



Article

An Ontology-Driven Personalized Faceted Search for Exploring Knowledge Bases of Capsicum

Zaenal Akbar ^{1,*}, Hani Febri Mustika ¹, Dwi Setyo Rini ², Lindung Parningotan Manik ¹, Ariani Indrawati ¹, Agusdin Dharma Fefirenta ² and Tutie Djarwaningsih ²

¹ Research Center for Informatics, Indonesian Institute of Sciences, Jakarta 12710, Indonesia; hani.febri.mustika@lipi.go.id (H.F.M.); lindung.parningotan.manik@lipi.go.id (L.P.M.); ariani.indrawati@lipi.go.id (A.I.)

² Research Center for Biology, Indonesian Institute of Sciences, Jakarta 12710, Indonesia; dwi.setyo.rini@lipi.go.id (D.S.R.); agus123@lipi.go.id (A.D.F.); tutie.djarwaningsih@lipi.go.id (T.D.)

* Correspondence: zaenal.akbar@lipi.go.id; Tel.: +62-21-87917216

Abstract: Capsicum is a genus of flowering plants in the Solanaceae family in which the members are well known to have a high economic value. The Capsicum fruits, which are popularly known as peppers or chili, have been widely used by people worldwide. It serves as a spice and raw material for many products such as sauce, food coloring, and medicine. For many years, scientists have studied this plant to optimize its production. A tremendous amount of knowledge has been obtained and shared, as reflected in multiple knowledge-based systems, databases, or information systems. An approach to knowledge-sharing is through the adoption of a common ontology to eliminate knowledge understanding discrepancy. Unfortunately, most of the knowledge-sharing solutions are intended for scientists who are familiar with the subject. On the other hand, there are groups of potential users that could benefit from such systems but have minimal knowledge of the subject. For these non-expert users, finding relevant information from a less familiar knowledge base would be daunting. More than that, users have various degrees of understanding of the available content in the knowledge base. This understanding discrepancy raises a personalization problem. In this paper, we introduce a solution to overcome this challenge. First, we developed an ontology to facilitate knowledge-sharing about Capsicum to non-expert users. Second, we developed a personalized faceted search algorithm that provides multiple structured ways to explore the knowledge base. The algorithm addresses the personalization problem by identifying the degree of understanding about the subject from each user. In this way, non-expert users could explore a knowledge base of Capsicum efficiently. Our solution characterized users into four groups. As a result, our faceted search algorithm defines four types of matching mechanisms, including three ranking mechanisms as the core of our solution. In order to evaluate the proposed method, we measured the predictability degree of produced list of facets. Our findings indicated that the proposed matching mechanisms could tolerate various query types, and a high degree of predictability can be achieved by combining multiple ranking mechanisms. Furthermore, it demonstrates that our approach has a high potential contribution to biodiversity science in general, where many knowledge-based systems have been developed with limited access to users outside of the domain.

Keywords: ontology; personalized faceted search; knowledge base; Capsicum



Citation: Akbar, Z.; Mustika, H.F.; Rini, D.S.; Manik, L.P.; Indrawati, A.; Fefirenta, A.D.; Djarwaningsih, T. An Ontology-Driven Personalized Faceted Search for Exploring Knowledge Bases of Capsicum. *Future Internet* **2021**, *13*, 172. <https://doi.org/10.3390/fi13070172>

Academic Editor: András Micsik

Received: 2 June 2021

Accepted: 24 June 2021

Published: 30 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Along with the advancement of Information and Communication Technology (ICT), scientists have generated a tremendous amount of data, including biodiversity data. The era of Biodiversity Big Data has already emerged [1]. Biodiversity data cover a wide range of life forms on Earth within its many regions, ecosystems, and habitats. Multiple new challenges have been introduced including data collection and processing, mobilization, imputation, sharing, and integration [2]. Therefore, more advanced technologies are needed

to manage it. Data-intensive science [3], which is also recognized as the fourth paradigm of scientific discovery, serves as a scientific methodology to analyze the large volume of biodiversity data.

One type of biodiversity data is the characteristics of living organisms at the morphological level. Specific morphological characteristics provide basic information for understanding the structure of an organism, the relationship between structure and function, as well as plant classification. The morphological characteristics also provide plant biologists and taxonomists with a framework to assess the differences or similarities between species. Therefore, the considerable data of characteristics serve as a favorable parameter for accurate identification and description of the plant species. As an example, biologists have used big data and genetic approaches to understand the evolution of the plant form and physiology [4].

Capsicum is a genus of flowering plants in the Solanaceae family in which the members are well known for having a high economic value. Capsicum fruits, which are popularly known as peppers or chili, have been widely used by people as spices in many cuisines worldwide. Chilli also can be used as a raw material for many products such as sauces, food coloring, and medicine. The genus of Capsicum consists of nearly 30 species and approximately 50,000 varieties [5]. Certainly, there is great variation between and within the pepper species.

The morphological characteristics of Capsicum can be used as essential information to identify the species of a plant. That information has been collected and shared through multiple knowledge-based systems, online databases, or information systems. Regarding data-intensive science, the challenge lies in the data-sharing approach. Most of the data are shared among scientists or experts who know the meaning of the data. Hence, the data are difficult to be consumed by non-expert users or the public. On the other hand, the latter group of users has been recognized as an essential constituent in the scientific discovery process, especially in collecting substantial amounts of data and engaging with the public [6].

In this paper, we introduce a method to overcome this challenge. Our method utilizes a faceted search to organize information such that data in a repository can be explored more systematically. This method would help non-expert users to start and focus on finding relevant information from a large number of knowledge bases that contain characteristics of Capsicum.

1.1. Motivation

A solution to overcome the knowledge understanding discrepancy of a domain of interest is to use an ontology-based approach. An ontology—shared, explicit and formal conceptualizations of a domain [7]—describes a knowledge-based program through the definition of a set of representational terms. As a set of objects and relationships among them, a common ontology is highly suitable for a variety of knowledge-sharing activities to guarantee consistency.

To share knowledge about plant anatomy, morphology, and the stages of plant development, the Plant Ontology (PO) has been developed and continuously expanded [8]. PO adopts the data model of the Gene Ontology (GO) [9] to annotate gene expression and phenotype data of plant structures and stages of plant development. As a common reference ontology for plant structures and development stages, the PO solves terminology disparities used by scientists from different projects and groups. The PO has been used to integrate multiple online plant genomics portals and databases, such as the Arabidopsis Information Resource (TAIR) (<https://arabidopsis.org> (accessed on 28 June 2021)), the Sol Genomics Network (SGN) (<https://solgenomics.net> (accessed on 28 June 2021)), the Maize Genetics and Genomics Database (MaizeGDB) (<https://maizegdb.org> (accessed on 28 June 2021)), the Oryzabase Database (<http://shigen.nig.ac.jp> (accessed on 28 June 2021)), and the Gramene Database (<http://gramene.org> (accessed on 28 June 2021)) to name a few.

The PO has been used successfully to share knowledge among scientists, for example, to accurately describe the plant development stages across species [10]. Unfortunately, the

ontology was designed to share knowledge among scientists who have knowledge about the subject. It is not necessarily consumable by non-expert users, those who are not familiar with the subject. These non-expert users include students, young scientists, or even citizen scientists concerned with the knowledge. This group of users has been recognized as an essential research tool due to its capability in providing data at an extensive scale and fine-grained resolution [11]. Besides being powerful in providing large amounts of data, citizen scientists could convey information to the public more conveniently [6].

Based on the previously described situations, we outlined a few things that motivated our work as follow:

1. The PO can be used to describe plant characteristics, from anatomy and morphology to the stages of plant development. It is suitable to share knowledge among scientists but not necessarily with non-expert users.
2. For non-expert users, when describing a less familiar object (for example, a flower of a plant), they tend to describe it based on generic properties or attributes. For example, to describe the petal of a flower, they would describe it based on familiar properties such as color, size, texture, etc.

1.2. Challenges

For non-expert users, finding relevant information from a less familiar knowledge base would be a daunting task from the beginning. How to formulate search keywords, refine search keywords, and filter the results, are examples of challenges facing this group of users. In this work, we formulated these challenges as follow:

1. How to start to explore a knowledge base of Capsicum by describing a generic morphological character. Searching should start from a point, for example, by defining at least one plant character. The start point could be any point in the knowledge base, regardless of its generality or specificity.
2. How to refine the search results by selecting the most relevant criteria/group. Finding the most relevant criteria/group is the main challenge.
3. How to sort multiple results to be presented to the users. When multiple criteria/groups are identified as relevant, they need to be sorted to provide users with the most relevant first. Finding the way to sort the results is the next challenge.

Our main goal is to enable non-expert users with minimal knowledge of plant characteristics to consume the collected knowledge of Capsicum. Our approach relies on an information searching technique, a so-called faceted search. Faceted search is a search technique organizing the search outputs into groups with different topics that enable users to filter the results and to find the desired information quickly [12]. However, in contrast with existing works that have utilized faceted search as a solution to information overload, we use the technique to help non-expert users explore the less familiar database. The method suggests the most relevant criteria/group that can be used to narrow down the search and to focus the data exploration process.

The rest of the paper is organized as follows. A few related works are discussed in Section 2, where we also outline our contributions. In Section 3, we describe our approach, especially the proposed algorithm that consists of two parts matching and ranking procedures. The implementation of the algorithm is described in Section 4. We evaluate the algorithm and discuss our findings in Section 5 before finally summarizing this paper with a few conclusions and future works in Section 6.

2. Related Work

We align our work with two broad research areas: the development of ontology as a bridge to unify diverse terminologies in plant science and ontology-driven faceted search. This section describes several related works from each area and outlines our contribution at the end.

Ontology has been recognized as a vital component for interoperability across knowledge-based systems [7]. Ontologies are fundamental for unifying diverse terminologies and

are increasingly used by scientists in many fields, including the online web search engines [13]. An ontology-underpinned emergency response system for water pollution accidents has been proposed to meet the demand of the government and public users for sustainable monitoring and real-time emergency response [14]. An ontological model can also represent context, driven by events, in academic domains by integrating five modular contexts, namely Person, temporal (time), physical space (location), network, and academic events [15]. Furthermore, an ontology also can be used to integrate a variety of quality assessment methods into a unified model for assessing the quality of a website [16].

Over the past years, many structured vocabularies, databases, and information systems have been developed to allow scientists to exchange knowledge about plant traits [17]. Ontology-based solutions are widely used to represent knowledge in this domain, which includes a set of terms to describe the classes in the domain and the relationships among terms [18]. For example, a schema to represent data from multiple biodiversity information systems that are available on the Web was constructed to enable Linked Biodiversity Data [19]. The International Rice Information System (IRIS) was developed to handle rice functional genomics data diversity, including genomic sequence data, molecular genetic data, expression data, and proteomic information [20]. The BRENDA enzyme information system (<https://brenda-enzymes.org/> (accessed on 28 June 2021)) was developed from a database to a competence center for enzyme-related information, which combines manually curated enzyme data with proteomic and genomic information [21]. The Pepper Expressed Sequence Tags (EST) database was constructed that consists of 122,582 sequenced ESTs and 116,412 refined ESTs that are available from 21 pepper EST libraries [22]. Much work has also been conducted utilizing local resources. For example, the development of an ontology for an Indonesian medicinal plant [23], an ontology for Thai Zingiberaceae [24], and an ontology for plant genetic resources in the Gene bank of the Institute of Plant Genetic Resources in the town of Sadovo [25]. Specifically for the Plant Ontology (PO), it can be used to describe not only plant anatomy and morphology but also the stages of plant development [8]. Knowledge and information detail about plant traits, genotype, and phenotype are usually used as basic information in PO. It should facilitate a formal description of phenotypes and standardized annotation of plant traits that accurately describe plant anatomy and morphology. In the previous release of PO, the so-called Plant Structure Ontology (PSO), there were three types of parent–child relationships in PSO that can be used to associate two terms, namely *is_a*, *part_of*, and *develops_from* [26]. PO in combination with another ontology can also be used to extract entity–quality relationships from digitized taxon descriptions [27].

On the other hand, faceted search as a human–computer interaction technique has been identified as a practical approach to handle a vast amount of data [28]. It provides interactivity and easy-to-use data visualization solution to guide the decision-making process through the classification of information, so-called “facets”. It has been used in many domains, for example, to provide dynamic faceted search solutions over enterprise databases for a domain-independent system [29] and to ensure a high-quality recommendation of scientific articles based on a query paper [30]. A visual recommendation system can also be built based on a faceted browsing technique that provides interactive navigation of automatically generated visualizations [31]. In the bioenergy domain, a faceted search system can be used to eliminate sense disambiguation of search terms [32]. A faceted search visualization technique can also support categorized access to heterogeneous and unstructured biomedical data sources [33]. This technique can also support a metasearch system to search and gather data resources from multiple Open Linked Data Projects [34]. Furthermore, it facilitates a coarse-grained and fine-grained exploration of geographic maps through interactive widgets [35].

In most of the faceted-search solutions, ontology plays a significant role. An ontology can be integrated with faceted navigation to improve information retrieval results through a query expansion mechanism [36]. By representing entities of an ontology as the facets and ontological instances as facets values, a personalized search interface can be constructed

by matching the facets and user profile [37]. An Unsupervised Ontology was extracted from heterogeneous and unstructured biomedical content to be aligned with multiple existing biomedical ontologies to enable data exploration from heterogeneous biomedical resources [33]. For ontology-driven solutions, multiple search techniques/algorithms have been used to explore relatively large search spaces, for example, using the Hill-Climbing algorithm to learn the Bayesian network of a mushrooms dataset to identify the most relevant content [28]. A random walk-based framework has been used to induce a sub-network consisting of related nodes of the scientific article citations or content similarity network [30]. Relevant facets can also be predicted by using a Random Forest (RF) model [38], based on the most frequent queries of the most similar users [34], based on the ability of the user to provide desired values for each facet [29], as well as based on statistical and perceptual measures [31].

As mentioned in Section 1, several ontologies have been developed for plant research. The most prominent one is the Plant Ontology (PO), which adopts a data model of Gene Ontology (GO) that covers flowering plants in general. GO, for example, has been used to cluster and assemble the Pepper EST database to contribute to the analysis of gene synteny as part of a chili pepper sequencing project [22]. Using PO as the basis for an ontology-driven faceted search is an ideal solution. However, as mentioned in Section 1, the ontology is too complex and hardly understandable by non-expert users.

In this work, we attempt to construct a search-based method that can be used to explore a knowledge base of Capsicum. The goal is to provide a faceted search that enables non-expert users to explore knowledge bases with less prior knowledge about characteristics of Capsicum. Our work contributes in at least two aspects:

1. Development of an ontology that intends to communicate knowledge to non-expert users. Instead of using existing ontologies, we developed a small yet powerful ontology to describe the characteristics of Capsicum. The ontology was not intended to be complete but to be easily consumed by non-expert users.
2. Utilization of faceted search technique to drive search process. This technique has been widely used to overcome information overload or searching from a large amount of data. In contrast, our faceted search was intended to search from an unfamiliar database where the amount of data is not necessarily significant.

3. Method

In this section, we discuss our research methodology. First, we introduce our solution, followed by our research procedures. After that, we explain each research step especially for knowledge modeling, matching, and ranking procedures. At the end, we describe our method to evaluate the results.

3.1. Capsicum Search: A Personalized Faceted Search

We propose a personalized faceted search solution for exploring the knowledge bases of Capsicum. The proposed solution consists of two parts: an ontology as a shared common understanding of the subject and a search algorithm. The ontology will be used as the base to produce a list of facets, while the search algorithm will filter and order the list further according to specific criteria. It is expected that the constructed list of facets is highly appropriate for the need of users. The search algorithm can be explained as follows:

1. A domain ontology is represented as a Directed Acyclic Graph (DAG) that consists of vertices and edges starting from the most generic part of a plant. The leaves of the graph are the most specific parts of the plant.
2. A list of queries is represented as path traversal procedures, where the encoded entities and properties can be located correctly in the graph.
3. Based on the graph and path traversal procedures, relevant entities are identified based on entities' relationships in the graph, for example, based on relationships of siblings, sub-graphs, etc.

- All relevant entities become the list of facets to be presented to the users that can be used to refine their search results further.

3.2. Research Procedures

Figure 1 shows our research procedures, which consist of several steps. Before explaining each step, we present some basic definitions that we borrow from graph theory [39]. To avoid confusion, we use entity and node interchangeably, where an entity is represented as a node in the graph.

Definition 1 (Knowledge Base as Graph). *Knowledge Base is a collection of knowledge obtained from experts. The knowledge is represented as a graph, where nodes of the graph are related to each other using specific relationships. A tree is a special kind of graph that contains no cycles.*

Definition 2 (Level). *A level of a node in a tree is how far the node is from the root, where the root has level 0. The higher the level, the further the node from the root.*

Definition 3 (Path). *A path between two nodes is the possible way to reach one node from another.*

Definition 4 (Branch). *The branch of a tree, a sub-tree consists of smaller connected nodes. Two nodes are in the same branch if there is a path between them without going through the root.*

We explain several most important research steps in the following sub-sections, including knowledge modeling, matching, and ranking procedures.

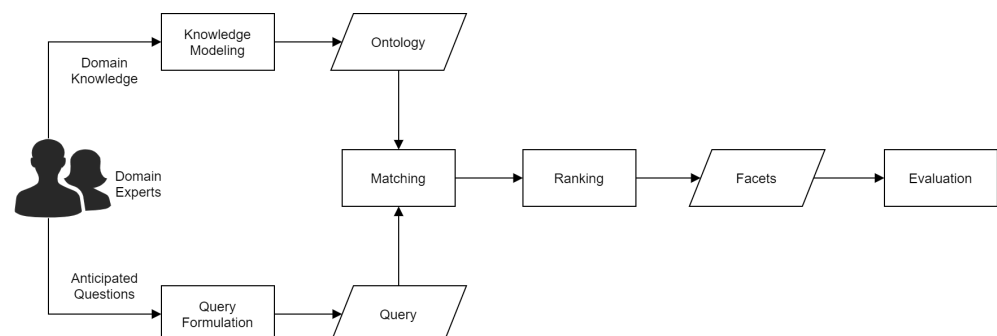


Figure 1. Our research procedures and its evaluation.

3.3. Knowledge Modeling and Query Formulation

In this step, knowledge from domain experts was acquired through a series of intensive discussions and become the base of our ontology development. There are two groups of ontology development models that are widely used, namely waterfall and iterative-increment life-cycle models [40,41]. Multiple prominent ontology engineering methodologies come under the latter group, including Methontology [42] and NeOn [40]. For example, Methontology has been successfully used to develop ontologies in the legal domain [43], while NeOn methodology in the human resources domain [44] and healthcare domain [41]. A combination of both methodologies was also possible, for example, integrating methods and models of assessing the quality of Internet services [45].

To build our ontology, we adopted the ontology engineering methodology presented by Uschold and Gruninger [46]. It belongs to the waterfall model group, where the stages of the ontology engineering process were performed sequentially [41]. The methodology was adopted because it provides flexibility to the ontology developers in our team to lead and drive discussions with domain experts. We performed the following stages:

- Identification of the purpose and scope of the ontology. As mentioned in Section 1, we share knowledge about the characteristics of Capsicum with non-expert users. Therefore, we expected that the ontology should cover characteristics of Capsicum identifiable by this group of users.

2. Building the ontology, which covers the ontology capture, coding, and integration with existing ontologies. We identified entities, properties, and data types, including how entities are related to each other. After that, we represented the identified objects using the Resource Description Framework (RDF) [47] and Ontology Web Language (OWL) [48]. For coding the ontology, we actively used the Protégé ontology editor [49]. For integration with existing ontologies, we adopted the terms from the Plant Ontology [8].
3. Evaluation. We evaluated the ontology by using competence questions [46] to carry out reasoning with different characteristics of Capsicum. This evaluation ensures that a list of correct entities can be obtained when a common characteristic is provided. To order the obtained entities as facets, we use a ranking mechanism explained in Section 3.5.
4. Documentation. We generated the documentation of our ontology by using the WIDOCO tool [50]. It generated human-readable descriptions of terms and summaries with integration with other external information.

In the query formulation step, a list of anticipated questions was formulated. We consider entities and properties from each question. To be precise, from each obtained question, we extracted entities, properties, and values from every property. As a result, a query list, represented as tuples (entity, property, and values), is obtained.

3.4. Matching

In this step, the constructed ontology and the obtained list of queries are aligned to identify the matching pairs. As the representation of the degree of understanding, the queries of users are various. Each user has a different level of background knowledge about the subject. Therefore, the matching procedure should be adapted to fit those variations of degrees. We characterize users based on how they define their queries in four groups as follow:

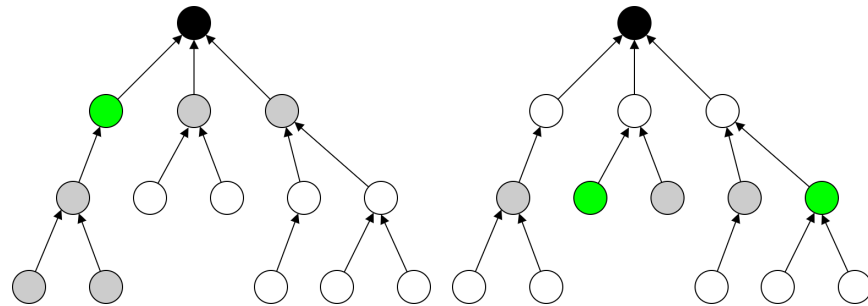
1. Familiar with only one part of the plant. Users in this type only provide the description of a specific part and ignore other generic or more specific parts.
2. Familiar with the generic parts of the plant. Users in this type provide a relatively generic description of the whole plant without focusing on a specific part.
3. Focused on small parts of the plant. Users in this type provide more specific descriptions that are related to each other.
4. Combination of generic and focused. Users in this type provide random descriptions of the plant.

Figure 2 depicts the relevant four types of possible cases when users provide descriptions. The knowledge base is represented as a tree, where the root (color black) is the plant itself and other nodes are parts of it. The relationship (represented as an arrow from one node to another) represents if the lower node is part of the higher node. The nodes with green color are two defined nodes (based on user description), and gray are the related nodes.

1. Matching #1, users describe a plant using only one entity. In this case, we selected the entities at the same level as well as entities in the same branch, as shown in Figure 2a. We called this matching mechanism a single-entity personalization method.
2. Matching #2, users describe a plant using two entities located at the same level in the graph. In this case, we selected entities that are located at the same level, as shown in Figure 2b. We called this matching mechanism a level-based personalization method.
3. Matching #3, users describe a plant using two entities not located at the same level but the same branch. In this case, we selected entities that are located at the same branch, as shown in Figure 2c. We called this matching mechanism a branch-based personalization method.
4. Matching #4, users describe a plant using two entities not located at the same level or the same branch. In this case, we selected entities at the same level from both entities

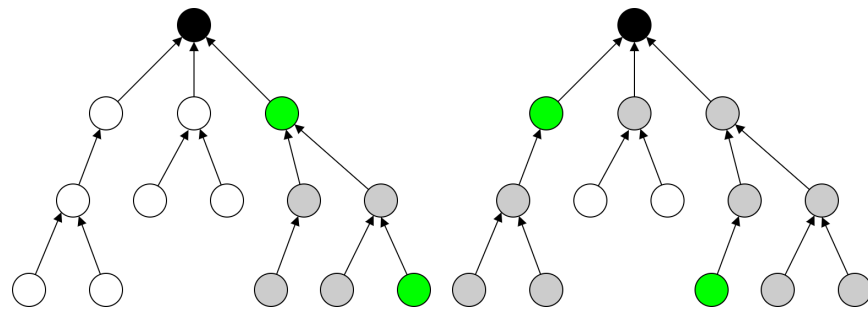
as well as entities from the branches, as shown in Figure 2d. We called this matching mechanism a level-and-branch-based personalization method.

All selected entities were then used as the list of facets to be presented to users. To this point, the order of the facets is random, so it is unpredictable. Therefore, a ranking mechanism is essential to ensure that the identical output is produced when using the identical query. This mechanism is explained further in the following sub-section.



(a) Matching #1. The question contains only one entity, so we consider both level and branch.

(b) Matching #2. The question contains more than one entities and both entities are at the same level.



(c) Matching #3. The question contains more than one entities and both entities are at the same branch.

(d) Matching #4. The question contains more than one entities which are not at the same level, nor at the same branch.

Figure 2. Four matching mechanisms for identifying relevant entities based on the matching of one or more defined entities. The entity with black color is the root, green is the selected nodes, and gray is the identified nodes.

3.5. Ranking

A ranking mechanism is required to anticipate how to deliver multiple results to users. In this step, all obtained entities are ranked to determine their orders according to a specific criterion. As final results, we obtained a list of ordered entities to be used as facets in our faceted search engine. Back to our example in Figure 2, we obtained 4, 3, 4, and 9 facets to be presented to users as shown in Figure 2a–d, respectively. In a graph with complex relationships, the number of facets can be enormous, and therefore, a mechanism to order them is necessary.

Our ranking mechanism utilizes the relationships between entities as well as properties from each entity. We defined three ranking mechanisms to order the obtained facets as follow:

1. Ranking #1: Select the matched entities with similar properties to the provided question. For example, if the query contains the property “Color”, then all entities with “Color” are ordered first.
2. Ranking #2: Select the matched entities with a higher number of properties. An entity with a more detailed description (based on the number of available properties) is ordered first.

- Ranking #3: Select the more generic entity first. Since the generality of entities can be obtained through their levels, a lower-level entity is ordered first.

We expected that the list of facets to be presented to users should be complete and predictable.

3.6. Evaluation

For evaluation, we used a randomized comparator under the assumption that if the proposed system is robust, it should produce an identical list of facets for an identical query regardless of its execution time. In this case, we computed the degree of predictability of produced results. If two facets x and y have an equal chance to be ordered first, then the degree of predictability is low, meaning that the produced results can be randomly ordered. Given F as the list of produced facets and $R \in F$ as the list of facets that was randomly ordered according to the employed ranking mechanisms m , the degree of predictability is computed as follow:

$$P_m = 1 - \frac{|R_m|}{|F|} \tag{1}$$

As shown in Equation (1), a degree predictability P is computed according to the ranking mechanism m . It works by computing the fraction of randomly ordered items R from the total items F . The value of P is between 0, representing fully random (zero predictability) to 1 (high predictability).

4. Result

In this section, we describe the results of our work. We start by discussing the developed ontology, followed by a prototype of our faceted search implementation.

4.1. Ontology

Figure 3 shows the developed ontology to deliver knowledge of Capsicum to non-expert users. In its initial version, the ontology consists of 21 entities, 2 object properties, 11 data type properties, and 4 species as individuals (*Capsicum annuum*, *Capsicum frutescens*, *Capsicum chinense*, and *Capsicum pubescens*). It simplifies the Plant Ontology, where it focused only on the fundamental characteristics that most non-expert users can recognize.

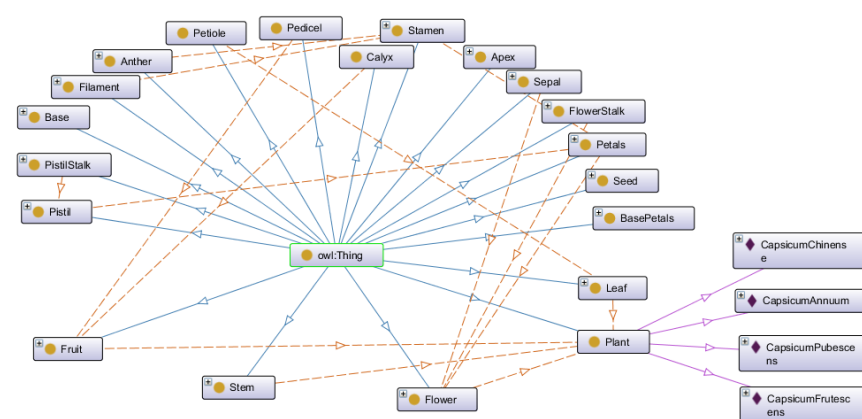


Figure 3. OntoCapsicum, our ontology to share knowledge of Capsicum to non-expert users.

Table 1 shows the list of main entities available in our ontology. They can be arranged into five levels, where level 0 belongs to the root of the plant. Object relationships in our ontology are partOf and its inverse hasPart to represent if an entity is part of another entity.

Table 1. The list of main entites in our ontology and their levels.

Level 0	Level 1	Level 2	Level 3	Level 4
Plant	Stem			
	Leaf	Apex Base Petiole		
	Flower	Flower Stalk Sepal Petals	Base Petals Pistil Stamen	Pistil Stalk Anther Filament
	Fruit	Pedicel Calyx Seed Ripe Fruit Raw Fruit		

Table 2 displays the list of data type properties in our ontology. It consists of a few general characteristics that can be used to describe Capsicum plant. The table also shows the number of entities that are correlated with each relevant property. Property “Color” can be used to describe 13 entities, followed by “Shape” and “Length”, each with 9 and 7 entities, respectively.

Table 2. The list of data type properties in our ontology.

No.	Property	# Entities as Domain
1	Color	13
2	Diameter	1
3	Length	7
4	Number of Seed	1
5	Number of Stalk Segment	1
6	Position	4
7	Shape	9
8	Spot	2
9	Surface	1
10	Texture	2
11	Width	2

Table 3 shows the defined values for several properties in our ontology. We update the list regularly, and the latest version of the ontology is available online (<https://ricover.hpc.lipi.go.id/ontocapsicum/> (accessed on 28 June 2021)).

Table 3. The list of pre-defined values for several properties in our ontology.

Property	Pre-Defined Values
Texture	bare; coarse; hairy; hairless; slippery;
Color	blue; bluish; dark green; green; greenish-white; greenish-yellow; pale green; purple; red; slightly purplish; white; yellow; yellowish;
Shape	bell shape; cuff; elongated; lanceolate; star-like; rounded; rounded eggs; triangular-like;
Surface	smooth;
Position	hanging; fixed; upright;

4.2. Search Algorithm

Algorithm 1 displays the algorithm used to produce an ordered list of facets to be presented to the users to refine their search further. Given G as the graph representation of

the ontology, Q as the query that represented as a list of tuple (entity, property, and value), and R as the ranking mechanisms, its primary goal is to find a list of relevant entities F to be used as facets. The algorithm works by seeking the matching entities, given the graph and the list of queries (line no. 1). The matched entities obtained will determine which matching mechanisms utilized. If the size of matched entities is only 1, then the mechanism #1 is executed (line no. 18). Otherwise, a further determination is executed (line no. 2). Based on the determination for level and branch, the algorithm executes ranking mechanisms #4 or #3 or #2 (line no. 11, 13, and 15, respectively). Finally, after all relevant entities were collected, the ranking mechanism is executed further to ensure the predictability of the results (line no. 23).

Algorithm 1: Capsicum Search Algorithm.

```

input :  $G, Q, R$ 
output:  $F$ 

1 match  $\leftarrow$  FindMatchEntities( $G, Q$ );
2 if Size(match) > 1 then                                     // Use mechanism #2, #3, #4
3   for  $i \leftarrow 0$  to Size(match) do
4     level  $\leftarrow$  CollectSameLevelEntities( $G, \text{match}[i]$ );
5     branch  $\leftarrow$  CollectSameBranchEntities( $G, \text{match}[i]$ );
6      $L = \text{Union}(L, \text{level})$ ;
7      $B = \text{Union}(B, \text{branch})$ ;
8   end
9   sameLevel  $\leftarrow$  CheckSameLevel( $G, \text{match}$ );
10  sameBranch  $\leftarrow$  CheckSameBranch( $G, \text{match}$ );
11  if sameLevel == True And sameBranch == True then           // #4
12    |  $F = \text{Union}(L, B)$ ;
13  else if sameBranch == True then                               // #3
14    |  $F = B$ ;
15  else if sameLevel == True then                               // #2
16    |  $F = L$ ;
17 end
18 else                                                         // Use mechanism #1
19   level  $\leftarrow$  CollectSameLevelEntities( $G, \text{match}[0]$ );
20   branch  $\leftarrow$  CollectSameBranchEntities( $G, \text{match}[0]$ );
21    $F = \text{Union}(\text{level}, \text{branch})$ ;
22 end
23  $F = \text{OrderItems}(F, R)$ 

```

5. Evaluation

The evaluation was conducted by generating a list of possible queries performed by a person and by measuring their effects on produced relevant facets. First, the relevant facets are identified through the matching between queries and the ontology. After that, the facets are ordered according to specific criteria. When a similar query is executed multiple times on a faceted search engine application, the list of produced facets should be the same and ordered consistently. This evaluation calculates how consistent the produced facets when multiple types of queries are applied.

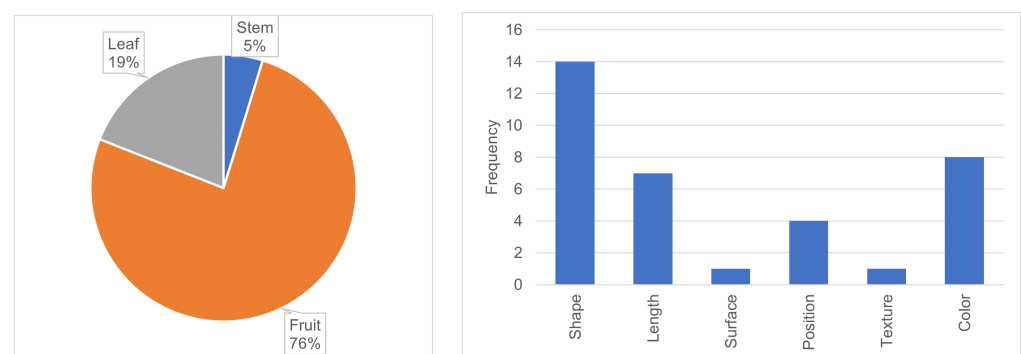
We consulted domain experts and collected a list of search queries that were possibly provided by non-expert users. This consultation is required to ensure that a query can be answered by our matching procedure (in-scope query). We did not consider the out-of-scope queries because the matching procedure does not produce any results for this kind of query. According to our ontology, we identified entities and properties from each query and found that most of the queries consist of one entity and up to three properties. Figure 4 shows the composition of entities and properties in the collected queries. For entity, 76% of them were about "Fruit", followed by 19% and 5% about "Leaf" and "Stem" respectively

as shown in Figure 4b. The composition of properties is shown in Figure 4a, where the most used property was “Shape”, followed by “Color”, “Length”, “Position”, etc.

For evaluation, we execute the algorithm to measure the degree of predictability of the produced results from two main steps in our algorithm, namely matching and ranking, as explained in Section 3.2. Based on the defined four types of matching mechanisms, we investigated a few queries that fit each mechanism as follow:

1. Case 1. The query contains only one node (fit with the matching mechanism #1). Testing case 1 uses permutation of three entities, namely “Fruit”, “Leaf”, and “Stem”. All of them belong to level 1, and they are suitable for case 1.
2. Case 2. The query contains two nodes, where both nodes are at the same level (fit with the matching mechanism #2). Testing case 2 is conducted by using a combination of “Petals”, and “Seed”.
3. Case 3. The query contains two nodes, where both nodes are at the same branch (fit with the matching mechanism #3). Testing case 3 uses a combination of nodes in the same branch with multiples levels from three entities, such as “Fruit”, “Stamen”, and “Flower”.
4. Case 4. The query contains two nodes, where both nodes are neither at the same level nor at the same branch (fit with the matching mechanism #4). Testing case 4 is conducted with a combination of entities as nodes.

Based on the composition of entities and properties extracted from queries as shown in Figure 4, we selected a list of queries representing all four cases described above. As a result, 19 queries were selected that fairly represent all four cases and four individuals that are available in our ontology, as shown in Table 4.



(a) The composition of the identified entities in the available queries

(b) The composition of the identified properties in the available queries

Figure 4. The composition of identified entities and properties from the available queries.

For each query, we ran the algorithm to obtain the list of facet candidates. We collected the candidates after the matching and before a ranking mechanism was applied. Our intention was to identify the best combination of ranking mechanisms providing the highest degree of predictability. These intermediate results are shown in Table 5.

Table 4. List of search queries.

No.	Case	Search queries	Individuals
1	Case 1	Elongated fruit shape	<i>Capsicum frutescens</i>
2	Case 1	Lanceolate leaf shape	<i>Capsicum annuum</i>
3	Case 1	Hairy stems	<i>Capsicum pubescens</i>
4	Case 2	Greenish-yellow petals and yellowish seeds	<i>Capsicum frutescens</i>
5	Case 2	Elongated fruit shape and yellowish seeds	<i>Capsicum frutescens</i>
6	Case 3	Bell fruit shape and yellowish seeds	<i>Capsicum annuum</i>
7	Case 3	Hanging flower position and a few centimeters steam length	<i>Capsicum annuum</i>
8	Case 3	Star-like flower shape and a few millimeters pistil length	<i>Capsicum chinense</i>
9	Case 3	Star-like flower shape and yellow anthers	<i>Capsicum chinense</i>
10	Case 3	Greenish-yellow petals and star-like flower shape	<i>Capsicum chinense</i>
11	Case 4	Elongated fruit shape and greenish-white petals	<i>Capsicum frutescens</i>
12	Case 4	Elongated fruit shape and triangular-like leaf shape	<i>Capsicum frutescens</i>
13	Case 4	Elongated fruit shape and upright fruit position	<i>Capsicum frutescens</i>
14	Case 4	Elongated fruit shape and hanging fruit position	<i>Capsicum annuum</i>
15	Case 4	Star-like flower shape and rounded leaf shape	<i>Capsicum pubescens</i>
16	Case 4	Green leafy ripe fruit and lanceolate leaf shape	<i>Capsicum annuum</i>
17	Case 4	Dark green leaves and white petals	<i>Capsicum annuum</i>
18	Case 4	Smooth leaf surface and bluish anthers	<i>Capsicum chinense</i>
19	Case 4	Light green leaves and yellowish seeds	<i>Capsicum frutescens</i>

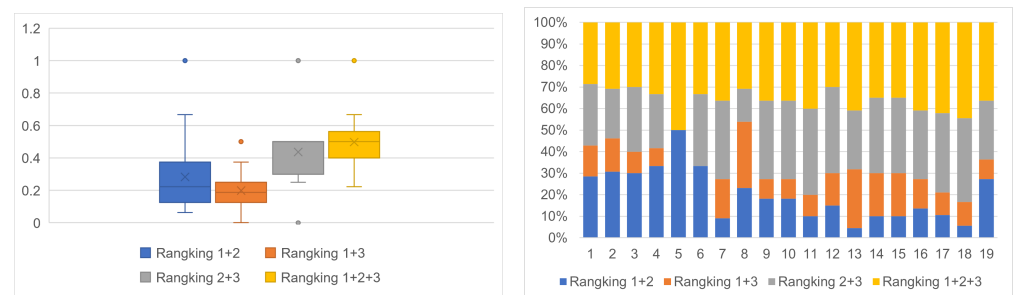
Table 5. Search results after a matching mechanism was applied.

No.	Results after Matching Step
1	Stem, Leaf, Flower, Ripe Fruit, Raw Fruit, Seed, Calyx, Pedicel
2	Stem, Fruit, Flower, Apex, Base, Petiole
3	Leaf, Fruit, Flower
4	Apex, Base, Petiole, Ripe Fruit, Raw Fruit, Seed, Calyx, Pedicel, Sepal, Flower Stalk
5	Apex, Base, Petiole, Ripe Fruit, Raw Fruit, Flower Stalk, Calyx, Pedicel, Petals
6	Ripe Fruit, Raw Fruit, Calyx, Pedicel
7	Sepal, Flower Stalk, Petals, Pistil, Anthers, Filament, Base Petals, Pistil Stalk
8	Sepal, Flower Stalk, Petals, Stamen, Anthers, Filament, Base Petals, Pistil Stalk
9	Sepal, Flower Stalk, Petals, Pistil, Stamen, Filament, Base Petals, Pistil Stalk
10	Sepal, Flower Stalk, Pistil, Stamen, Anthers, Filament, Base Petals, Pistil Stalk
11	Stem, Leaf, Flower, Ripe Fruit, Raw Fruit, Seed, Calyx, Pedicel, Sepal, Flower Stalk, Base Petals, Pistil Stalk, Pistil, Stamen, Anthers, Filament
12	Stem, Leaf, Flower, Ripe Fruit, Raw Fruit, Seed, Calyx, Pedicel, Petals, Flower Stalk, Base Petals, Pistil Stalk, Pistil, Stamen, Anthers, Filament
13	Stem, Leaf, Flower, Ripe Fruit, Raw Fruit, Seed, Calyx, Pedicel, Petals, Flower Stalk, Base Petals, Pistil Stalk, Sepal, Stamen, Anthers, Filament
14	Stem, Leaf, Fruit, Ripe Fruit, Seed, Calyx, Pedicel, Petals, Flower Stalk, Base Petals, Pistil Stalk, Sepal, Stamen, Anthers, Filament, Pistil
15	Stem, Leaf, Fruit, Ripe Fruit, Raw Fruit, Calyx, Pedicel, Petals, Flower Stalk, Base Petals, Pistil Stalk, Sepal, Stamen, Anthers, Filament, Pistil
16	Stem, Leaf, Fruit, Ripe Fruit, Seed, Calyx, Raw Fruit, Petals, Flower Stalk, Base Petals, Pistil Stalk, Sepal, Stamen, Anthers, Filament, Pistil
17	Stem, Fruit, Flower, Apex, Base, Petiole, Sepal, Flower Stalk, Base Petals, Pistil Stalk, Stamen, Anthers, Filament, Pistil
18	Stem, Fruit, Flower, Apex, Base, Petiole, Sepal, Flower Stalk, Base Petals, Pistil Stalk, Stamen, Petals, Filament, Pistil
19	Stem, Fruit, Flower, Apex, Base, Petiole, Ripe Fruit, Raw Fruit, Calyx, Pedicel

Table 5 shows the search results after applying a relevant matching mechanism for every query. The produced results are randomly ordered, and therefore, it is necessary to apply a ranking mechanism to ensure its predictability.

We applied the three ranking mechanisms (based on property similarity, number of properties, and levels) described in Section 3.5 and measured their degrees of predictability. The results are shown in Figure 5. Figure 5a shows the distribution of degree

predictability when using the combination of the three ranking mechanisms. We obtained the highest degree of predictability when combined all three ranking mechanisms (value of mean = 0.49). The second higher degree was obtained by combining ranking mechanisms #2 (number of properties) and #3 (levels) with a mean = 0.44. A combination of ranking mechanism #1 (property similarity) and #2 produced a lower degree of predictability (value of mean = 0.29), followed by a combination of ranking mechanism #1 and #3 with a value of mean = 0.19. Furthermore, a detailed comparison for every query using every possible combination of ranking mechanism is shown in Figure 5b. In most of the queries, a combination of all three ranking mechanisms and a combination of ranking mechanism #2 and #3 are superior over the others. Surprisingly, the combination of ranking mechanism #1 and #2 was performed well when the number of matched entities is low, as shown in the results for questions no. 1 to 10.



(a) Degree predictability for each ranking mechanism (b) Detailed degree of predictability for each query

Figure 5. Degree predictability for combination of ranking mechanisms and for each query.

Finally, we summarize a few things from the results explained above. First, the simplification of the complex ontology is critical to ensure that the ontology can fit targeted users. We used a simplified version of the Plant Ontology to allow for a knowledge-sharing process about morphological characteristics of Capsicum species to non-expert users. Second, the alignment between the constructed ontology and queries is possible through multiple matching mechanisms. All queries can be aligned into one of four available matching mechanisms. Third, a combination of multiple ranking mechanisms is necessary to increase the degree of predictability of produced results. There is no single ranking mechanism that is more important than the other. Overall, we demonstrated that the proposed solution can be implemented effectively with promising results.

6. Conclusions

In general, a Knowledge Base (KB) can be seen as a technology that captures and stores knowledge from human experts. A Knowledge-Based System (KBS) uses a KB to make intelligent decisions, for example, to support decision-making processes, learning, and other activities. In most cases, a KB is highly domain-oriented, based on the expertise of experts in a specific domain. It is necessary to have sufficient knowledge about the field to be able to consume a KB properly.

A challenge occurs when knowledge-sharing from a KB with persons who are not familiar with the domain, for example, citizen scientists. The involvement of citizen scientists has been recognized as an essential factor in science explorations. Their involvement is in collecting data and at the same time in consuming data, for example, to identify species. An approach for knowledge-sharing is through an ontology that can be seen as a shared common understanding about a domain. Using a common ontology for a variety of knowledge-sharing activities would guarantee consistency. In the domain of plant science, several ontologies have been developed to support knowledge-sharing among scientists. They were intended to be consumed by scientists who have knowledge about the domain but not by users who have less or even no knowledge about it.

In this paper, we tackled the challenge using an ontology-driven knowledge-sharing approach that is consumable by non-expert users. Our work started by eliciting knowledge from experts and formulated them in an ontology. The ontology was developed by focusing on generic characteristics and left out detailed information. Furthermore, we developed a faceted search application to consume the ontology and to provide relevant facets that users can use to refine their queries further. In this way, non-expert users should explore a KB about Capsicum with minimal background knowledge.

We characterized users into four groups based on their ways to describe objects in a KB, in our case, parts of a plant. As a result, we introduced four matching mechanisms to identify relevant entities in our ontology and three ranking mechanisms to order the identified entities. To evaluate our search system, we measured the degree of predictability of produced facets. Given an identical query, the system should provide an equal output. We found that different types of queries can be mapped into one of the available matching mechanisms. Furthermore, by combining all three ranking mechanisms, we obtained the highest degree of predictability of the list of facets.

In the future, we expand our current faceted search application vertically by using more sophisticated ontology and horizontally by using ontology from different areas. Specifically, we use a color index to avoid ambiguity in defining the value for the entity “Color”. In its current implementation, we use pre-defined values such as “Dark Green” where every person might have a different definition. Additionally, we standardize values for specific entities. Values for the entity “Shape” such as “rounded” and “bell shape”, should be standardized to avoid multiple interpretations. We are also considering transforming our ontology into a fuzzy ontology because fuzzy logic has suitable formalisms to handle imprecise and uncertain knowledge [51]. We envisioned that our approach has an essential impact on biodiversity science in general, where many KBs have been developed but they are difficult to be consumed by users outside of the domain. Further, non-expert users such as citizen scientists contribute significantly to knowledge-sharing and dissemination of science to the public.

Author Contributions: Conceptualization: Z.A. and D.S.R.; data curation: H.F.M., L.P.M. and A.I.; investigation: H.F.M., A.D.F. and T.D.; methodology: Z.A. and D.S.R.; resources: A.D.F. and T.D.; writing—original draft preparation: Z.A., H.F.M., D.S.R., L.P.M. and A.I.; writing—review and editing: Z.A. and D.S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Research and Technology of the Republic of Indonesia (Ristek)—Managing Agency of Endowment Fund for Education of the Republic of Indonesia (LPDP) with the contract number: 39/EI/PRN/2020.

Data Availability Statement: Not Applicable, the study does not report any data.

Acknowledgments: We thank all members of the Knowledge Engineering Research Group at the Research Center for Informatics and the Plant Physiology Research Group, Plant Systematic Research Group, and Plant Ecology Research Group from the Research Center for Biology, Indonesian Institute of Sciences, for their valuable feedback.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Farley, S.S.; Dawson, A.; Goring, S.J.; Williams, J.W. Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience* **2018**, *68*, 563–576. [[CrossRef](#)]
2. König, C.; Weigelt, P.; Schrader, J.; Taylor, A.; Kattge, J.; Kreft, H. Biodiversity data integration—The significance of data resolution and domain. *PLoS Biol.* **2019**, *17*, e3000183. [[CrossRef](#)] [[PubMed](#)]
3. Bell, G.; Hey, T.; Szalay, A. Beyond the Data Deluge. *Science* **2009**, *323*, 1297–1298. [[CrossRef](#)] [[PubMed](#)]
4. Das Gupta, M.; Tsiantis, M. Gene networks and the evolution of plant morphology. *Curr. Opin. Plant Biol.* **2018**, *45*, 82–87. [[CrossRef](#)]
5. Antonio, A.S.; Wiedemann, L.S.M.; Veiga Junior, V.F. The genus Capsicum: A phytochemical review of bioactive secondary metabolites. *RSC Adv.* **2018**, *8*, 25767–25784. [[CrossRef](#)]

6. Peter, M.; Diekötter, T.; Kremer, K. Participant Outcomes of Biodiversity Citizen Science Projects: A Systematic Literature Review. *Sustainability* **2019**, *11*, 2780. [[CrossRef](#)]
7. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum. Comput. Stud.* **1995**, *43*, 907–928. [[CrossRef](#)]
8. Cooper, L.; Walls, R.L.; Elser, J.; Gandolfo, M.A.; Stevenson, D.W.; Smith, B.; Preece, J.; Athreya, B.; Mungall, C.J.; Rensing, S.; et al. The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* **2013**, *54*, e1. [[CrossRef](#)]
9. Consortium, T.G.O. The Gene Ontology project in 2008. *Nucleic Acids Res.* **2007**, *36*, D440–D444. [[CrossRef](#)]
10. Walls, R.L.; Cooper, L.; Elser, J.; Gandolfo, M.A.; Mungall, C.J.; Smith, B.; Stevenson, D.W.; Jaiswal, P. The Plant Ontology Facilitates Comparisons of Plant Development Stages Across Species. *Front. Plant Sci.* **2019**, *10*, 631. [[CrossRef](#)]
11. Burgess, H.; DeBey, L.; Froehlich, H.; Schmidt, N.; Theobald, E.; Ettinger, A.; HilleRisLambers, J.; Tewksbury, J.; Parrish, J. The science of citizen science: Exploring barriers to use as a primary research tool. *Biol. Conserv.* **2017**, *208*, 113–120. [[CrossRef](#)]
12. Mahdi, M.N.; Ahmad, A.R.; Ismail, R.; Natiq, H.; Mohammed, M.A. Solution for Information Overload Using Faceted Search—A Review. *IEEE Access* **2020**, *8*, 119554–119585. [[CrossRef](#)]
13. Lens, F.; Cooper, L.; Gandolfo, M.A.; Groover4, A.; Jaiswal, P.; Lachenbruch, B.; Spicer, R.; Staton, M.E.; Stevenson, D.W.; Walls, R.L.; et al. An extension of the Plant Ontology project supporting wood anatomy and development research. *IAWA J.* **2012**, *33*, 113–117. [[CrossRef](#)]
14. Meng, X.; Xu, C.; Liu, X.; Bai, J.; Zheng, W.; Chang, H.; Chen, Z. An Ontology-Underpinned Emergency Response System for Water Pollution Accidents. *Sustainability* **2018**, *10*, 546. [[CrossRef](#)]
15. Padilla-Cuevas, J.; Reyes-Ortiz, J.A.; Bravo, M. Ontology-Based Context Event Representation, Reasoning, and Enhancing in Academic Environments. *Future Internet* **2021**, *13*, 151. [[CrossRef](#)]
16. Ziembra, P.; Wątróbski, J.; Jankowski, J.; Wolski, W. Construction and Restructuring of the Knowledge Repository of Website Evaluation Methods. In *Information Technology for Management*; Ziembra, E., Ed.; Springer International Publishing: Cham, Switzerland, 2016; pp. 29–52. [[CrossRef](#)]
17. Avraham, S.; Tung, C.W.; Ilic, K.; Jaiswal, P.; Kellogg, E.A.; McCouch, S.; Pujar, A.; Reiser, L.; Rhee, S.Y.; Sachs, M.M.; et al. The Plant Ontology Database: A community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res.* **2008**, *36*, D449–D454. [[CrossRef](#)]
18. Arnaud, E.; Cooper, L.; Shrestha, R.; Menda, N.; Nelson, R.T.; Matteis, L.; Skofic, M.; Bastow, R.; Jaiswal, P.; Mueller, L.; et al. Towards a Reference Plant Trait Ontology for Modeling Knowledge of Plant Traits and Phenotypes. In Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain, 4–7 October 2012.
19. Akbar, Z.; Kartika, Y.A.; Ridwan Saleh, D.; Mustika, H.F.; Parningotan Manik, L. On Using Declarative Generation Rules To Deliver Linked Biodiversity Data. In Proceedings of the 2020 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Tangerang, Indonesia, 18–20 November 2020; pp. 267–272. [[CrossRef](#)]
20. McLaren, C.G.; Bruskiwich, R.M.; Portugal, A.M.; Cosico, A.B. The International Rice Information System. A Platform for Meta-Analysis of Rice Crop Data. *Plant Physiol.* **2005**, *139*, 637–642. [[CrossRef](#)]
21. Schomburg, I.; Jeske, L.; Ulbrich, M.; Placzek, S.; Chang, A.; Schomburg, D. The BRENDA enzyme information system—From a database to an expert system. *J. Biotechnol.* **2017**, *261*, 194–206. [[CrossRef](#)]
22. Kim, H.J.; Baek, K.H.; Lee, S.W.; Kim, J.; Lee, B.W.; Cho, H.S.; Kim, W.T.; Choi, D.; Hur, C.G. Pepper EST database: Comprehensive in silico tool for analyzing the chili pepper (*Capsicum annuum*) transcriptome. *BMC Plant Biol.* **2008**, *8*, 101. [[CrossRef](#)]
23. Silalahi, M.; Cahyani, D.E.; Sensuse, D.I.; Budi, I. Developing Indonesian medicinal plant ontology using socio-technical approach. In Proceedings of the 2015 International Conference on Computer, Communications, and Control Technology (I4CT), Kuching, Malaysia, 21–23 April 2015; pp. 39–43. [[CrossRef](#)]
24. Kaewboonma, N.; Supnithi, T.; Panawong, J. Developing ontology for Thai Zingiberaceae: From taxonomies to ontologies. In Proceedings of the 2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 27–30 June 2017; pp. 596–599. [[CrossRef](#)]
25. Stoyanova-Doycheva, A.; Ivanova, V.; Doychev, E.; Spassova, K. Development of an Ontology in Plant Genetic Resources. In Proceedings of the 2020 IEEE 10th International Conference on Intelligent Systems (IS), Varna, Bulgaria, 28–30 August 2020; pp. 246–251. [[CrossRef](#)]
26. Ilic, K.; Kellogg, E.A.; Jaiswal, P.; Zapata, F.; Stevens, P.F.; Vincent, L.P.; Avraham, S.; Reiser, L.; Pujar, A.; Sachs, M.M.; et al. The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant. *Plant Physiol.* **2006**, *143*, 587–599. [[CrossRef](#)]
27. Hoehndorf, R.; Alshahrani, M.; Gkoutos, G.V.; Gosline, G.; Groom, Q.; Hamann, T.; Kattge, J.; de Oliveira, S.M.; Schmidt, M.; Sierra, S.; et al. The flora phenotype ontology (FLOPO): Tool for integrating morphological traits and phenotypes of vascular plants. *J. Biomed. Semant.* **2016**, *7*, 65. [[CrossRef](#)]
28. Simonini, G.; Zhu, S. Big data exploration with faceted browsing. In Proceedings of the 2015 International Conference on High Performance Computing Simulation (HPCS), Amsterdam, The Netherlands, 20–24 July 2015; pp. 541–544. [[CrossRef](#)]
29. Roy, S.B.; Wang, H.; Nambiar, U.; Das, G.; Mohania, M. DynaCet: Building Dynamic Faceted Search Systems over Databases. In Proceedings of the 2009 IEEE 25th International Conference on Data Engineering, Shanghai, China, 29 March–2 April 2009; pp. 1463–1466. [[CrossRef](#)]

30. Chakraborty, T.; Krishna, A.; Singh, M.; Ganguly, N.; Goyal, P.; Mukherjee, A. FeRoSA: A Faceted Recommendation System for Scientific Articles. In *Advances in Knowledge Discovery and Data Mining*; Bailey, J., Khan, L., Washio, T., Dobbie, G., Huang, J.Z., Wang, R., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 528–541. [[CrossRef](#)]
31. Wongsuphasawat, K.; Moritz, D.; Anand, A.; Mackinlay, J.; Howe, B.; Heer, J. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 649–658. [[CrossRef](#)]
32. Farazi, F.; Chapman, C.; Raju, P.; Melville, L. Ontology-based faceted semantic search with automatic sense disambiguation for bioenergy domain. *Int. J. Big Data Intell.* **2018**, *5*, 62–72. [[CrossRef](#)]
33. De Maio, C.; Fenza, G.; Loia, V.; Parente, M. Biomedical data integration and ontology-driven multi-facets visualization. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015; pp. 1–8. [[CrossRef](#)]
34. Sánchez-Cervantes, J.L.; Colombo-Mendoza, L.O.; Alor-Hernández, G.; García-Alcaráz, J.L.; Álvarez-Rodríguez, J.M.; Rodríguez-González, A. LINDASearch: A faceted search system for linked open datasets. *Wirel. Netw.* **2020**, *26*, 5645–5663. [[CrossRef](#)]
35. Mauro, N.; Ardissono, L.; Lucenteforte, M. Faceted search of heterogeneous geographic information for dynamic map projection. *Inf. Process. Manag.* **2020**, *57*, 102257. [[CrossRef](#)]
36. Thadeu Ferreira da Silva, S.; Apolonio, S.d.O.; Vivacqua, A.S.; Oliveira, J.; Xexéo, G.B.; Campos, M.L.M. Ontoogle: Enhancing retrieval with ontologies and facets. In Proceedings of the 2011 15th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Laussane, Switzerland, 8–10 June 2011; pp. 192–199. [[CrossRef](#)]
37. Le, T.; Vo, B.; Duong, T.H. Personalized Facets for Semantic Search Using Linked Open Data with Social Networks. In Proceedings of the 2012 Third International Conference on Innovations in Bio-Inspired Computing and Applications, Kaohsiung, Taiwan, 26–28 September 2012; pp. 312–317. [[CrossRef](#)]
38. Niu, X.; Fan, X.; Zhang, T. Understanding Faceted Search from Data Science and Human Factor Perspectives. *ACM Trans. Inf. Syst.* **2019**, *37*, 1–27. [[CrossRef](#)]
39. Bondy, J.; Murty, U. *Graph Theory*, 1st ed.; Springer Publishing Company, Incorporated: London, UK, 2008.
40. Suárez-Figueroa, M.C.; Gómez-Pérez, A.; Fernández-López, M. The NeOn Methodology framework: A scenario-based methodology for ontology development. *Appl. Ontol.* **2015**, *10*, 107–145. [[CrossRef](#)]
41. Spoladore, D.; Pessot, E. Collaborative Ontology Engineering Methodologies for the Development of Decision Support Systems: Case Studies in the Healthcare Domain. *Electronics* **2021**, *10*, 1060. [[CrossRef](#)]
42. Fernández-López, M.; Gómez-Pérez, A.; Juristo, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series, Palo Alto, CA, USA, 24–26 March 1997.
43. Corcho, O.; Fernández-López, M.; Gómez-Pérez, A.; López-Cima, A. Building Legal Ontologies with METHONTOLOGY and WebODE. In *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications*; Benjamins, V.R., Casanovas, P., Breuker, J., Gangemi, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 142–157. [[CrossRef](#)]
44. Villazón-Terrazas, B.; Ramírez, J.; Suárez-Figueroa, M.C.; Gómez-Pérez, A. A network of ontology networks for building e-employment advanced systems. *Expert Syst. Appl.* **2011**, *38*, 13612–13624. [[CrossRef](#)]
45. Ziemba, P.; Jankowski, J.; Wątróbski, J.; Becker, J. Knowledge Management in Website Quality Evaluation Domain. In *Computational Collective Intelligence*; Núñez, M., Nguyen, N.T., Camacho, D., Trawiński, B., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 75–85. [[CrossRef](#)]
46. Uschold, M.; Gruninger, M. Ontologies: Principles, methods and applications. *Knowl. Eng. Rev.* **1996**, *11*, 93–136. [[CrossRef](#)]
47. Pan, J.Z. Resource Description Framework. In *Handbook on Ontologies*; Staab, S., Studer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 71–90. [[CrossRef](#)]
48. Antoniou, G.; van Harmelen, F. Web Ontology Language: OWL. In *Handbook on Ontologies*; Staab, S., Studer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 67–92. [[CrossRef](#)]
49. Musen, M.A. The Protégé Project: A Look Back and a Look Forward. *AI Matters* **2015**, *1*, 4–12. [[CrossRef](#)] [[PubMed](#)]
50. Garijo, D. WIDOCO: A Wizard for Documenting Ontologies. In *The Semantic Web—ISWC 2017*; d’Amato, C., Fernandez, M., Tamma, V., Lecue, F., Cudré-Mauroux, P., Sequeda, J., Lange, C., Heflin, J., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 94–102. [[CrossRef](#)]
51. Díaz Rodríguez, N.; Cuéllar, M.P.; Lilius, J.; Delgado Calvo-Flores, M. A fuzzy ontology for semantic modelling and recognition of human behaviour. *Knowl. Based Syst.* **2014**, *66*, 46–60. [[CrossRef](#)]