledge bases. Despite the advantages of linked vocabularies in terms of interoperability, Semantic Web reasoning techniques are often associated with performance issues for IoT platforms [125], which brings new challenges in embedding linked vocabularies in more efficient data analytics approaches (e.g., RDF stream processing [160]).

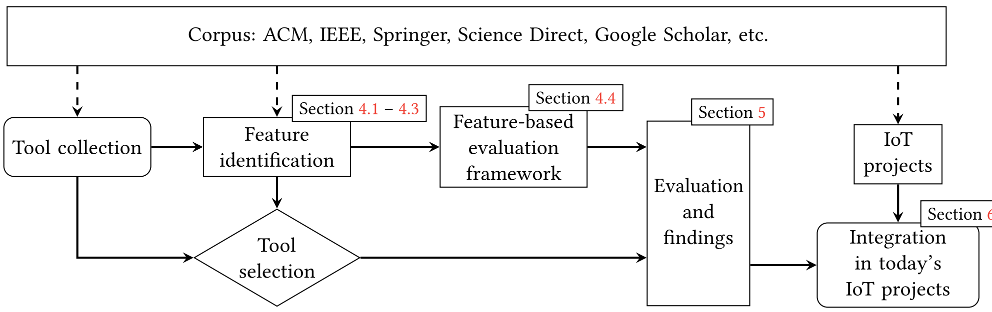


Fig. 3. Evaluation methodology of this survey.

The contribution of this article is to thoroughly analyze existing VRTs and assess whether these tools are appropriate to the constraints of the aforementioned IoT ecosystem vision. The evaluation framework developed in this survey can guide the development of new VRTs, which, for example, are able to recommend best-suited ontologies for IoT use cases. It should be noted that the evaluation of existing IoT vocabularies and identifying best-suited vocabularies for IoT domains is out of the scope of this survey.

3 EVALUATION METHODOLOGY

As a typical purpose for systematic reviews [69], this survey aims to compare existing approaches in terms of advantages and disadvantages, provide a joint conceptualization of the various approaches in the field, and identify open challenges. The methodology followed in this survey is illustrated in Figure 3.

First, tools related to vocabulary recommendation have been collected from well-known digital libraries and search engines. An analysis over this set exposed different dimensions that inherit various challenges for these tools. Subsequently, based on the dimensions of vocabulary recommendation presented in Section 2.1, key features have been identified in an exhaustive manner through corpus refinement, as detailed in Sections 4.1–4.3. These features represent the unified aggregation of relevant tools and serve as criteria of the evaluation framework specification, as presented in Section 4.4. The comparison study, findings, and consideration of IoT ecosystem aspects are discussed in Section 5. Last, the integration of vocabulary recommendation in IoT projects and the feasibility of surveyed VRTs for IoT scenarios are discussed in Section 6.

Relevant tools and respective publications for the comparison in Section 5 were selected following the PRISMA methodology [94]. To be included in the evaluation analysis, the recommendation tool must satisfy the requirement that it both/either propose a discovery mechanism and/or a selection mechanism for linked vocabularies. Studies presented in doctoral dissertations, master's theses, textbooks, and non–peer-reviewed papers were ignored. Further, the following (closely related) tool types were excluded:

- Expert vocabulary collections with no selection mechanism being offered (e.g., LOV4IoT [49], Protege Online Library,[4] vocab.org,[5] ontologi.es,[6] joinup,[7] and SWEET ontologies [116]);

[4]Protege Ontology Library: https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library#OWL_ontologies – accessed 09/2018.
[5]vocab.org: http://purl.org/vocab/ – accessed 09/2018.
[6]ontologi.es: http://ontologi.es – accessed 09/2018.
[7]joinup core vocabularies: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/core-vocabularies – accessed 09/2018.

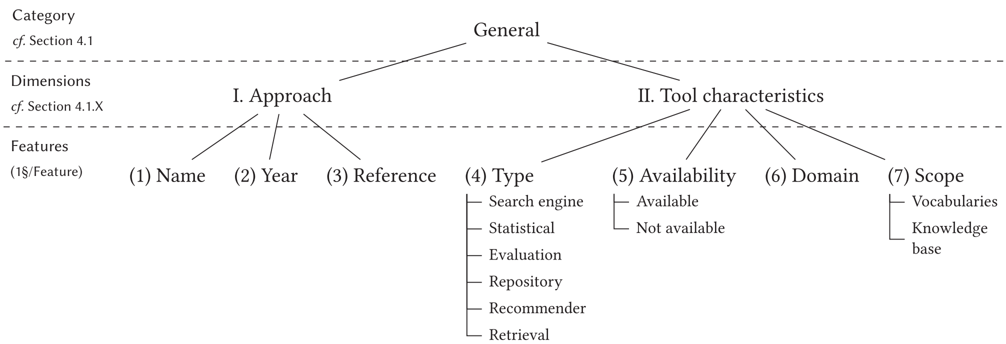


Fig. 4. General dimensions and features to characterize VRTs.

- Tools that solely focus on the validation of a single vocabulary which, however, could support evaluation in VRTs (e.g., OntoCheck [130] and Oops! [109]);
- Analytical tools that could also provide valuable inputs for VRTs like metadata extraction (e.g., Aether [83]);
- Tools computing plain schema-related statistics of a single vocabulary (e.g., RDFStats [76]).

Eventually, 40 tools with 45 associated studies were selected, published in the following scientific libraries: Springer (~22.2%), IOS Press (~13.3%), CEUR (~13.3%), Science Direct (~8.9%), ACM (~6.7%), IEEE (~6.7%), AAAI (~4.4%), IGI Global (~4.4%), IADIS (~4.4%), Wiley (~4.4%), and others (~11%).

4 EVALUATION FRAMEWORK SPECIFICATION FOR VRTS

This section provides a more in-depth discussion of vocabulary recommendation by identifying key features for each dimension. To help the reader follow the features that are introduced and discussed with regard to each category (i.e., for general features in Section 4.1, discovery features in Section 4.2, and selection features in Section 4.3), an at-a-glance overview in the form of a tree graph (e.g., Figure 4) is given in each of the following sections. The enumeration of dimensions/features shown in the graph is kept throughout the survey. Section 4.4 presents the resulting evaluation framework.

4.1 General Features

Before going into detail on vocabulary discovery and selection, general dimensions and features are defined to characterize VRTs. Features were associated with two general dimensions—namely, *approach* and *tool characteristics*—as summarized in Figure 4.

4.1.1 Approach. This first dimension is introduced to present the approaches, considering the following features:

Name: The name of the tool as used in its publication, to uniquely identify the approach.

Year: Year of the first associated publication that serves for identification of trends in proposed tools over time.

Reference: Scientific reference of the study used as basis for the evaluation.

4.1.2 Tool Characteristics. Second, more detailed characteristics of the tools are considered. These general features include:

Type: Different types of vocabulary libraries have been identified [34]. However, as a broader scope of recommendation tools is considered, an novel classification scheme was used that is based on the dimension(s) a tool puts particular emphasis on. Six different types of VRTs were identified:

- *Search engine:* Focus on vocabulary collection (e.g., discovery and indexing of semantic documents through Web crawling).
- *Statistical tool:* Focus on vocabulary collection and evaluation (e.g., analyzing the usage of vocabularies, often extracted from LOD sources, to guide end-users in choosing an appropriate vocabulary/term).
- *Evaluation tool:* Focus on vocabulary evaluation [e.g., assessing quality of the discovered vocabularies to give recommendations (considering a given set of metrics)].
- *Repository:* Focus on vocabulary curation (e.g., the provision of a platform for a community to manually collect and review vocabularies based on predefined requirements).
- *Recommender:* Focus on vocabulary ranking (e.g., by applying information filtering techniques or learning over LOD datasets to recommend most suitable vocabularies/terms).
- *Retrieval tool:* Focus on vocabulary interaction and matching (e.g., by proposing advanced means for querying, exploring, and matching candidates on a query from an existing set of vocabularies).

Availability: Whether the tool is *available* or *not available,*[8] which also indicates whether it could be evaluated experimentally. It is determined by checking whether an active website or download of the tool could be found by following URLs in the publication(s) and via a web search with the tool's name.

Domain: Covered domains of the vocabulary collection (if not independent). This is concluded from the vocabularies that are maintained in the tool's repository.

Scope: Indicates whether the approach focuses on *vocabularies* or further supports *knowledge bases* (since schemas and data in the Semantic Web are based on the same formalism). The scope is inferred by checking whether the tool's repository exclusively contains vocabularies.

4.2 Discovery Features

The discovery of vocabularies is a fundamental process of VRTs, as only discovered vocabularies can be in the set of potential candidates to be recommended upon a query. Three dimensions with regard to discovery are discussed—namely, *collection*, *evaluation*, and *curation*—as summarized in Figure 5.

4.2.1 Collection. The first step of the recommendation process is concerned with the collection process of available vocabularies. Two distinct features and associated approaches were identified through the tools' evaluation and based on the discussion on ontology collections in d'Aquin and Noy [34], namely:

Crawling: In this process, the Web is browsed systematically by a software system realized through Semantic Web crawling, reusing common Web search engines, processing LOD sources and endpoints, or accessing APIs of existing vocabulary collections and VRTs. These approaches rely on a fundamental best practice that states that vocabularies should be hosted and made publicly accessible at the URI of the vocabulary.

Semantic Web crawlers, also referred to as RDF crawlers, harvest data from Semantic Web documents (SWD) to discover linked vocabularies or data. These crawlers focus on extracting RDF-based data that can be found in various formats (e.g., RDF/XML, turtle, JSON-LD) or embedded in other documents (e.g., RDFa in HTML). Existing VRTs also exploit conventional *web search engines* and associated crawlers (Google, etc.) in order to discover semantic web documents on the web

[8] Availability as of September 2018.

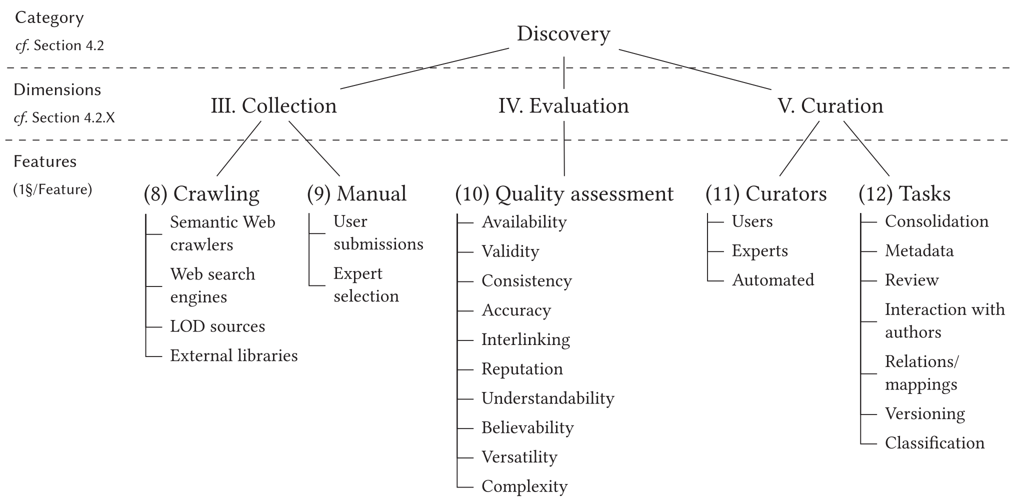


Fig. 5. Dimensions and features related to vocabulary discovery.

(e.g., by filtering specific document types such as *filetype:rdf* and *filetype:owl*). Deploying an RDF crawler faces various design and implementation challenges, as discussed in Harth et al. [55]. As an alternative way to browsing the whole web, accessing *LOD endpoints* or data dumps to extract used vocabulary terms is another way employed for collection. However, this approach is only capable of discovering vocabularies that have been used to model data in the analyzed linked dataset. Last, some approaches crawl vocabularies from the APIs of existing *external libraries*.

Manual: In contrast to automatic approaches, those that rely on a manual collection process do not aim to discover all available vocabularies on the web but rather seek to fulfill one of the following goals: (i) present a proposed vocabulary to the community, (ii) keep supervised control over the maintained candidate set, or (iii) provide a platform for community consensus. Manual collection can be achieved either through *user submissions* or *expert selection*. Submission-based approaches are more flexible, and facilitate the evolution of the vocabulary collection. Expert collections are often maintained by an official body.

4.2.2 Evaluation. The second dimension of vocabulary discovery is concerned with the assessment of the quality and correctness of vocabularies [60]. In the vocabulary recommendation process, the purpose of evaluating vocabularies are twofold: (i) assuring a certain quality for the selected vocabulary candidates and (ii) giving the best recommendation for selection [122, 148]. Thus, as illustrated in Figure 1, evaluation serves as an input for *curation* as well as *ranking*, which are respectively discussed in Sections 4.2.3 and 4.3.3. In this survey, the focus is on evaluation aspects relevant for vocabulary recommendation rather than on ontology evaluation for the vocabulary engineering process, which has already been the subject of studies in the literature [17, 41, 60, 139].

In this respect, the Semantic Web community has proposed various best practices and guidelines for vocabulary design, development, publication, and reuse. These documents often cover both schema- and data-related aspects while providing a source for identifying quality criteria for vocabularies. One may cite, among other examples, the 5-star Linked Data model [14], consumer/publisher recommendations [61], Linked Data design considerations [57], five-star rating for vocabulary use [65], OntoClean methodology [45], guidelines for Linked Data generation and publication [114], ontology pitfalls [110], W3C best practices recipes [16], or still the best practices

Table 2. Evaluation Criteria of Linked Vocabularies and Their Consideration in VRT Processes

Criteria	Synonyms	Description	Implementation	Used as criteria for: Curation	Ranking
Availability [14, 143, 165]	Dereferencability [57, 61, 65], Accessibility [61]	Whether the vocabulary is accessable and dereferencable through its URI.	HTTP requests, openness of vocabulary license.	✓	✓
Validity [112]	Syntactic accuracy/correctness [14, 143, 165], Syntax evaluation [61, 139]	Whether the vocabulary is syntactically correct.	Parsing the vocabulary.	✓	✓
Consistency [37, 41, 60, 61, 165]	Machine-readable [65]	Whether the vocabulary is free of logical contradictions with regard to its underlying representation (RDFS, OWL-variant, etc.).	Applying reasoners.	✓	✓
Accuracy [60, 165]	Domain cohesion [60, 139], Veracity [143] (Re-)usability [37]	Whether the schema correctly represents a real-world domain.	Human judgment.	✓	✓
Interlinking [14, 57, 65, 165]	Connectedness [19], Coupling [60], Structural evaluation [139]	The extent to which the vocabulary includes sufficient semantic relations to external vocabularies.	Counting in- and out-links at the schema level.	✓	✓
Popularity [122]	Usage statistics [139]	The extent to which the vocabulary/term is often used to model data of the domain it describes.	Analyzing LOD datasets for instantiations of the vocabulary, counting its presence in ontology repositories, or taking into account the number of local requests.	✗	✓
Reputation [165]	-	Whether users judge the vocabulary to be of integrity.	User reviews and ratings.	✓	✓
Understandability [57, 165]	Practical quality [143], Interpretability [165], Clarity [60], Metadata [65, 139]	Whether the vocabulary can be understood without ambiguity; e.g., through annotation properties like rdfs:label and rdfs:comment.	Counting annotation properties.	✓	✓
Believability [165]	Provenance metadata [57]	Whether the provenance / metadata about the vocabulary indicates that it comes from credible source.	Checking author information and history.	✓	✓
Versatility [165]	-	Whether the vocabulary is available in different languages and serialization formats.	Checking labels with language property (@en etc.).	✓	✗
Richness [37, 122]	Complexity [143], Density [3], Informativeness [6]	The extent to which concepts in the vocabulary are described and specified.	Measure based on number of properties, siblings, subclasses, and superclasses per concept.	✗	✓
Centrality [22]	Betweenness [3]	The extent to which a concept is central in the vocabulary graph.	Measure based on amount of relations of a concept and/or the count of shortest paths within the vocabulary that go through it.	✗	✓
Importance [36]	-	A combination of popularity and interlinking, meaning that the importance depends on the quality of the link.	Measures of interlinking while taking into account the popularity of the source for in-links, e.g., PageRank algorithm.	✗	✓

applied to IoT [50]. Furthermore, quality assessment of vocabularies has been extensively studied in the literature [37].

Quality assessment: The quality attributes considered in the evaluation framework are listed in Table 2. The selection of criteria is mainly based on the comprehensive review presented in Zaveri

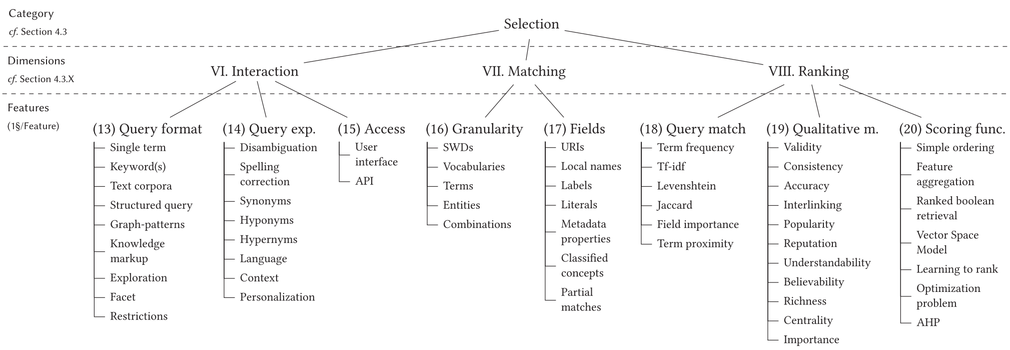


Fig. 6. Dimensions and features related to vocabulary selection.

et al. [165] and was complemented with those stemming from best practices and those considered by the VRTs that are the subjects of the evaluation. From the aforementioned sources, only quality attributes that are concerned with the schema (TBox) and deemed relevant for vocabulary recommendation were selected. In addition, we assess whether the quality criteria is typically considered for curation and/or ranking.

4.2.3 Curation. The last identified dimension of the discovery process is vocabulary curation, which refers to the management and maintenance of vocabulary candidates from the internal repository. Indeed, curation is often a collaborative effort to ensure and improve the quality of formalized knowledge [43]. Two features were identified based on the survey in Groza et al. [43] and the reviewed tools, namely:

Curators: This feature indicates who oversees the curation and maintenance of the vocabulary candidate collection, which could be fulfilled by *users*, *experts* or through *automated processes*. Whereas human-based curation offers means to improve vocabularies based on reviews and discussions, automated curation is able to handle large sets of discovered vocabularies more efficiently.

Tasks: The curation process can cover different aspects, including *metadata* completion and maintenance, *review* of newly discovered vocabularies, the addition of *semantic relations* to other vocabularies of the corpus, the *consolidation* of discovered vocabularies, and support of *versioning*, as well as *classification*. It should be noted that the curation process strongly influences the quality and extent of the vocabulary candidate repository, which could potentially enhance vocabulary selection (e.g., the classification of vocabularies can be exploited to filter domains).

4.3 Selection Features

The second fundamental task of vocabulary recommendation is to select the most appropriate candidate from the repository. The dimensions with regard to selection are *interaction*, *matching*, and *ranking*, as summarized in Figure 6 and discussed in this section.

4.3.1 Interaction. Recommendation approaches need to provide means to interact with users and agents to query the recommendation service. This dimension is broken down to the following features:

Query format: A VRT interface could offer the following means/formats for querying, as identified from evaluated tools (in particular from Harth [53]) and the discussions in Henzinger [58]: *single terms*, *keyword*-based search, *text corpora* (so-called free text retrieval), *logical/structured query* (e.g., SPARQL), *graph-patterns*, and tasks expressed with *knowledge markup*. Whereas keyword-based search is the most popular and easy to use interface, it is often argued that it does not allow for a precise specification of the information need [58]. Search- and recommendation-based approaches often require multiple interactions for users to reach their goal and thus further employ interaction models that include *exploration* of the vocabulary collection (e.g., object focus, path traversal [53]), specification of *restrictions*, and offering of *facets*.

Query expansion: The idea of query expansion is to express the user's information need more precisely to enhance retrieval results. Available approaches, based on the comprehensive survey in Carpineto and Romano [26], include *disambiguation* options (e.g., defining the sense of a keyword in case it has multiple meanings [47, 162]), *spelling checks* on the user's input, and expanding keywords with *synonyms*, *hyponyms,* or *hypernyms* (e.g., through WordNet [91]). Some approaches aim to *personalize* the query and retrieve more suitable results for a specific user. In semantic search, such user preferences could represent a user's interest in a certain concept. Instead of explicitly defining preferences, another approach takes into account the *contextual parameters* of a request to improve the recommendation.

Access: VRTs are often designed to provide a search oriented toward humans and thus often offer a *user interface* (UI) to access information. In more detail, yet beyond the scope of this survey, UIs are concerned with result presentation and visualization (list, graphs, trees, etc.). However, the need for Semantic Web applications to discover vocabularies led to the need for RESTful APIs to access the service, which further allows the aggregation of VRTs.

4.3.2 Retrieval. The second dimension of vocabulary selection is concerned with:

Granularity: The importance of granularity for vocabulary recommendation is discussed elsewhere [118, 122] and served as motivation to collect granularity levels that were considered by the evaluated tools. Retrieval in VRTs can be done on the level of matching *SWDs*, *vocabularies*, *terms*, *entities*, or *combinations* from different vocabularies. The granularity strongly impacts the nature of the recommendation process: Whereas some approaches aim at recommending complete vocabularies that have the best coverage of the queried domain/concepts (also through recommending combinations of vocabularies), other approaches seek to find a single best term or entity. Indeed, it is not trivial to decide whether it is best to reuse as few vocabularies as possible for a user's intended task or to combine "better" terms from various vocabularies [127].

Matching fields: The matching process is concerned with retrieving candidates from the repository that match the query. This feature shows the detail to which tools match a query against the information contained in a vocabulary. Due to the inherent structure of RDF vocabularies, matches can be performed on different fields and properties including *URIs*, *local names*, *labels*, *literals*, and *metadata properties* such as basic properties like *rdfs:comment* or properties from vocabularies such as the Dublin Core[9] and SKOS.[10] Moreover, in case vocabularies are classified during the curation process, vocabularies of a *matching domain/concept* can be retrieved. In the case of logical queries, the match is returned from processing the query based on its underlying language. Furthermore, *partial matches* could also be taken into account, even though one may dispute their real impact on the search quality [66]. The matching fields considered by an approach are relevant for vocabulary selection since there is no one way or guarantee that vocabularies are annotated in the same way. Valuable information could reside in different fields/properties that could help to identify a matching candidate.

[9]Dublin Core vocabulary: http://purl.org/dc/terms/ – accessed 09/2018.
[10]SKOS vocabulary: https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html – accessed 09/2018.

4.3.3 Ranking. Algorithms to rank matched candidates form a key component of VRTs. Ranking aims at determining best candidates from the vocabularies that matched the query by taking into account various measures, including:

Query match measures: These enable the ranking of the candidate set determined by the query match through content- and graph-based similarity measures. Due to the huge amount of similarity measures in the literature, the evaluation framework is limited to those found in the evaluated tools. However, those were aligned to the literature [21, 85]. A common approach is to compute the *term frequency* in a vocabulary for all words in a query, which is often combined with the inverse document frequency of the terms (otherwise, rare terms would have no power to influence the query relevancy), also known as *term frequency–inverse document frequency (tf–idf)* [123]. Another way to calculate similarity between query and candidate at the string level is to simply compute the edit distance (*Levenshtein*). The *Jaccard coefficient* allows to measure the overlap between sets and thus can be applied on the set of words from the query and those from the vocabulary. Further, finding a match in some field types of a vocabulary might be of higher importance than others (e.g., a match in the name is more valuable than in the metadata), which can be represented through assigning different weights to field types (i.e., *field importance* or weighted zone ranking [85]). The last query match measure found is concerned with the *proximity* of multiple query term matches within the vocabulary graph. This measure is calculated by identifying the shortest path between matched fields [85].

Query-independent measures: Qualitative measures: These aim to compute a score for a vocabulary or term independent from the query. The approach to the collection of qualitative features is discussed in Section 4.2.2, and those quality criteria relevant for ranking are indicated in Table 2. One scoring algorithm standing out is PageRank [98] to measure the importance of a document, a popular one for ranking Web pages that can be adapted to the needs for SWDs. PageRank falls into link analysis, being based on a random surfer who, starting from one page/vocabulary, randomly follows a link. The more often a node is visited by the random surfer during his walk, the more important it is. For SWDs, the random surfer needs to consider the semantics of the followed links.

Scoring functions: Last, previously presented measures are used as inputs of the scoring function to compute an overall ranking of matching candidates for a query. Due to the large amount of approaches to achieve a ranking, only those found in the evaluated tools are listed in the evaluation framework. Further reading includes general scoring functions [85] and ranking of vocabularies [27, 141].

Ranking with only one feature requires *simple ordering*. In a straightforward manner, multiple features can be *aggregated* (e.g., through weighted or unweighted sums or factorization). In the *vector space model* [124], documents are represented as vectors with each component representing a document term, which could be computed based on the tf-idf-like measures. Vector space scoring calculates the similarity between two documents (e.g., between a query vector and a vocabulary vector) by calculating the cosine similarity, which, however, is expensive to compute [85]. Weights of ranking features can be defined through experts or learned from a training dataset (*learning to rank* [80]), with algorithms such as LambdaMART [159]. Calculating a score for candidates could also be seen as an *optimization problem* by formulating features as cost functions. Last, features could also be aggregated through the *analytical hierarchy process (AHP)* [121].

4.4 Evaluation Framework

The overall evaluation framework consists of the different sets of features with regard to the different vocabulary recommendation dimensions, which have all been summarized in Table 3. This framework is then used as a basis for evaluating and comparing various recommendation tools.

Table 3. Evaluation Framework

Category	Dimension	Feature	Description
General	I. Approach	(1) Name	Tool's name.
		(2) Year	Year of first relevant publication.
		(3) Reference	Reference to the tool's academic publication.
	II. Tool characteristics	(4) Type	The category the tool falls into: Search Engine (■), Statistical (■), Evaluation (■), Repository (■), Recommender (■), Retrieval (■).
		(5) Availability	Whether the tool is available (✓), not available (✗).
		(6) Domain	Domains covered, if not domain-independent.
		(7) Scope	Whether tool focuses on Vocabularies (Voc), or on Knowledge Bases (KB).
Discovery	III. Collection	(8) Crawling	Automatic collection of vocabularies through: Semantic Web crawler (Cr), Web search engines (SE), LOD Endpoints (LOD), and/or External Libraries (Lib).
		(9) Manual	Manual collection through: User Submission (U), and/or Expert Collection (E).
	IV. Evaluation	(10) Quality assessment	Assessment of discovered vocabularies for curation purposes, based on: Availability (Ava), Validity (Val), Consistency (Con), Accuracy (Acc), Interlinking (Int), Reputation (Rep), Understandability (Und), Believability (Bel), Versatility (Ver), or Richness (Rich).
	V. Curation	(11) Curators	Curation handled Automatically (A), by Experts (E), or Peers/Users (U).
		(12) Tasks	Curation tasks covered: Consolidation (Con), Metadata (Met), Content Review (Rev), Interaction with authors (Int), defining Relations and Mappings (Rel), maintain Versions (Ver), add Classifications (Clas).
Selection	VI. Interaction	(13) Query format	Ways to query the recommendation service, including: Single Term (Term), Keywords (Key), Text Corpora (TC), Structured (QL), Graph-pattern (Gra), Knowledge Markup (KM), Exploration (Exp), Restrictions (Res), Facets (Fac).
		(14) Query expansion	Means to improve the query formulated by the user: Disambiguation (Dis), Spelling Correction (Spe), Synonyms (Syn), Hyponyms (Hypo), Hypernyms (Hyper), Language (Tra), Context (Con), or Personalization (Per).
		(15) Access	Whether information access is provided for Users (UI), and/or Agents (API).
	VII. Matching	(16) Granularity	To which granularity a query is matched and retrieved from the corpus: Vocabulary (Voc), Terms (Term), Entities (Ent), SWDs (SWD), or support of Combinations of these (Comb).
		(17) Fields	To which fields of a vocabulary a query is matched: URIs (URI), Local names (Nam), Literals (Lit), Labels (Lab), Metadata Properties (Met), Classified Concepts (Con), Partial Matches (Par).
	VIII. Ranking	(18) Query match measure	Assessment of the similarity between query and vocabularies in the corpus, based on: Term Frequency (TF), TF-IDF (TF-IDF), Levensthein (Lev), Jaccard (Jac), taking into account the Field Importance (FI), and/or the Term Proximity (TP) of matches in a vocabulary.
		(19) Qualitative measures	Assessment of vocabularies in the corpus to calculate a quality score based on: Validity (Val), Consistency (Con), Accuracy (Acc), Interlinking (Int), Popularity (Pop), Reputation (Rep), Understandability (Und), Believability (Bel), Richness (Rich), Centrality (Cen), or Importance (Imp).
		(20) Scoring function	Approach to calculate a final rank based on the used measures: Simple Ordering (Ord), Feature Aggregation (Agg), Ranked Boolean Retrieval (RBR), Vector Space Model (VSM), Learning to Rank (L2R), Optimization Problem (Opt), Analytical Hierarchy Process (AHP).

5 EVALUATION OF VRTS

This section presents the evaluation of existing VRTs based on the proposed framework. Tables 4 and 5 present the results with regard to the general/discovery and selection dimensions, respectively. In the following sections, the findings of the evaluation are discussed.

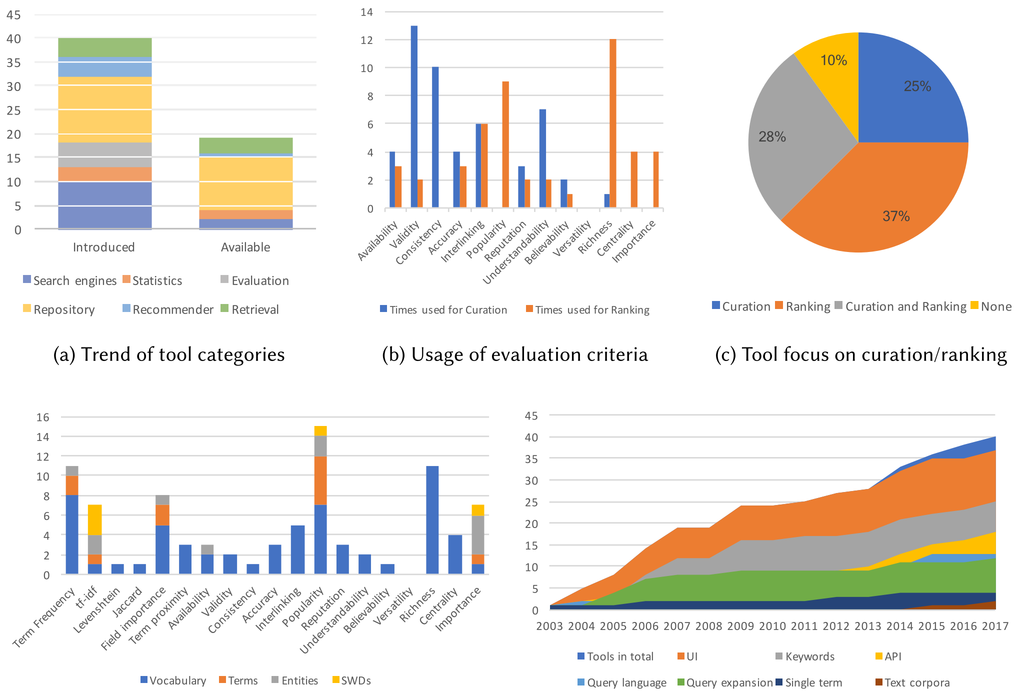
To help readers extract quick and meaningful information, some results of the evaluation (i.e., from Tables 4 and 5) have been displayed in the form of charts in Figure 7, including:

Table 4. Evaluation of VRTs Regarding General and Discovery Features

I. Approach	II. Tool characteristics				III. Collection			IV. Evaluation		V. Curation	
(1) Name	(2) Year	(3) Reference	(4) Type	(5) Availability	(6) Domain	(7) Scope	(8) Crawled	(9) Manual	(10) Quality assessment	(11) Curators	(12) Tasks
Ontokhoj	2003	[101]	🟥	✗	-	Voc	Cr	-	-	A	Clas
OntoSelect	2004	[19]	🟥	✗	-	Voc	Cr	U	-	-	-
Swoogle	2004	[36]	🟥	✓	-	Voc, KB	Cr, SE	U	Val, Int	A	Met, Rel, Ver
Ontosearch	2005	[166]	🟥	✗	-	Voc, KB	SE	-	-	-	-
SWSE + ReConRank	2007	[54]	🟥	✗	-	Voc, KB	Cr	-	-	A	Con, Rel
Sindice	2007	[151]	🟥	✗	-	Voc, KB	Cr	U	Val, Con	-	-
Watson	2007	[32]	🟥	✓	-	Voc, KB	Cr, SE, Lib *Swoogle*	-	Val, Con, Rep	A	Con, Rel, Met, Clas
Falcons Concept & Entity Search	2009	[28, 113]	🟥	✗	-	Voc, KB	Cr	-	Val	A	Clas
VisiNav	2009	[52]	🟥	✗	-	Voc, KB	Cr, LOD	-	-	-	-
WebOWL	2012	[12]	🟥	✗	-	Voc, KB	Cr	-	Val, Con	-	-
LODstats	2012	[8]	🟧	✓	-	Voc, KB	LOD	-	-	A	Met
vocab.cc	2013	[136]	🟧	✓	-	Voc	LOD	-	-	-	-
OUSAF	2015	[5]	🟧	✗	-	Voc	*Watson, Sindice*	-	Val, Con, Int, Und	-	
Supekar et al.	2004	[143]	🟩	✗	-	Voc	*Onthokoj*	-	-	A	Clas
OntoMetric	2004	[81]	🟩	✗	-	Voc	U	-	-	-	-
Ontology Auditor	2005	[20]	🟩	✗	-	Voc	Lib	-	Val, Con, Acc		-
OntoQA	2005	[147]	🟩	✗	-	Voc, KB	*Swoogle*	-	Rich, Int, Und	-	-
Knowledge Zone + TS-ORS	2006	[77, 144]	🟩	✗	Biomed.	Voc	-	U	Acc, Und, Rep, Bel	U	Clas, Ver, Rev, Met
Open Metadata Registry	2006	[59]	🟦	✓	-	Voc	-	U	-	U	Ver, Met
Ontosearch2	2006	[100]	🟦	✗	-	Voc	-	U	Val, Con	-	-
Oyster	2006	[99]	🟦	✓	-	Voc	-	U	-	U	Met
OBO Foundry	2007	[135]	🟦	✓	Biomed.	Voc	-	U	Ava, Int, Und	U	Met, Ver, Clas, Rev
BioPortal	2009	[97]	🟦	✓	Biomed.	Voc	-	U	Int, Acc, Und	U	Rev, Met, Ver, Rel
Cupboard	2009	[33]	🟦	✗	-	Voc	-	U	*TS-ORS*	U	*TS-ORS, Oyster*, Rel
MMI	2009	[120]	🟦	✓	Marine	Voc	-	U	Val, Con	U	Met, Rel, Ver
Ontobee	2011	[161]	🟦	✓	Biomed.	Voc	*OBO Foundry*	E	-	-	-
BiOSS	2010	[87]	🟦	✗	Biomed.	Voc	-	E	-	-	-
Manchester OWL Repository	2014	[88]	🟦	✓	-	Voc	Cr, SE, Lib	U	Ava, Val, Con	A	Con
smartcity.linkeddata.es	2014	[108]	🟦	✓	IoT	Voc	-	E, U	Ava	E	Int, Met, Rev
LOV	2014	[154]	🟦	✓	-	Voc	-	U	Ava, Val, Und, Int, Bel	E, A	Met, Rev, Ver, Int
Ontology Lookup Service	2015	[67]	🟦	✓	Biomed.	Voc	-	U	Val, Con	E, A	Ver
Ontohub	2017	[29]	🟦	✓	-	Voc	-	U	Val, Con	U	Rev, Ver, Met, Rel
(Web)CORE	2006	[25, 38]	🟪	✗	-	Voc	Lib	U	Acc, Und, Rep	U	Rev, Clas
DWRank	2014	[22, 23]	🟪	✗	-	Voc	Lib	-	-	-	-
TermPicker	2016	[128]	🟪	✗	-	Voc	LOD	-	-		-
NCBO 2.0	2017	[86]	🟪	✓	Biomed.	Voc	*BioPortal*	-	-	-	-
AKTiveRank	2006	[3]	🟨	✗	-	Voc	*Swoogle*	-	-	-	-
(combi)SQORE	2007	[152, 153]	🟨	✗	-	Voc	*Watson*	-	-	-	-
LOVR	2015	[138]	🟨	✓	-	Voc, KB	*vocab.cc, LOV*		-		-
RecoOn	2016	[24]	🟨	✓	-	Voc	Lib	-	-	-	-

Table 5. Evaluation of VRTs Regarding Selection Features

I. Approach	**VI. Interaction**			**VII. Matching**			**VIII. Ranking**	
(1) Name	(13) Query format	(14) Query expansion	(15) Access	(16) Granularity	(17) Fields	(18) Query match measure	(19) Qualitative measure	(20) Scoring function
Ontokhoj	Term, QL	Con, Dis, Syn, Hyper	UI, API	Voc	Nam	-	Imp	Ord
OntoSelect	KM, Exp	-	UI	Voc	Lab, Par	TF	Int, Rich	Agg
Swoogle	Key, QL, Fac	-	UI, API	SWD, Term	Lab	TF-IDF	Imp	Ord
Ontosearch	Key	Per	UI	SWD	Con	TF-IDF	-	VSM
SWSE + ReConRank	Key, Fac	-	UI, (API)	Ent	Nam, Lab, Met, Lit	TF-IDF	Imp	Ord
Sindice	Key	-	UI	Ent	URI, Nam, Lab	TF-IDF	Ava	Agg
Watson	Key, Exp, QL	-	UI, API	Voc, Ent, Comb	Nam, Lab, Met, Lit, Par, Con	-	Rich	-
Falcons Concept & Entity Search	Key	-	UI	SWD, Voc, Ent	Nam, Lab, Lit	TF-IDF	Pop	VSM
VisiNav	Key, Fac, Exp	-	UI	Ent	Nam, URI, Lab, Lit	-	Imp	Ord
WebOWL	QL	-	UI, API	Ent	-	-	Imp	
LODstats	Term, Exp, QL	-	UI	Voc, Term, Ent	Nam, URI,	-	Pop	-
vocab.cc	Key	-	UI, API	Term	Nam, URI, Lab	-	Pop	Ord
OUSAF	QL	-	UI	Term	-	-	Pop, Rich	Agg
Supekar et al.	-	-	*Ontokhoj*	Voc	Con	-	Ava, Val, Acc, Rich	Agg
OntoMetric	-	-	UI	Voc	-		Ava, Acc, Und, Rich	AHP
Ontology Auditor	Exp	Syn, Hyper, Hypo	UI, API	Voc	Con	-	Val, Und, Rich	Agg
OntoQA	Key	Syn	UI	Voc	Nam	-	Int, Pop, Rich, Cen	Agg
							Cen	
Knowledge Zone + TS-ORS	Key, Exp	Per, Hypo	UI	Voc	Nam, Met, Con	-	Acc, Rich, Rep, Bel	Agg
Open Metadata Registry	Key, Exp, QL	-	UI	Voc, Term	Lab, Con, Met	-	-	Ord
Ontosearch2	Key, QL, Res	Syn	UI	Term, Ent	URI, Nam, Lab	TF, FI	-	Agg
Oyster	Term	-	UI	Voc	Nam	-	-	-
OBO Foundry	Exp	-	UI	Voc	-	-	-	-
BioPortal	Key, Exp, Fac	Syn	UI, API	Voc, Term	Nam, Lit	-	-	-
Cupboard	*Watson*	-	UI	Voc	*Watson*	*TS-ORS*	-	*TS-ORS*
MMI	Key, Fac, Exp, QL	-	UI, API	Voc, Term	-	-	-	Ord
Ontobee	Key, Exp, QL	-	UI	Voc	Lab, Par	-	-	Ord
BiOSS	Key	Dis, Syn	UI, API	Voc, Comb	Nam, Con, Lab, Par	TF	Rich, Pop	Agg
Manchester OWL Repository	QL, Exp	-	UI, API	Voc	-	-	-	-
smartcity.linkeddata.es	Term, Exp	-	UI	Voc	Con	-	-	Ord
LOV	Key, QL	-	UI, API	Voc, Term	Nam, Lab, Met	TF-IDF, FI	Pop	Agg
Ontology Lookup Service	Key, Exp, QL	-	UI, API	Voc, Ent	URI, Nam, Lab, Par	-	-	-
Ontohub	Key, Exp	-	UI, API	Voc	URI, Nam, Met	-	-	-
(Web)CORE	Key	Per, Dis	UI	Voc	Nam, Par	TF, FI, Lev	Rep	VSM
DWRank	Key	Syn	-	Voc	Nam, Met, Par, Con	TF, FICen	Int, Cen	L2R
TermPicker	Gra	-	UI	Term	Nam	-	Pop	L2R
NCBO 2.0	Key, TC	Syn	UI, API	Voc, Comb	Met	TF, FI	Pop, Rich, Cen	Agg
AKTiveRank	Key	-	UI	Voc	URI, Lab, Par	TF, FI, TP	Int, Cen	Agg
(combi)SQORE	Key, Res	Syn, Hyper, Hypo	UI, API	Voc, Comb	Nam	TF, TP	Rich	Agg
LOVR	TC	-	UI, API	Term	*LOV, vocab.cc*	*LOV, vocab.cc*	*LODStats*	Agg
RecoOn	Key	-	API	Voc	Nam, Par	Jac, TP	Int, Pop, Rich	Opt



(a) Trend of tool categories (b) Usage of evaluation criteria (c) Tool focus on curation/ranking

(d) Ranking features per selection granularity (e) Trend of interaction features

Fig. 7. Analysis of the VRTs evaluation.

- Figure 7(a): Number of tools of each type that have been introduced and are still available;
- Figure 7(b): How often evaluation criteria were used, and whether they are for quality assessment or ranking;
- Figure 7(c): Whether tools focus on quality assessment, ranking, or both;
- Figure 7(d): To show which ranking features were used for each matching granularity;
- Figure 7(e): Trends of how often interaction features are used.

Shift from semantic search engines and evaluation to repositories, recommenders, and retrieval: One trend that can be observed is the shift from the development of search engines and evaluation-focused tools to recommender and retrieval systems. As depicted in Figure 7(a), only a few search engines and evaluation tools included in the survey are still available (about 50% of the reviewed VRTs). Even though a similar observation can be made about recommenders, it should be noted that three out of four tools from this category have been introduced only in recent years, indicating a growing interest. Semantic search engines are often not solely focused on vocabularies but also on Linked Data; however, conventional Web search engines like Google increasingly incorporate the capabilities of retrieving Semantic Web content [96]—a well-known example is the schema.org vocabulary embedded in websites, which is supported by many conventional Web search engines. Evaluation tools are able to thoroughly assess vocabularies; however, they are often inefficient in finding suitable candidates from a vocabulary reuse standpoint. Indeed, with the huge amount of vocabularies published on the Semantic Web, the challenge lies not in discovering as many as possible, but rather in selecting efficiently as few as possible well-fitting and requirements-meeting vocabularies/terms.

Curation vs. ranking: A fundamental aspect of recommendation is the assessment of vocabulary quality. Figure 7(b) shows how often the evaluation criteria are used for curation and/or ranking. Most curation approaches focus on ensuring validity (13 times), consistency (10), and understandability (7) of newly collected vocabularies, whereas ranking models rather take into account the richness (12), popularity (9), and interlinking degree (6). Considering Figure 7(c), it can be added that VRTs often focus on either the curation of the vocabulary collection (25%) or efficient ranking of queries (37%). However, the combination of both, which is implemented by 28% of the reviewed VRTs, would naturally increase the quality of the recommendation service. An example thoroughly considering both approaches is the Linked Open Vocabularies (LOV) platform.

Limited support of combined recommendations: As previously mentioned, it is not trivial to decide whether a recommendation should be made on a vocabulary or term/entity level. When publishing IoT data, rarely will a single vocabulary cover all required terms. The most common selection granularity of the surveyed tools is a complete vocabulary. Identifying the combination of most suitable and interlinked terms for an existing nonlinked schema cannot be easily achieved with existing VRTs because engineers are still required to pick and combine suitable terms. Combined recommendations are especially useful when the recommendation service can take into account all the terms (or other data structures) that the user/agent is looking for. In this evaluation, only two approaches offer recommendations based on text corpora, namely NCBO 2.0 and LOVR, whereas the latter takes HTML as input with the goal of semantically annotating websites.

Impact of qualitative measures on ranking quality remains unclear: Despite the various evaluations for recommendations presented in the selected studies, the general importance of qualitative measures to achieve a better quality of ranking remains unclear as most approaches focus on a limited set of criteria, and different metrics are used for same criteria. The selection approach by Semantic Web users is often driven by the popularity of a vocabulary [127]. The evaluation reveals that popularity is also among the most used criteria for selection in the surveyed VRTs (Figure 7(d)). However, to achieve good results, these query-independent features need to be combined at least with a reliable query match measure [21]. In general, establishing the correct weight between features to optimize the ranking model is a tedious task [80], and there is no common agreement on the importance of each ranking feature for ranking models. Only two approaches of the survey use learning-based approaches to assign weights to features; namely, TermPicker, which focuses on different metrics related to popularity, and DWRank, which focuses on learning the weights for features like centrality and importance. Furthermore, the aggregation of features is also dependent on selection granularity. Figure 7(d) shows the features used for ranking per selection granularity. It can be observed that a large variety of features is only considered (and suitable) when ranking complete vocabularies/SWDs, whereas only a limited number of features is used for ranking terms/entities.

Simplicity for interaction: The trend of some interaction features is displayed in Figure 7(e). It can be observed that simple interfaces are more popular. Most approaches are keyword-based and increasingly offer APIs. On the other hand, the use of query expansion features to refine queries for users and alternative query formats (e.g., query languages or text corpora) are less often implemented by the surveyed VRTs.

6 VOCABULARY RECOMMENDATION IN TODAY'S IOT

We propose the conceptual integration of vocabulary recommendation in IoT ecosystems in Section 6.1, while the consideration of VRTs in today's IoT platforms is analyzed in Section 6.2.

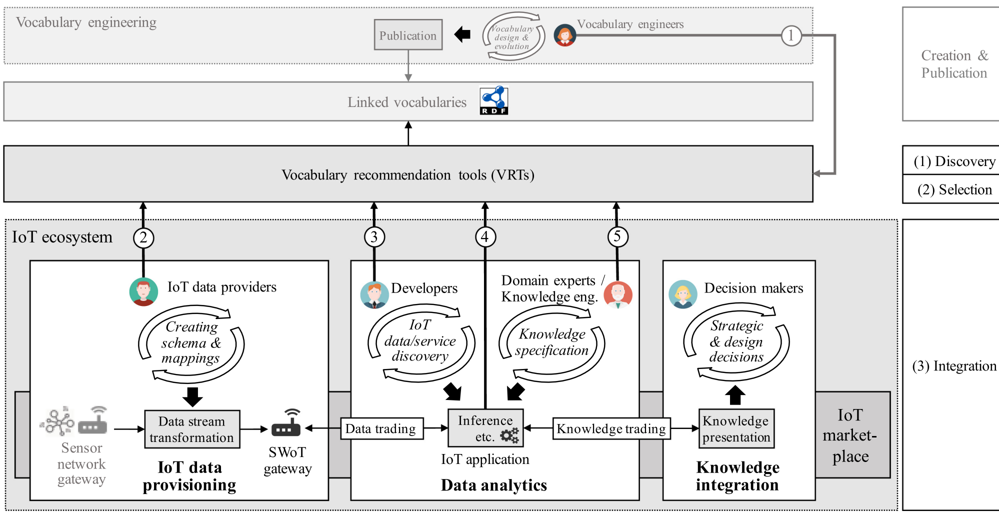


Fig. 8. Integration of vocabulary recommendation in the IoT ecosystem model.

6.1 Conceptual Integration of VRTs in IoT Ecosystems

As mentioned, the process of recommending linked vocabularies can be structured in three steps: (i) discovery of available vocabularies, (ii) selection of the most suitable candidates for user queries, and (iii) integration of the recommendation for the user's task at hand. Figure 8 illustrates the *integration* of VRTs in the scope of an IoT ecosystem model that is dependent on user's intent.

Five distinct use cases (denoted by ① – ⑤ in Figure 8) provide insight into different users/agents who query vocabulary recommendation for different purposes. Case ① shows the most fundamental use case from the Semantic Web in which vocabulary engineers use recommendation tools during the vocabulary development process to link to and potentially extend already existing definitions during the development process instead of redefining them. Cases ② – ⑤ show integration cases in the IoT ecosystem model. Case ② shows the case of smart object owners who use the recommendation not only to define a semantic schema, but also to create mappings from local sensor data to the newly defined schema. With these mapping rules, sensor data streams can be transformed and published with semantic annotations, which is a core requirement to efficiently join IoT ecosystems (e.g., to be easily and efficiently discovered). Cases ③ and ⑤ represent queries from developers and domain experts who respectively intend to discover available IoT data/services and specify knowledge for an intelligent IoT application. Some of these processes could also be automated through artificial agents requesting vocabulary recommendation (④). Eventually the vocabulary recommendation fosters interoperability and allows for more efficient and elaborate knowledge extraction.

6.2 Consideration of VRTs in IoT Platforms

As thoroughly reviewed in Mineraud et al. [92], existing open IoT platforms lack unified and interoperable data models. Ongoing analysis, discussions, and IoT project efforts with regard to linked vocabularies for the IoT domain indicate that this issue is thoroughly addressed by the community, as evidenced by other studies [9, 39]. Despite the fact that increasingly more IoT platforms support SWTs and that vocabularies to describe *Things* are becoming more mature, the sole use of linked vocabularies is not enough to achieve global interoperability [11]. As a first step, this only makes those platforms interoperable that either use the same vocabularies or vocabularies that have been successfully mapped/matched. Recent efforts in this regard include, for example, the Fiesta IoT project that achieved semantic interoperability between the FIWARE and OneM2M platforms (both using different data models/vocabularies) [75]. Earlier, the SPITFIRE project was concerned with aligning IoT vocabularies [107].

VRTs form a key building block to support users of semantic-aware IoT platforms for all the IoT ecosystem use cases that were introduced in Section 6.1 (i.e., linked sensor data publication/transforming sensor data streams, discovering IoT data/services, defining domain knowledge). However, existing IoT platforms—despite the support of SWTs—do not yet follow the IoT ecosystem model as presented in Section 2.2, and do not consider vocabulary recommendation in their scope of tools and platforms. One reason that could explain this is that vocabularies for IoT-related domains (mobility, city, home, etc.) have not yet reached full maturity, with many still being under specification and development (e.g., the MobiVoc[11] vocabulary for the mobility domain). However, the expectation is that developers can easily extend platforms to their needs, integrating data from and modeling data in a format that is understood by various platforms [92]. VRTs, in their essence, support this goal through collecting and offering the means for selecting appropriate vocabularies. The recent and promising Industry Ontologies Foundry (IOF) initiative[12] to some extent agrees with this vision and aims to adapt the success story of the OBO Foundry from the medical domain to the industrial domain (including IoT) in order to provide a collaborative tool suite that helps to build and collect jointly interoperable vocabularies. Nonetheless, the idea of sharing and reusing data models defined by the community has already found its way to the IoT; an example is the information model repository (based on a domain-specific language) of the Eclipse Vorto tool.[13]

Despite the lack of consideration of VRTs in IoT platforms, they have been considered in other SWT-based tools. For example, Schaible [126] describes the integration of TermPicker in Karma [70], which is a linked data integration tool based on mapping rules. Still, such tools do not satisfy IoT specific requirements (e.g., applying the transformation on data streams, while considering specific characteristics of sensor data streams [10], and publishing the data in a SWoT gateway). On the other hand, IoT-specific tools often do not consider vocabulary recommendation. Instead, they are built on a preselected set of vocabularies, like the tools/approaches presented elsewhere [51, 74, 95, 102]. In a recent work [71], a tool to generate a SWoT gateway based on term-level recommendations from the LOV platform has been proposed in the framework of the bIoTope H2020 project.[14]

In an open IoT ecosystem in which data are not modeled to suit a single IoT platform but instead based on common, community-based vocabularies that could be understood by many platforms, VRTs are essential. The VRTs surveyed in this article could be used for this purpose; however, the variety of tools and that the recommendations differ based on the chosen tool (due to different collection and selection capabilities) can be frustrating for users. Further, the success of ontology usage in the biomedical domain indicates that domain-dependent VRTs have a higher chance to be used and adopted by the community and to achieve a consensus. A VRT specialized on IoT-related domains could provide unique collection and selection features (e.g., taking into account the number of IoT platforms and their capabilities that comply with a certain vocabulary).

[11]MobiVoc: http://schema.mobivoc.org/ – accessed 09/2018.
[12]IOF: https://www.youtube.com/watch?v=y0TeTfoFdSA – accessed 09/2018.
[13]Eclipse Vorto: http://www.eclipse.org/vorto/ – accessed 09/2018.
[14]bIoTope: http://www.biotope-project.eu/ – accessed 09/2018.

7 RESEARCH CHALLENGES AND DIRECTIONS

This section summarizes the identified research challenges and directions derived from the evaluation and discussions of this survey with regard to both vocabulary discovery (Section 7.1) and vocabulary selection (Section 7.2).

7.1 Vocabulary Discovery for IoT

The following discussion on challenges for vocabulary discovery is based on the dimensions introduced in the previous sections (Figure 5); namely, collection, evaluation, and curation.

Collection: The collection of vocabularies for IoT domains is a challenging task because most vocabularies are still being proposed in the scope of research projects. LOV4IoT [49], for example, is dedicated to classifying proposed vocabularies and making them accessible by integrating them into the LOV platform [154]. One challenge for vocabulary collection of IoT domains is the restriction to certain domains as sensors are deployed in an increasing number of settings (e.g., in cities and factories) and thus in new contexts that require modeling. However, projects such as LOV4IoT indicate that vocabulary collections can be eventually maintained domain-independently. Future efforts to collect vocabularies for IoT domains will help recommendation tools to build a better vocabulary candidate set.

Evaluation: The evaluation of IoT vocabularies as such can rely on the quality criteria identified in the survey. However, since many vocabularies are still being proposed for the same domains and due to the rapid pace of developments in the IoT, vocabularies are being continuously improved. Hence, more emphasis can be put on the evolution of vocabularies (i.e., focusing on vocabularies that are being actively maintained and extended), while, on the other hand, neglecting those out-of-date. The most critical qualitative evaluation of a vocabulary—its accuracy—requires human judgment. Future evaluation tools, designed as collaborative platforms, will help to achieve a community consensus about proposed IoT vocabularies.

Curation: The amount of vocabularies available, on the one hand, and the complex task of reviewing vocabularies, on the other hand, call for semi-automated curation processes. A particular challenge for the IoT is to keep track of developments, classify and collect the metadata of proposed vocabularies that can be of interest to users, and provide valuable information for matching and ranking vocabularies. Despite its importance, a trend was identified in which recent tools instead focus only on curation or on ranking of vocabularies. The combination of both and the provisioning of curated data to ranking models will benefit future recommendation tools. Moreover, the importance of matching existing, well-known vocabularies to achieve interoperability has been highlighted in the survey. A collaborative curation platform could support the vocabulary matching process and could serve to document the achieved matches, which can be taken into account when recommending a vocabulary based on a query.

7.2 Vocabulary Selection for IoT

The subsequent discussion on challenges for vocabulary selection is based on the dimensions introduced in the previous sections (Figure 6); namely, interaction, matching, and ranking.

Interaction: One challenge is the requirement for more expressive ways to formulate the information needs of the different IoT ecosystem stakeholders. This could, for example, correspond to query formats based on outputs from IoT gateways with proprietary data models (i.e., text corpora such as JSON, XML, etc.) and the specification of the intended use, such as data stream annotation, knowledge specification for a context-aware system, and IoT data discovery. Such improvements of a tool's interfaces will foster the adoption of vocabulary recommendation by users and developers in IoT settings.

Matching: As highlighted previously, it is not a trivial task to decide whether a recommendation should be made on a vocabulary term/entity level or based on combinations from different vocabularies. This may not only depend on the interaction mechanisms provided but also on the intentions of the user. The development of more sophisticated matches with different levels of granularity and the consideration of the user's intent, such as its IoT use cases, will help to optimize the recommendation task.

Ranking: The survey revealed that the popularity of vocabularies/terms is a desirable feature and is often used for vocabulary recommendation. However, in the surveyed tools, this feature is computed only by analyzing LOD datasets, which do not represent semantically annotated IoT data. This may result in miscalculated qualitative scores for IoT vocabularies and does not provide an objectively appropriate ranking. One challenge is thus to define a popularity measure that is suitable for IoT vocabularies. Possible directions include the employment of modern information retrieval techniques, such as analyzing the user click behavior of existing VRTs (that contain IoT vocabularies and are used by IoT stakeholders) to calculate the popularity of vocabularies and terms. Last, existing VRTs do not consider more complex features for advanced users of linked vocabularies, such as the reasoning complexity of a vocabulary [13, 163]. Understanding how the vocabulary recommendation influences reasoning capabilities and constraints in IoT applications is not trivial and opens new challenges. Specialized ranking models for IoT use cases (e.g., through additional information generated during the curation process) will significantly improve the overall recommendation and foster further convergence to most suitable vocabularies for IoT use cases.

8 CONCLUSION

In this survey, the process of vocabulary recommendation was thoroughly reviewed and placed into the context of IoT ecosystems. VRTs help to guide stakeholders of IoT ecosystems when publishing, discovering, and integrating IoT data and services from heterogeneous sources. A comprehensive evaluation framework was defined based on dimensions regarding the discovery (i.e., collection, evaluation, and curation) and selection (i.e., interaction, matching, and ranking) of appropriate vocabularies. This framework served to evaluate 40 vocabulary recommendation tools from the literature, and trends/findings with regard to the identified features were highlighted. Moreover, the conceptual integration of vocabulary recommendation in IoT ecosystem use cases and the consideration of VRTs in today's landscape of IoT platforms were discussed.

In conclusion, two dimensions of vocabularies recommendation are important: curating a vocabulary collection and providing simple, yet efficient selection mechanisms. The survey revealed that tools often focus on either one and that implemented strategies for both differ greatly. It is not completely clear, however, how different features impact the overall recommendation quality. Even though early advances in sharing and reusing data models defined by the community for the IoT could be evidenced, today's scope of IoT platforms does not yet consider VRTs. Whereas VRTs have been integrated into tools that support traditional Semantic Web use cases, only a few tools supporting use cases of IoT ecosystems with vocabulary recommendation could be found.

The presented framework is limited to functional requirements that impact the output of vocabulary recommendation. Nonfunctional requirements (e.g., performance, reliability, scalability) impose additional challenges (e.g., efficient indexing of vocabularies) on the implementation of VRTs which have not been considered in the scope of this survey.

The evaluation presented in this survey can support Semantic Web developers and IoT researchers in getting an overview of the state-of-the-art in vocabulary recommendation and helping to choose the most appropriate tool. Furthermore, the presented evaluation framework can be used to compare newly proposed approaches to improve vocabulary recommendation with previous work. In our vision, a tool that serves as a platform to share, extend, curate, and recommend the vocabularies of IoT-related domains could serve as a fundamental building block for the convergence to interoperable IoT ecosystems.

APPENDIX

A ACRONYMS

An overview of all acronyms used in this article is given in Table 6.

Table 6. Acronym Table

Acronym	Description	Acronym	Description
ABox	Assertional Box	SSN	Semantic Sensor Network
API	Application Programming Interface	SWD	Semantic Web document
IoT	Internet of Things	SWoT	Semantic Web of Things
JSON	JavaScript Object Notation	SWRL	Semantic Web Rule Language
JSON-LD	JavaScript Object Notation for Linked Data	SWT	Semantic Web Technology
KB	Knowledge base	TBox	Terminological Box
LOD	Linked Open Data	UI	User Interface
OWL	Web Ontology Language	URI	Unified Resource Identifier
RDF	Resource Description Format	VRT	Vocabulary Recommendation Tool
RDFS	RDF Schema	WoT	Web of Things
REST	Representational State Transfer	WSDL	Web Service Description Language
SMEs	Small and Medium-sized Enterprises	W3C	World Wide Web Consortium
SPARQL	SPARQL Protocol and RDF Query Language	XML	Extensible Markup Language

ACKNOWLEDGMENTS

The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions.

REFERENCES

[1] AIOTI. 2015. Alliance for Internet of Things Innovation (AIOTI), European Commission. (2015).

[2] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. 2015. Internet of Things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials* 17, 4 (2015), 2347–2376.

[3] Harith Alani, Christopher Brewster, and Nigel Shadbolt. 2006. Ranking ontologies with AKTiveRank. In *Proceedings of the International Semantic Web Conference*. Springer, 1–15. https://doi.org/10.1007/11926078_1

[4] Grigoris Antoniou and Frank Van Harmelen. 2004. Web ontology language: Owl. In *Handbook on Ontologies*, Staab Steffen and Studer Rudi (Eds.). Springer, 67–92. https://doi.org/10.1007/978-3-540-24750-0_4

[5] Jamshaid Ashraf, Omar Khadeer Hussain, and Farookh Khadeer Hussain. 2015. Making sense from big RDF data: OUSAF for measuring ontology usage. *Software: Practice and Experience* 45, 8 (2015), 1051–1071.

[6] Ghislain Auguste Atemezing and Raphaël Troncy. 2014. Information content based ranking metric for linked open vocabularies. In *Proceedings of the 10th International Conference on Semantic Systems*. ACM, 53–56. https://doi.org/10.1145/2660517.2660533

[7] Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. The Internet of Things: A survey. *Computer Networks* 54, 15 (2010), 2787–2805.

[8] Sören Auer, Jan Demter, Michael Martin, and Jens Lehmann. 2012. LODStats–an extensible framework for high-performance dataset analytics. In *Proceedings of the International Conference on Knowledge Engineering and Knowledge Management*. Springer, 353–362. https://doi.org/10.1007/978-3-642-33876-2_31

[9] Garvita Bajaj, Rachit Agarwal, Pushpendra Singh, Nikolaos Georgantas, and Valerie Issarny. 2017. A study of existing ontologies in the IoT-domain. (2017). https://arxiv.org/abs/1707.00112.

[10] Payam Barnaghi, Wei Wang, Lijun Dong, and Chonggang Wang. 2013. A linked-data model for semantic sensor streams. In *Proceedings of the IEEE International Conference on Internet of Things (iThings/CPSCom), IEEE Green Computing and Communications (GreenCom), IEEE Cyber, Physical, and Social Computing*. IEEE, 468–475. https://doi.org/10.1109/GreenCom-iThings-CPSCom.2013.95

[11] Payam Barnaghi, Wei Wang, Cory Henson, and Kerry Taylor. 2012. Semantics for the Internet of Things: Early progress and back to the future. *International Journal on Semantic Web and Information Systems (IJSWIS)* 8, 1 (2012), 1–21.

[12] Alexandros Batzios and Pericles A Mitkas. 2012. WebOWL: A semantic web search engine development experiment. *Expert Systems with Applications* 39, 5 (2012), 5052–5060.

[13] Hamid R Bazoobandi, Jacopo Urbani, Frank van Harmelen, and Henri Bal. 2017. An empirical study on how the distribution of ontologies affects reasoning on the web. In *Proceedings of the International Semantic Web Conference*. Springer, 69–86. https://doi.org/10.1007/978-3-319-68288-4_5

[14] Tim Berners-Lee. 2006. Linked data: Design issues. *http://www.w3.org/DesignIssues/LinkedData.html* (2006).

[15] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The semantic web. *Scientific American* 284, 5 (2001), 34–43.

[16] Diego Berrueta, Jon Phipps, Alistair Miles, Thomas Baker, and Ralph Swick. 2008. Best practice recipes for publishing RDF vocabularies. *Working draft, W3C* (2008).

[17] Janez Brank, Marko Grobelnik, and Dunja Mladenić. 2005. A survey of ontology evaluation techniques. In *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD'05)*.

[18] Arne Bröring, Stefan Schmid, Corina-Kim Schindhelm, Abdelmajid Khelil, Sebastian Käbisch, Denis Kramer, Danh Le Phuoc, Jelena Mitic, Darko Anicic, and Ernest Teniente. 2017. Enabling IoT ecosystems through platform interoperability. *IEEE Software* 34, 1 (2017), 54–61.

[19] Paul Buitelaar, Thomas Eigner, and Thierry Declerck. 2004. OntoSelect: A dynamic ontology library with support for ontology selection. In *Proceedings of the Demo Session at the International Semantic Web Conference*.

[20] Andrew Burton-Jones, Veda C Storey, Vijayan Sugumaran, and Punit Ahluwalia. 2005. A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering* 55, 1 (2005), 84–102.

[21] Anila Sahar Butt, Armin Haller, and Lexing Xie. 2014. Ontology search: An empirical evaluation. In *Proceedings of the International Semantic Web Conference*. Springer, 130–147. https://doi.org/10.1007/978-3-319-11915-1_9

[22] Anila Sahar Butt, Armin Haller, and Lexing Xie. 2014. Relationship-based top-k concept retrieval for ontology search. In *Proceedings of the International Conference on Knowledge Engineering and Knowledge Management*. Springer, 485–502. https://doi.org/10.1007/978-3-319-13704-9_37

[23] Anila Sahar Butt, Armin Haller, and Lexing Xie. 2016. DWRank: Learning concept ranking for ontology search. *Semantic Web* 7, 4 (2016), 447–461.

[24] Anila Sahar Butt, Armin Haller, and Lexing Xie. 2016. RecOn: Ontology recommendation for structureless queries. *Applied Ontology* 11, 4 (2016), 301–324.

[25] Iván Cantador, Miriam Fernández Sánchez, and Pablo Castells. 2007. Improving ontology recommendation and reuse in WebCORE by collaborative assessments. In *Proceedings of the Workshop on Social and Collaborative Construction of Structured Knowledge at 16th International World Wide Web Conference*. CEUR. http://ceur-ws.org/Vol-273/.

[26] Claudio Carpineto and Giovanni Romano. 2012. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys (CSUR)* 44, 1 (2012), 1.

[27] Gong Cheng, Saisai Gong, and Yuzhong Qu. 2011. An empirical study of vocabulary relatedness and its application to recommender systems. *The Semantic Web–ISWC 2011* (2011), 98–113. https://doi.org/10.1007/978-3-642-25073-6_7

[28] Gong Cheng and Yuzhong Qu. 2009. Searching linked objects with falcons: Approach, implementation and evaluation. *International Journal on Semantic Web and Information Systems* 5, 3 (2009), 50–71.

[29] Mihai Codescu, Eugen Kuksa, Oliver Kutz, Till Mossakowski, and Fabian Neuhaus. 2017. Ontohub: A semantic repository engine for heterogeneous ontologies. *Applied Ontology* 12, 3-4 (2017), 275–298.

[30] Michael Compton, Cory Henson, Laurent Lefort, Holger Neuhaus, and Amit Sheth. 2009. A survey of the semantic specification of sensors. In *Proceedings of the 2nd International Conference on Semantic Sensor Networks*, Vol. 522. CEUR Workshop Proceedings, 17–32.

[31] Li Da Xu, Wu He, and Shancang Li. 2014. Internet of Things in industries: A survey. *IEEE Transactions on Industrial Informatics* 10, 4 (2014), 2233–2243.

[32] Mathieu d'Aquin, Claudio Baldassarre, Laurian Gridinoc, Marta Sabou, Sofia Angeletou, and Enrico Motta. 2007. Watson: Supporting next generation semantic web applications. In *Proceedings of the IADIS International Conference on WWW/Internet*. http://www.iadisportal.org/wwwinternet-2007-proceedings.

[33] Mathieu d'Aquin and Holger Lewen. 2009. Cupboard–a place to expose your ontologies to applications and the community. *The Semantic Web: Research and Applications* (2009), 913–918.

[34] Mathieu d'Aquin and Natalya F Noy. 2012. Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web* 11 (2012), 96–111.

[35] Giuseppe De Giacomo and Maurizio Lenzerini. 1996. TBox and ABox reasoning in expressive description logics. *AAAI Technical Report WS-96-05* 96, 316-327 (1996), 10.

[36] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs. 2004. Swoogle: A search and metadata engine for the semantic web. In *Proceedings of the 13th ACM International Conference on Information and Knowledge Management.* ACM, 652–659. https://doi.org/10.1145/1031171.1031289

[37] Astrid Duque-Ramos, Jesualdo Tomás Fernández-Breis, Miguela Iniesta, Michel Dumontier, Mikel Egaña Aranguren, Stefan Schulz, Nathalie Aussenac-Gilles, and Robert Stevens. 2013. Evaluation of the OQuaRE framework for ontology quality. *Expert Systems with Applications* 40, 7 (2013), 2696–2703.

[38] Miriam Fernández, Iván Cantador, and Pablo Castells. 2006. CORE: A tool for collaborative ontology reuse and evaluation. *Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web* (2006). http://wwwconference.org/proceedings/www2006/km.aifb.uni-karlsruhe.de/ws/eon2006/index.html.

[39] Maria Ganzha, Marcin Paprzycki, Wiesław Pawłowski, Paweł Szmeja, and Katarzyna Wasielewska. 2017. Semantic interoperability in the Internet of Things: An overview from the INTER-IoT perspective. *Journal of Network and Computer Applications* 81 (2017), 111–124.

[40] Birte Glimm, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. 2008. Conjunctive query answering for the description logic. *Journal of Artificial Intelligence Research* 31 (2008), 157–204.

[41] Asunción Gómez-Pérez. 2004. Ontology evaluation. In *Handbook on Ontologies.* Springer, 251–273. https://doi.org/10.1007/978-3-540-24750-0_13

[42] Asunción Gómez-Pérez, Mariano Fernández-López, and Oscar Corcho. 2006. *Ontological Engineering: With Examples From the Areas of Knowledge Management, E-Commerce and the Semantic Web.* Springer Science & Business Media.

[43] Tudor Groza, Tania Tudorache, and Michel Dumontier. 2013. State of the art and open challenges in community-driven knowledge curation. *Journal of Biomedical Informatics* 46, 1 (2013), 1–4.

[44] Thomas R Gruber. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5, 2 (1993), 199–220.

[45] Nicola Guarino and Christopher Welty. 2002. Evaluating ontological decisions with OntoClean. *Communications of the ACM* 45, 2 (2002), 61–65.

[46] Dominique Guinard and Vlad Trifa. 2016. *Building the Web of Things.* Manning Publications Co.

[47] Jingzhi Guo, Li Da Xu, Guangyi Xiao, and Zhiguo Gong. 2012. Improving multilingual semantic interoperation in cross-organizational enterprise systems through concept disambiguation. *IEEE Transactions on Industrial Informatics* 8, 3 (2012), 647–658.

[48] Didem Gürdür and Fredrik Asplund. 2017. A systematic review to merge discourses: Interoperability, integration and cyber-physical systems. *Journal of Industrial Information Integration* (2017).

[49] Amelie Gyrard, Christian Bonnet, Karima Boudaoud, and Martin Serrano. 2016. LOV4IoT: A second life for ontology-based domain knowledge to build semantic web of things applications. In *Proceedings of the IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud'16).* IEEE, 254–261. https://doi.org/10.1109/FiCloud.2016.44

[50] Amelie Gyrard, Martin Serrano, and Ghislain A. Atemezing. 2015. Semantic web methodologies, best practices and ontology engineering applied to Internet of Things. In *Proceedings of the 2nd World Forum on Internet of Things (WF-IoT).* IEEE, 412–417. https://doi.org/10.1109/WF-IoT.2015.7389090

[51] Sara Hachem, Thiago Teixeira, and Valérie Issarny. 2011. Ontologies for the Internet of Things. In *Proceedings of the 8th Middleware Doctoral Symposium.* ACM, 3. https://doi.org/10.1145/2093190.2093193

[52] Andreas Harth. 2009. VisiNav: Visual web data search and navigation. In *Database and Expert Systems Applications.* Springer, 214–228.

[53] Andreas Harth. 2010. VisiNav: A system for visual search and navigation on web data. *Web Semantics: Science, Services and Agents on the World Wide Web* 8, 4 (2010), 348–354.

[54] Andreas Harth, Aidan Hogan, Renaud Delbru, Jürgen Umbrich, Sean O'Riain, and Stefan Decker. 2007. SWSE: Answers before links! In *Proceedings of the 2007 International Conference on Semantic Web Challenge*, Vol. 295. CEUR, 137–144.

[55] Andreas Harth, Jürgen Umbrich, and Stefan Decker. 2006. MultiCrawler: A pipelined architecture for crawling and indexing semantic web data. In *Proceedings of the International Semantic Web Conference*, Vol. 4273. Springer, 258–271.

[56] Jens Hartmann, Raúl Palma, and Asunción Gómez-Pérez. 2009. Ontology repositories. In *Handbook on Ontologies.* Springer, 551–571. https://doi.org/10.1007/978-3-540-92673-3_25

[57] Tom Heath and Christian Bizer. 2011. Linked data: Evolving the web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology* 1, 1 (2011), 1–136.

[58] Monika Henzinger. 2007. Search technologies for the internet. *Science* 317, 5837 (2007), 468–471.

[59] Diane I. Hillmann, Stuart A. Sutton, Jon Phipps, and Ryan Laundry. 2006. A metadata registry from vocabularies up: The NSDL registry project. *Proceedings of the International Conference on Dublin Core and Metadata Applications* (2006).

[60] Hlomani Hlomani and Deborah Stacey. 2014. Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal* (2014), 1–5.

[61] Aidan Hogan, Andreas Harth, Alexandre Passant, Stefan Decker, and Axel Polleres. 2010. Weaving the pedantic web. *Proceedings of the 3rd International Workshop on Linked Data on the Web (LDOW2010)* (2010). http://ceur-ws.org/Vol-628/.

[62] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, et al. 2004. SWRL: A semantic web rule language combining OWL and RuleML. *W3C Member Submission* 21 (2004), 79.

[63] Heini Ikävalko, Petra Turkama, and Anssi Smedlund. 2018. Enabling the mapping of Internet of Things ecosystem business models through roles and activities in value co-creation. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. http://hdl.handle.net/10125/50509.

[64] Krzysztof Janowicz, Armin Haller, Simon J. D. Cox, Danh Le Phuoc, and Maxime Lefrancois. 2018. Sosa: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics* (2018).

[65] Krzysztof Janowicz, Pascal Hitzler, Benjamin Adams, Dave Kolas, and Charles Vardeman II. 2014. Five stars of linked data vocabulary use. *Semantic Web* 5, 3 (2014), 173–176.

[66] Matthew Jones and Harith Alani. 2006. Content-based ontology ranking. *Proceedings of the 9th International Protégé Conference* (2006). http://oro.open.ac.uk/id/eprint/21191.

[67] Simon Jupp, Tony Burdett, Catherine Leroy, and Helen E Parkinson. 2015. A new ontology lookup service at EMBL-EBI. In *Proceedings of Semantic Web Applications and Tools for Life Science International Conference*. 118–119. http://ceur-ws.org/Vol-1546/.

[68] Asad Masood Khattak, Dang Viet Hung, Phan Tran Ho Truc, Le Xuan Hung, D Guan, Zeeshan Pervez, Manhyung Han, Sungyoung Lee, Young-Koo Lee, et al. 2010. Context-aware human activity recognition and decision making. In *Proceedings of the 12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom'10)*. IEEE, 112–118.

[69] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele University Technical Report TR/SE-0401* (2004).

[70] Craig A Knoblock, Pedro Szekely, José Luis Ambite, Aman Goel, Shubham Gupta, Kristina Lerman, Maria Muslea, Mohsen Taheriyan, and Parag Mallick. 2012. Semi-automatically mapping structured sources into the semantic web. In *Extended Semantic Web Conference*. Springer, 375–390. https://doi.org/10.1007/978-3-642-30284-8_32

[71] Niklas Kolbe, Jérémy Robert, Sylvain Kubler, and Yves Le Traon. 2017. PROFICIENT: Productivity tool for semantic interoperability in an open IoT ecosystem. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. https://doi.org/10.1145/3144457.3144479

[72] Niklas Kolbe, Arkady Zaslavsky, Sylvain Kubler, Jérémy Robert, and Yves Le Traon. 2017. Enriching a situation awareness framework for IoT with knowledge base and reasoning components. In *Proceedings of the International and Interdisciplinary Conference on Modeling and Using Context*. Springer, 41–54. https://doi.org/10.1007/978-3-319-57837-8_4

[73] Maxim Kolchin, Nikolay Klimov, Alexey Andreev, Ivan Shilin, Daniil Garayzuev, Dmitry Mouromtsev, and Danil Zakoldaev. 2015. Ontologies for web of things: A pragmatic review. In *Proceedings of the International Conference on Knowledge Engineering and the Semantic Web*. Springer, 102–116. https://doi.org/10.1007/978-3-319-24543-0_8

[74] Sefki Kolozali, Maria Bermudez-Edo, Daniel Puschmann, Frieder Ganz, and Payam Barnaghi. 2014. A knowledge-based approach for real-time IoT data stream annotation and processing. In *Proceedings of the IEEE International Conference on Internet of Things (iThings 2014), Green Computing and Communications (GreenCom 2014), and Cyber-Physical-Social Computing (CPSCom'14)*. IEEE, 215–222. https://doi.org/10.1109/iThings.2014.39

[75] Erno Kovacs, Martin Bauer, Jaeho Kim, Jaeseok Yun, Franck Le Gall, and Mengxuan Zhao. 2016. Standards-based worldwide semantic interoperability for IoT. *IEEE Communications Magazine* 54, 12 (2016), 40–46.

[76] Andreas Langegger and Wolfram Woss. 2009. RDFStats: An extensible rdf statistics generator and library. In *Proceedings of the 20th International Workshop on Database and Expert Systems Application*. IEEE, 79–83. https://doi.org/10.1109/DEXA.2009.25

[77] Holger Lewen, Kaustubh Supekar, Natalya F. Noy, and Mark A. Musen. 2006. Topic-specific trust and open rating systems: An approach for ontology evaluation. In *Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web*. https://km.aifb.kit.edu/ws/eon2006/.

[78] Shancang Li, Li Da Xu, and Shanshan Zhao. 2018. 5G internet of things: A survey. *Journal of Industrial Information Integration* (2018).

[79] Fei Liu, Chee-Wee Tan, Eric TK Lim, and Ben Choi. 2017. Traversing knowledge networks: An algorithmic historiography of extant literature on the Internet of Things (IoT). *Journal of Management Analytics* 4, 1 (2017), 3–34.

[80] Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331.

[81] Adolfo Lozano-Tello and Asunción Gómez-Pérez. 2004. ONTOMETRIC: A method to choose the appropriate ontology. *Journal of Database Management (JDM)* 2, 15 (2004), 1–18.

[82] Alexander Maedche and Steffen Staab. 2001. Ontology learning for the semantic web. *IEEE Intelligent Systems* 16, 2 (2001), 72–79.

[83] Eetu Mäkelä. 2014. Aether–generating and viewing extended VoID statistical descriptions of RDF datasets. In *European Semantic Web Conference*. Springer, 429–433. https://doi.org/10.1007/978-3-319-11955-7_61

[84] Bogdan Manate, Victor Ion Munteanu, and Teodor Florin Fortis. 2014. Towards a smarter Internet of Things: Semantic visions. In *Proceedings of the Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*. IEEE, 582–587. https://doi.org/10.1109/CISIS.2014.84

[85] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, et al. 2008. *Introduction to Information Retrieval.* Vol. 1. Cambridge University Press.

[86] Marcos Martínez-Romero, Clement Jonquet, Martin J. O'Connor, John Graybeal, Alejandro Pazos, and Mark A. Musen. 2017. NCBO ontology recommender 2.0: An enhanced approach for biomedical ontology recommendation. *Journal of Biomedical Semantics* 8, 1 (2017), 21.

[87] Marcos Martínez-Romero, José M Vázquez-Naya, Javier Pereira, and Alejandro Pazos. 2014. BiOSS: A system for biomedical ontology selection. *Computer Methods and Programs in Biomedicine* 114, 1 (2014), 125–140.

[88] Nicolas Matentzoglu, Daniel Tang, Bijan Parsia, and Uli Sattler. 2014. The manchester OWL repository: System description. In *Proceedings of the 2014 International Semantic Web Conference on Posters & Demonstrations Track*. CEUR, 285–288. https://dl.acm.org/citation.cfm?id=2878525.

[89] Brian McBride. 2004. The resource description framework (RDF) and its vocabulary description language RDFS. In *Handbook on Ontologies*. Springer, 51–65. https://doi.org/10.1007/978-3-540-24750-0_3

[90] Eric Miller. 1998. An introduction to the resource description framework. *Bulletin of the Association for Information Science and Technology* 25, 1 (1998), 15–19.

[91] George A Miller. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38, 11 (1995), 39–41.

[92] Julien Mineraud, Oleksiy Mazhelis, Xiang Su, and Sasu Tarkoma. 2016. A gap analysis of Internet-of-Things platforms. *Computer Communications* 89 (2016), 5–16.

[93] R. Minerva. 2015. Towards a definition of the Internet of Things. *IEEE Internet Initiative* (2015).

[94] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, Prisma Group, et al. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine* 6, 7 (2009), 1–6.

[95] Michael Mrissa, Lionel Médini, and Jean-Paul Jamont. 2014. Semantic discovery and invocation of functionalities for the web of things. In *Proceedings of the 23rd International WETICE Conference (WETICE)*. IEEE, 281–286. http://doi.ieeecomputersociety.org/10.1109/WETICE.2014.50

[96] Sebastian Neumaier, Axel Polleres, Simon Steyskal, and Jürgen Umbrich. 2017. Data integration for open data on the web. In *Reasoning Web International Summer School*. Springer, 1–28. https://doi.org/10.1007/978-3-319-61033-7_1

[97] Natalya F. Noy, Nigam H. Shah, Patricia L. Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L. Rubin, Margaret-Anne Storey, Christopher G. Chute, et al. 2009. BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* 37 (2009).

[98] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report. Stanford InfoLab.

[99] Raúl Palma, Peter Haase, and Asunción Gómez-Pérez. 2006. Oyster: Sharing and re-using ontologies in a peer-to-peer community. In *Proceedings of the 15th International Conference on World Wide Web.* ACM, 1009–1010. https://doi.org/10.1007/11574620_77

[100] Jeff Z Pan, Edward Thomas, and Derek Sleeman. 2006. ONTOSEARCH2: Searching and querying web ontologies. *Proceedings of the IADIS International Conference on WWW/Internet* 2006 (2006), 211–218. http://www.iadisportal.org/e-society-2006-proceedings.

[101] Chintan Patel, Kaustubh Supekar, Yugyung Lee, and E. K. Park. 2003. OntoKhoj: A semantic web portal for ontology searching, ranking and classification. In *Proceedings of the 5th ACM International Workshop on Web Information and Data Management.* ACM, 58–61. https://doi.org/10.1145/956699.956712

[102] Pankesh Patel, Amelie Gyrard, Soumya Kanti Datta, and Muhammad Intizar Ali. 2017. SWoTSuite: A toolkit for prototyping end-to-end semantic web of things applications. In *Proceedings of the 26th International World Wide Web Conference.* 263–267. https://doi.org/10.1145/3041021.3054736

[103] Charith Perera. 2016. Sensing as a service (S2aaS): Buying and selling IoT data. *IEEE IoT Newsletter* (2016).

[104] Charith Perera, Yongrui Qin, Julio C. Estrella, Stephan Reiff-Marganiec, and Athanasios V. Vasilakos. 2017. Fog computing for sustainable smart cities: A survey. *ACM Computing Surveys (CSUR)* 50, 3 (2017), 32.

[105] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2014. Context aware computing for the internet of things: A survey. *IEEE Communications Surveys & Tutorials* 16, 1 (2014), 414–454.
[106] Charith Perera, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. 2014. Sensing as a service model for smart cities supported by Internet of Things. *Transactions on Emerging Telecommunications Technologies* 25, 1 (2014), 81–93.
[107] Dennis Pfisterer, Kay Romer, Daniel Bimschas, Oliver Kleine, Richard Mietz, Cuong Truong, Henning Hasemann, Alexander Kröller, Max Pagel, Manfred Hauswirth, et al. 2011. SPITFIRE: Toward a semantic web of things. *IEEE Communications Magazine* 49, 11 (2011), 40–48.
[108] María Poveda Villalón, Raul García Castro, and A Gómez-Pérez. 2014. Building an ontology catalogue for smart cities. *10th European Conference on Product & Process Modelling* (2014). http://oa.upm.es/36715/.
[109] María Poveda-Villalón, Mari Suárez-Figueroa, and Asunción Gómez-Pérez. 2012. Validating ontologies with OOPS! *Knowledge Engineering and Knowledge Management* (2012), 267–281.
[110] Maria Poveda Villalon, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 2010. A double classification of common pitfalls in ontologies. *Workshop on Ontology Quality (OntoQual 2010)* (2010). http://oa.upm.es/5413/.
[111] World Wide Web Consortium. 2008. *SPARQL Query Language for RDF*. https://www.w3.org/TR/rdf-sparql-query/.
[112] Eric Prud'hommeaux and Ryan Lee. 2007. *W3C RDF Validation Service*. https://www.w3.org/RDF/Validator/.
[113] Yuzhong Qu and Gong Cheng. 2011. Falcons concept search: A practical search engine for web ontologies. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* 41, 4 (2011), 810–816.
[114] Filip Radulovic, María Poveda-Villalón, Daniel Vila-Suero, Víctor Rodríguez-Doncel, Raúl García-Castro, and Asunción Gómez-Pérez. 2015. Guidelines for linked data generation and publication: An example in building energy consumption. *Automation in Construction* 57 (2015), 178–187.
[115] D. Raggett. 2015. The web of things: Challenges and opportunities. *Computer* 5 (2015), 26–32.
[116] Robert G Raskin and Michael J Pan. 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). *Computers & Geosciences* 31, 9 (2005), 1119–1125.
[117] S. Rhee. 2016. Catalyzing the Internet of Things and smart cities: Global city teams challenge. In *Proceedings of the 1st International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE) in Partnership with Global City Teams Challenge*. 1–4. https://doi.org/10.1109/SCOPE.2016.7515058
[118] Antonio J Roa-Valverde and Miguel-Angel Sicilia. 2014. A survey of approaches for ranking on the web of data. *Information Retrieval* 17, 4 (2014), 295–325.
[119] Jérémy Robert, Sylvain Kubler, Niklas Kolbe, Alessandro Cerioni, Emmanuel Gastaud, and Kary Främling. 2017. Open IoT ecosystem for enhanced interoperability in smart cities: Example of Métropole De Lyon. *Sensors* 17, 12 (2017), 2849.
[120] Carlos Rueda, Luis Bermudez, and Janet Fredericks. 2009. The MMI ontology registry and repository: A portal for marine metadata interoperability. In *Biloxi-Marine Technology for Our Future: Global and Local Challenges*. IEEE, 1–6.
[121] Thomas L. Saaty. 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences* 1, 1 (2008), 83–98.
[122] Marta Sabou, Vanessa Lopez, Enrico Motta, and Victoria Uren. 2006. Ontology selection: Ontology evaluation on the real semantic web. In *Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web*.
[123] Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24, 5 (1988), 513–523.
[124] Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM* 18, 11 (1975), 613–620.
[125] Eduardo Felipe Zambom Santana, Ana Paula Chaves, Marco Aurelio Gerosa, Fabio Kon, and Dejan S Milojicic. 2017. Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 78.
[126] Johann Schaible. 2017. *TermPicker: Recommending Vocabulary Terms for Reuse When Modeling Linked Open Data*. Ph.D. Dissertation. Christian-Albrechts Universität Kiel.
[127] Johann Schaible, Thomas Gottron, and Ansgar Scherp. 2014. Survey on common strategies of vocabulary reuse in linked open data modeling. In *Proceedings of the European Semantic Web Conference*. Springer, 457–472. https://doi.org/10.1007/978-3-319-07443-6_31
[128] Johann Schaible, Thomas Gottron, and Ansgar Scherp. 2016. TermPicker: Enabling the reuse of vocabulary terms by exploiting data from the linked open data cloud. In *Proceedings of the International Semantic Web Conference*. Springer, 101–117. https://doi.org/10.1007/978-3-319-34129-3_7
[129] François Scharffe, Ghislain Atemezing, Raphaël Troncy, Fabien Gandon, Serena Villata, Bénédicte Bucher, Fayçal Hamdi, Laurent Bihanic, Gabriel Képéklian, Franck Cotton, et al. 2012. Enabling linked-data publication

with the datalift platform. In *Proceedings of the AAAI Workshop on Semantic Cities*. http://research.ihost.com/semanticcities12/index.html.

[130] Daniel Schober, Ilinca Tudose, Vojtech Svatek, and Martin Boeker. 2012. OntoCheck: Verifying ontology naming conventions and metadata completeness in Protégé 4. *Journal of Biomedical Semantics* 3, 2 (2012).

[131] Michael Sheng, Yongrui Qin, Lina Yao, and Boualem Benatallah. 2017. *Managing the Web of Things: Linking the Real World to the Web*. Elsevier.

[132] Feifei Shi, Qingjuan Li, Tao Zhu, and Huansheng Ning. 2018. A survey of data semantization in Internet of Things. *Sensors* 18, 1 (2018), 313.

[133] Elena Simperl. 2009. Reusing ontologies on the semantic web: A feasibility study. *Data & Knowledge Engineering* 68, 10 (2009), 905–925.

[134] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. 2007. Pellet: A practical OWL-DL reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web* 5, 2 (2007), 51–53.

[135] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J. Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, et al. 2007. The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25, 11 (2007), 1251.

[136] Steffen Stadtmüller, Andreas Harth, and Marko Grobelnik. 2013. Accessing information about linked data vocabularies with vocab.cc. In *Semantic Web and Web Science*. Springer, 391–396. https://doi.org/10.1007/978-1-4614-6880-6_34

[137] John A Stankovic. 2014. Research directions for the internet of things. *IEEE Internet of Things Journal* 1, 1 (2014), 3–9.

[138] Ioannis Stavrakantonakis, Anna Fensel, and Dieter Fensel. 2015. Linked open vocabulary recommendation based on ranking and linked open data. In *Proceedings of the Joint International Semantic Technology Conference*. Springer, 40–55. https://doi.org/10.1007/978-3-319-31676-5_3

[139] Darijus Strasunskas and Stein L Tomassen. 2008. On significance of ontology quality in ontology-driven web search. In *Proceedings of the World Summit on Knowledge Society*. Springer, 469–478. https://doi.org/10.1007/978-3-540-87781-3_51

[140] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi. 2012. *Ontology Engineering in a Networked World*. Springer Science & Business Media.

[141] R Subhashini, J. Akilandeswari, and V. Sinthuja. 2011. A review on ontology ranking algorithms. *International Journal of Computer Applications* 33, 4 (2011), 6–11.

[142] Harald Sundmaeker, Patrick Guillemin, Peter Friess, and Sylvie Woelfflé. 2010. *Vision and Challenges for Realising the Internet of Things*. Publications Office of the European Union.

[143] Kaustubh Supekar, Chintan Patel, and Yugyung Lee. 2004. Characterizing quality of knowledge on semantic web. In *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference*. AAAI, 472–478. https://aaai.org/Library/FLAIRS/flairs04contents.php.

[144] Kaustubh Supekar, Daniel Rubin, Natasha Noy, and Mark Musen. 2007. Knowledge zone: A public repository of peer-reviewed biomedical ontologies. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*. IOS Press, 812.

[145] J. Swetina, G. Lu, P. Jacobs, F. Ennesser, and J. Song. 2014. Toward a standardized common M2M service layer platform: Introduction to oneM2M. *IEEE Wireless Communications* 21, 3 (2014), 20–26.

[146] Ioan Szilagyi and Patrice Wira. 2016. Ontologies and semantic web for the Internet of Things: A survey. In *Proceedings of the 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON'16)*. IEEE, 6949–6954. https://doi.org/10.1109/IECON.2016.7793744

[147] Samir Tartir, I. Budak Arpinar, Michael Moore, Amit P. Sheth, and Boanerges Aleman-Meza. 2005. OntoQA: Metric-based ontology quality analysis. In *Proceedings of IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*. 45–53. https://pdfs.semanticscholar.org/58de/060fc301c7797b3ae9909255ff4dfc8821e3.pdf?_ga=2.185590377.1788324653.1547124271-1988443978.1547124271.

[148] Samir Tartir, I. Budak Arpinar, and Amit P. Sheth. 2010. Ontological evaluation and validation. In *Theory and Applications of Ontology: Computer Applications*. Springer, 115–130.

[149] Andreas Tolk, Saikou Diallo, and Charles Turnitsa. 2007. Applying the levels of conceptual interoperability model in support of integratability, interoperability, and composability for system-of-systems engineering. *Journal of Systemics, Cybernetics and Informatics* (2007).

[150] Nguyen Khoi Tran, Quan Z. Sheng, Muhammad Ali Babar, and Lina Yao. 2017. Searching the web of things: State of the art, challenges, and solutions. *ACM Computing Surveys (CSUR)* 50, 4 (2017), 55.

[151] Giovanni Tummarello, Renaud Delbru, and Eyal Oren. 2007. Sindice.com: Weaving the open linked data. In *The Semantic Web*. Springer, 552–565. https://doi.org/10.1007/978-3-540-76298-0_40

[152] Rachanee Ungrangsi, Chutiporn Anutariya, and Vilas Wuwongse. 2007. combiSQORE: An ontology combination algorithm. *The Semantic Web* (2007), 566–579.

[153] Rachanee Ungrangsi, Chutiporn Anutariya, and Vilas Wuwongse. 2009. SQORE: An ontology retrieval framework for the next generation Web. *Concurrency and Computation: Practice and Experience* 21, 5 (2009), 651–671.

[154] Pierre-Yves Vandenbussche, Ghislain A Atemezing, María Poveda-Villalón, and Bernard Vatant. 2017. Linked open vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web. *Semantic Web* 8, 3 (2017), 437–452.

[155] Ovidiu Vermesan and Peter Friess. 2016. *Digitising the Industry: Internet of Things Connecting the Physical, Digital and Virtual Worlds*. River Publishers.

[156] Kim Viljanen, Jouni Tuominen, Eetu Mäkelä, and Eero Hyvönen. 2012. Normalized access to ontology repositories. In *IEEE Sixth International Conference on Semantic Computing (ICSC)*. IEEE, 109–116. https://doi.org/10.1109/ICSC.2012.56

[157] Wei Wang, Suparna De, Ralf Toenjes, Eike Reetz, and Klaus Moessner. 2012. A comprehensive ontology for knowledge representation in the Internet of Things. In *Proceedings of the11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 1793–1798. https://doi.org/10.1109/TrustCom.2012.20

[158] Andrew Whitmore, Anurag Agarwal, and Li Da Xu. 2015. The Internet of Things: A survey of topics and trends. *Information Systems Frontiers* 17, 2 (2015), 261–274.

[159] Qiang Wu, Christopher J. C. Burges, Krysta M, Svore, and Jianfeng Gao. 2010. Adapting boosting for information retrieval measures. *Information Retrieval* 13, 3 (2010), 254–270.

[160] Marcin Wylot, Manfred Hauswirth, Philippe Cudré-Mauroux, and Sherif Sakr. 2018. RDF data storage and query processing schemes: A survey. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 84.

[161] Zuoshuang Xiang, Chris Mungall, Alan Ruttenberg, and Yongqun He. 2011. Ontobee: A linked data server and browser for ontology terms. In *Proceedings of the International Conference on Biomedical Ontology*. http://ceur-ws.org/Vol-833/.

[162] Guangyi Xiao, Jingzhi Guo, Zhiguo Gong, and Renfa Li. 2016. Semantic input method of Chinese word senses for semantic document exchange in e-business. *Journal of Industrial Information Integration* 3 (2016), 31–36.

[163] Juan Ye, Stamatia Dasiopoulou, Graeme Stevenson, Georgios Meditskos, Efstratios Kontopoulos, Ioannis Kompatsiaris, and Simon Dobson. 2015. Semantic web technologies in pervasive computing: A survey and research roadmap. *Pervasive and Mobile Computing* 23 (2015), 1–25.

[164] Andrea Zanella, Nicola Bui, Angelo Castellani, et al. 2014. Internet of things for smart cities. *IEEE Internet of Things Journal* 1, 1 (2014), 22–32.

[165] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2016. Quality assessment for linked data: A survey. *Semantic Web* 7, 1 (2016), 63–93.

[166] Yi Zhang, Wamberto Vasconcelos, and Derek Sleeman. 2005. OntoSearch: An ontology search engine. In *Research and Development in Intelligent Systems XXI*. Springer, 58–69. https://doi.org/10.1007/1-84628-102-4_5

Received March 2018; revised September 2018; accepted September 2018