

Ontology Merging: A Practical Perspective

Niladri Chatterjee^(✉), Neha Kaushik, Deepali Gupta,
and Ramneek Bhatia

Department of Mathematics, Indian Institute of Technology Delhi,
Hauz Khas, New Delhi 110016, India
niladri.iitd@gmail.com, swami.neha@gmail.com,
deepaligupta0737@gmail.com, ramneek1995@gmail.com

Abstract. Digitization of data has now come in a big way in almost every possible aspects of modern life. Agriculture as a domain is no exception. But digitization alone does not suffice, efficient retrievability of the information has to be ensured for providing web services including question-answering. However, building an ontology for a vast domain as a whole is not straightforward. We view creation of an ontology as an incremental process, where small-scale ontologies for different sub-domains are expected to be developed independently, to be merged into a single ontology for the domain. The paper aims at designing a framework for ontology merging. The method is described with agriculture as the primary domain with several subdomains such as crop, fertilizer, as subdomains among others. The supremacy of the scheme over Protégé, a well-known ontology management software is demonstrated.

Keywords: Ontology · Ontology merging · Protégé

1 Introduction

With the advent of more data-centric applications, ontologies as a knowledge representation technique have gained much popularity in the last one decade or so. Ontologies allow creation of annotations in which information is organized as a machine readable and machine understandable content. An ontology [1], by definition, is explicit specification of

- Concepts (classes) in a domain, e.g. Crop, Soil
- Relationships that exist between concepts, e.g. grows_in (Crop, Soil) gives information on which Crop grows well in which type of Soil.
- Attributes (also called as roles, properties or slots) of the concepts, e.g. SowingTime (for Crop), Moisture_Content (for Soil)
- Instances, e.g., Broccoli (for Crop), Loamy (for Soil)

Ontology development is rapidly growing to facilitate reuse of knowledge. Many applications, such as information retrieval, question-answering, document retrieval, text summarization, are carried out efficiently using the domain ontologies. Domain-specific ontology development has taken up a fast speed and can result in a

gamut of ontologies in a particular domain. These ontologies are needed for various web services including query-answering among others.

Ontology development is sprouting in agricultural domain also. Agriculture is a vast domain consisting of many subdomains, with varying terms being used across regions and with time. It is difficult to build an ontology for such a domain at one go. Hence, a practical way is to build the ontologies incrementally in various subdomains and then merge them for deriving results. Figure 2 shows 12 core subdomains of Indian agriculture.

[2, 3] outline the need for designing an agricultural ontology. End user queries involve information retrieval from different subdomains. One example query is “*Which fertilizer is good for wheat crop?*”

Ontologies aid in efficient query processing and information retrieval. Merging of similar and cross-/overlapping- domain ontologies is required to effectively solve the purpose of ontologies in agriculture. Distributed and heterogeneous ontologies should be inter-related to make them interoperable. Various operations to inter-relate two ontologies O1 and O2 are: *merging, mapping, alignment, refinement, unification, and integration* [4].

Merging means “coming together (the act of joining together as one)”. In merging, two original ontologies, O1 and O2, are joined together to create a single merged ontology. The original ontologies cover similar or overlapping (sub) domains. For example, with respect to agriculture, both O1 and O2 may belong to crops subdomain. However, ontologies of overlapping subdomains such as fertilizers and crops can also be merged.

Alignment in ontology stands for creating links between O1 and O2. Ontology alignment aims to achieve consistency between O1 and O2. It does not unite the two ontologies into one. Ontology alignment is carried out between ontologies of the complementary domains. For example, in agricultural domain, one may choose to have O1 and O2 from soil and crops domain respectively, keeping them separate, still serving to answer queries like which crop grows well in which soil.

The present paper focuses on merging of agricultural ontologies belonging to different subdomains such as crops, fertilizers, and soil. Merging can be performed only after accurately aligning the concepts of the source ontologies.

Section 2 outlines various tools and methods available for ontology merging. Section 3 presents the motivation behind this research work. Section 4 presents the proposed scheme. A review and analysis of the proposed scheme is presented in Sect. 5. Section 6 concludes the paper with some future directions.

2 Literature Survey

Many algorithms and tools for ontology merging have been worked upon. It creates a problem for naïve researchers in the field as they face exorbitant text relating to ontology merging tools and methods. [4–6] provide a comprehensive survey of ontology merging and alignment methods and tools.

One of the earliest tools in ontology merging is SMART [7]. It identifies linguistically similar classes and creates a list of initial linguistic similarity based on

class-name similarity. Examples of linguistic similarity measures used are synonym, common suffix and prefix, and shared sub-string etc. It is a semi-automated tool and generates suggestions for matches, users need to validate those suggestions.

AnchorPROMPT [8] is based on graph structure of ontologies. It traverses the path between related term-pairs, called as anchors in [8], in the source ontologies and identifies the similar terms along this path. Using this information, AnchorPROMPT finds new anchors.

PROMPT [9] is an ontology management tool. It facilitates ontology merging, alignment, and versioning. PROMPT provides merging suggestions to the user. These merging suggestions are based on linguistic and structural knowledge. PROMPT also presents aftereffects of applying these merging suggestions to the ontology.

SMART, anchorPROMPT and PROMPT are developed as a plugin for Protégé-2000¹. PROMPT is a popularly used tool for ontology merging.

A knowledge based translator, named OntoMorph, to facilitate ontology merging is presented in [10]. OntoMorph specifies mappings in the form of rule language. It uses both, syntactic and semantic rewriting. Syntactic rewriting uses pattern matching and works with sentence level transformations. Semantic rewriting uses logical inference on semantic models.

HICAL (Hierarchical Concept Alignment System) employs machine learning techniques for alignment of concept hierarchies [11]. It infers mappings from the overlap of data instances between two taxonomies.

CMS (Crosi Mapping System) uses semantics of the OWL constructs for structure matching in ontology alignment [12]. FCA-Merge [13] is based on bottom-up ontology merging. A merged concept lattice is obtained using formal concept analysis and application-specific ontology instances (belonging to the ontologies which are to be merged) which is converted to a merged ontology by human intervention. Chimaera [14] is an interactive tool for ontology merging. It assists the users for ontology editing, merging and testing.

Protégé, an ontology editing environment, also provides automatic ontology merging service among other options. It provides GUI based ontology merging.

Ontology Alignment Evaluation Initiative (OAEI)² is a standard platform for evaluating ontology merging/matching/mapping/alignment tools. OAEI aims at improving ontology matchers by assessing their weaknesses and strengths. It also provides comparison of various matchers. OAEI benchmark datasets available for 2016 campaign³ are *benchmark*, *anatomy*, *conference*, *multiform*, *interactive machine evaluation*, *large biomedical ontologies*, *disease and phenotype*, *process model matching*, and *instance matching*. Agriculture domain is not present in these benchmark datasets yet, we look forward to test the system developed during current work in OAEI campaigns.

Availability of so many tools and methods for ontology merging proves to be a motivation for identifying an optimal solution for merging cross-domain agricultural ontologies.

¹ <http://protege.stanford.edu>.

² <http://oaei.ontologymatching.org/>.

³ <http://oaei.ontologymatching.org/2016/>.

3 Motivation

There was a need to develop a new algorithm and tool for merging as none of the tools and research work highlighted in the previous section are currently functional. This is attributed primarily to the fact that most of the tools for ontology merging are developed as part of some research activity. On an average, the research goes on for 3–4 years on a specific target [15]. Thereafter, the tools/plugins get outdated and hence discarded. One such example is PROMPT⁴ which is a pioneer work in ontology merging. The website shows PROMPT 3.0 as the last release of the plugin which is compatible with Protégé3.3.1. The plugin is not available for download now (last checked on 20 January 2017). It has been 10 years since the website was updated. The Refactor merging tool in Protégé does not perform merging correctly and suffers from various problems, which the proposed scheme aims to overcome, as discussed further in Sect. 5.

Another motivation is to meet the demand of web services. Although knowledge resources like AGROVOC⁵, NAL thesaurus⁶, Agropedia⁷ are available for agricultural domain, these do not fulfil the requirements of answering queries by common users and suffer from some limitations. AGROVOC is a vast thesaurus and hence contains many terms which are not relevant from farmers' perspective of agricultural domain. Examples are *curriculum*, *indigenous knowledge*, *computer software*, *vocabulary* etc. It also lacks some relevant agricultural terms, e.g., *coriander* (an Indian herb), *neutral fertilizer* (type of fertilizer), *straight fertilizer* (type of fertilizer), *complex fertilizer* (type of fertilizer) etc. Target users of NAL thesaurus are agricultural researchers as it contains too scientific terms. For example, NAL thesaurus provides two options for *search type- terms contain text* and *terms begin with text*. When we searched the term *coriander* with both options, NAL thesaurus displayed thousands of results with none matching the word “*coriander*”. Agropedia provides knowledge models for few crops as image and pdf form. These knowledge models are created by teams of domain experts with great efforts. Ontology creation for agricultural domain is still in its infancy and requires attention. WordNet can also be used to aid the merging process.

4 The Proposed Scheme

The following algorithm (Fig. 1) illustrates the proposed scheme of ontology merging. The algorithm works with ‘n’ number of input ontologies (in the form of owl files) and gives a final merged ontology as output. The algorithm makes use of element-level as well as structure level matching techniques for merging the concepts, instances and relations (as explained in Sect. 1) of the input ontologies.

⁴ <http://protegewiki.stanford.edu/wiki/PROMPT>).

⁵ <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>.

⁶ <http://agclass.nal.usda.gov/>.

⁷ <http://agropedia.iitk.ac.in/>.

1. Consider owl ontologies O_a , $1 \leq a \leq n$, n is the number of ontologies to be merged. O_M is the final merged ontology.
2. Parse owl files to extract the set of concepts and relations in O_a , $1 \leq a \leq n$.
3. Use string based and informal resource based matching techniques to match the similar concepts in O_a and O_b , $1 \leq a \leq n$, $1 \leq b \leq n$.
4. Resolve the matching concepts C_{a_i} and C_{b_j} into C_{m_k} in O_M , where $1 \leq i, j \leq n_m$, where n_m is the no. of concepts in m^{th} ontology
5. Find the matching relation pairs in O_a , $1 \leq a \leq n$, using structure-level matching techniques.
6. Resolve the matching relations of all the ontologies.
7. Find out the related concepts and their relations among the unresolved concepts and relations using relevant portions of the text. This is done using NLP techniques and statistical analysis.
8. Merge the related concepts into O_M .
9. Copy the remaining concepts and relations from O_a , $1 \leq a \leq n$, into O_M .
10. Generate the owl file of the final merged ontology using an Ontology Editor.

Fig. 1. The proposed algorithm

The proposed scheme applies basic alignment schemes like prefix and suffix matching, edit distance, n-grams (3-grams in particular), tokenization and lemmatization and common knowledge thesauri (WordNet). After applying these techniques on the concepts of ontologies, the alignment scores between different pairs of concepts, normalized between 0 and 1 are obtained. A value closer to 1 suggests better alignment. These techniques are then combined after analysing the scores obtained using different matching techniques. It is observed that WordNet shows almost all words as similar because they belong to the same domain- agriculture. Hence WordNet similarity is used only as threshold maintained at 0.9. The degree of similarity is calculated by assigning equal weight to other techniques and checking with a threshold value of 1.2. Duplication of aligned classes is removed and symmetricity of alignment is ensured, i.e., the order in which the sources ontologies are loaded in the tool should not matter.

Element-level techniques do not suffice for merging ontologies. For example, *WeatherCondition* in *Ontology1* (Fig. 3) is very similar to *EnvironmentalFactor* of *Ontology2* (Fig. 4). However, this is not apparent by any element-level technique. *Type* as sub-class of *Soil* in *Ontology1* is merged with *Type* (sub-class of *fertilizer*) in *Ontology2*. This should not be happening as *Type* holds different meanings in the two domains. Structure-level techniques are used to deal with such issues by:

- Addition of Concepts- All non-aligned concepts are aligned only if they have a large co-topic similarity and a substantial number of common children concepts
- Removal of Concepts- Alignments obtained using element-level techniques are removed if the concepts have very less co-topic similarity and non-matching children concepts.

These techniques are first applied on concepts of the source ontologies and subsequently on the properties and relations of the concepts in the source ontologies.

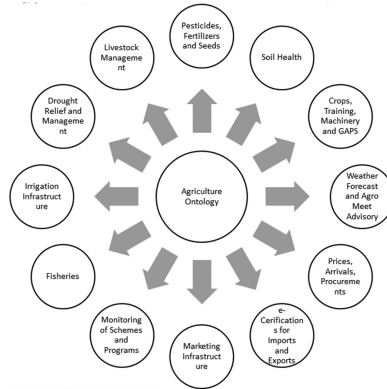


Fig. 2. Subdomains of Indian agriculture, Source: [16, 17]

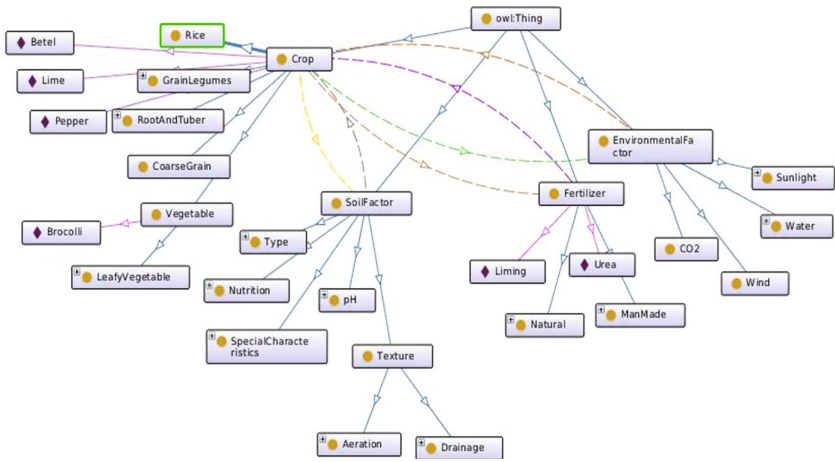


Fig. 3. Graphical representation of *Ontology1*: Source Ontology

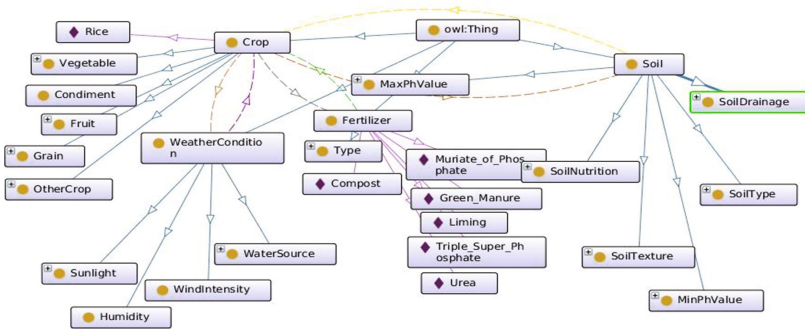


Fig. 4. Graphical representation of *Ontology2*: Source Ontology

5 Results and Analysis

In this section, we present the details of results obtained with ontology merging using Protégé and the proposed scheme. A lot of merging tools and methods have already been worked upon as outlined in Sect. 2. Protégé, a well-known system for ontology management is also bundled with an ontology merging feature.

We have examined the performance for merging two ontologies *Ontology1* (Fig. 3) and *Ontology2* (Fig. 4). These ontologies have been generated from agricultural text available over websites, such as, agricoop.nic.in, farmer.gov.in, agrifarming.in to name a few. The algorithm presented in [18] has been used to extract terms and relationships for construction of ontologies. The extracted terms and relationships are then fed to Protégé to generate owl files of the ontologies. The graphical view of ontologies presented here is generated using OntoGraf plugin of Protégé. Table 1 explains the representation of ontologies using OntoGraf with examples from *Ontology1* shown in Fig. 3. Same representation scheme has been followed for *Ontology2*, and also for *Ontology3*, and *Ontology4* which are discussed below.

Table 1. Representation of ontologies using OntoGraf

Meaning	Example from <i>ontology1</i> in Fig. 3	Type of element
Ontology concept	Crop, SoilFactor, Fertilizer, Vegetable, Rice, Wind	Rectangular node having yellow dot with the label
Instances in the ontology	Pepper, Lime, Broccoli, Urea	Rectangular node having purple diamond with the label
Hierarchical relations between concepts	Leafy Vegetable is a subclass of Vegetable, Vegetable is a subclass of Crop	Directed solid links
Domain-specific relations between the concepts	Purple dotted line represents the relation works_on(Fertilizer, Crop)	Directed dotted links

These ontologies are created from the domains of Crop, Weather, Fertilizer and Soil. These are created in a way so as to enable checking of merging scheme for removing duplicates, alignment accuracy, detection of matching concepts with dissimilar names, etc. The proposed scheme has been implemented using Python, Owlready⁸ library to extract the concepts, object properties and data properties from the source ontologies. Similarity techniques (as discussed in Sect. 4) are implemented in Python to find the alignments between the extracted elements of the ontologies. After applying the alignments obtained, owl file for the final merged ontology is also

⁸ <https://pypi.python.org/pypi/Owlready>.

Table 2. Metrics of *Ontologies*

	<i>Ontology1</i>	<i>Ontology2</i>	<i>Ontology3</i>	<i>Ontology4</i>
No. of concepts	22	22	44	29
No. of object properties	6	6	12	6
No. of data properties	7	8	15	8
No. of instances	41	43	84	56

exported to *Ontology4.owl* using Owlready in Python. Table 2 shows some essential metrics of the two source ontologies and resultant ontologies.

It can be seen in Fig. 5 that Protégé is just inserting the concepts of one ontology into another. Thus, the resulting ontology’s metrics are simply a sum of the source ontologies’ metrics. Also, it results in duplication of concepts as can be seen for *crop*, *sunlight*, *fertilizer*, etc. In *Ontology1* and *Ontology2*, *crop* is a common concept. It is expected that it is considered as one concept in *merged Ontology*. However, Protégé shows *crop* (*Ontology1*) and all its subclasses as different from *crop* (*Ontology2*) in the *Ontology3* (Fig. 5). This tool does not link the concepts which are expected to play a similar role on a structural level. It is also to be noted that Protégé does not merge the relations and instances of concepts in the source ontologies. This is evidenced in Fig. 5.

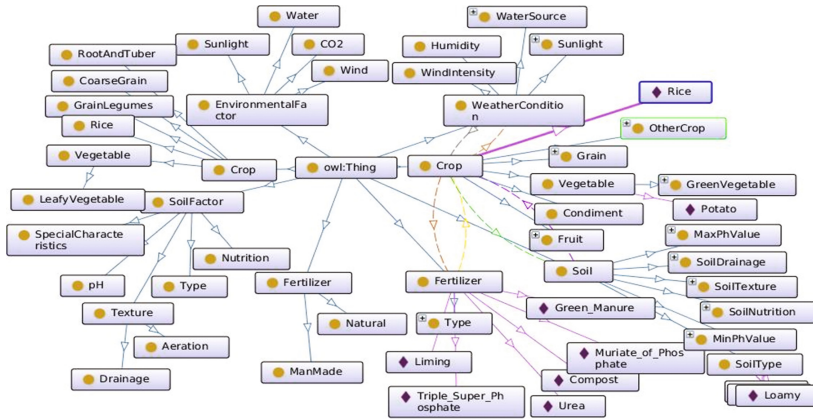


Fig. 5. Graphical representation of *Ontology3*: merged ontology using Protégé

Figure 6 shows *Ontology4*, merged ontology obtained using the proposed scheme. The metrics for this ontology are smaller than the sum of the original ones, hence, this tool is memory efficient. This scheme removes duplicate concepts in the merged ontology and also merges structurally similar concepts like *WeatherCondition* and *EnvironmentalFactor*. It also merges the instances and relations of concepts in the source ontologies. The scheme ensures that every aspect of source ontologies is present in the merged ontology. Thus, the proposed scheme obtained much better merged ontology in comparison with the merged ontology obtained using Protégé.

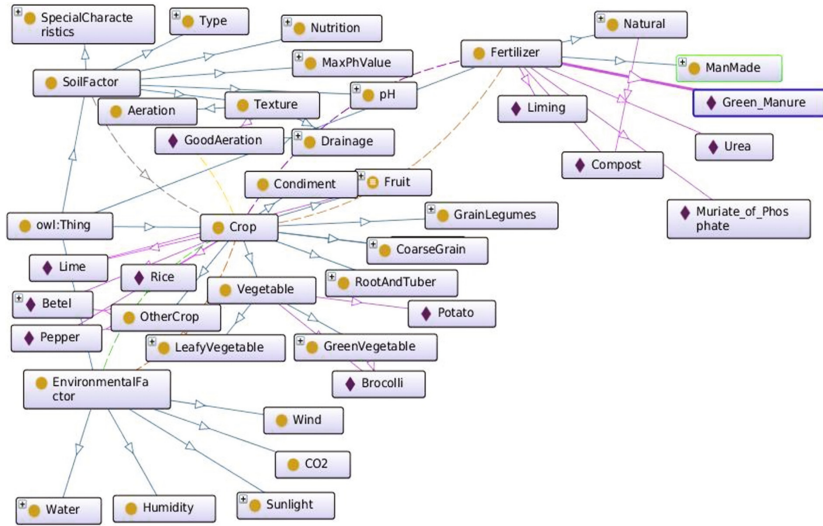


Fig. 6. Graphical representation of *Ontology4*: merged ontology using the proposed scheme

6 Conclusion and Future Work

The present work examines existing tools and proposes a new scheme for ontology merging. Protégé provides a good interface for ontology creation. It automatically generates OWL code from the information provided by the user regarding the classes, data properties, object properties, annotations etc. A lot of literature about ontology merging tools is available. However, the literature does not provide guidance for practical applications of the same. This paper equips the reader about a practical experience of ontology merging.

Protégé gives just a concatenation of source ontologies after merging. Moreover, it does not merge the instances and relations of the concepts of source ontologies. The scheme presented in this paper uses both element-level as well as structure-level techniques for alignment of concepts, instances and relations. It takes care of duplicity of concepts as well as incorrect alignments in the merged ontology. The merged ontology retains every aspect of the source ontologies.

The paper presents a good scheme for ontology merging taking practical example from agriculture domain. The scheme overcomes the shortcomings faced while using an existing tool for ontology merging. This work is being more rigorously tested using data-driven ontology evaluation and application-based ontology evaluation techniques for future work and improvement in the scheme.

Acknowledgements. The work has been supported by Department of Electronics and Information Technology, Ministry of Communication and Information Technology, Government of India in the form of a sponsored project entitled “Development of Tools for Automatic Term Extraction and RDFization of Agriculture Terms with focus on Crops sub-domain”.

References

1. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum. Comput. Stud.* **43**(5), 907–928 (1995)
2. Lata, S., Sinha, B., Kumar, E., Chandra, S., Arora, R.: Semantic web query on e-governance data and designing ontology for agriculture domain. *Int. J. Web Semant. Technol.* **4**(3), 65 (2013)
3. Malik, N., Sharan, A., Hijam, D.: Ontology development for agriculture domain. In: 2nd International Conference Computing for Sustainable Global Development (INDIACom), pp. 738–742. IEEE (2015)
4. Choi, N., Song, I.-Y., Han, H.: A survey on ontology mapping. *ACM Sigmod Rec.* **35**(3), 34–41 (2006)
5. Predoiu, L., Feier, C., Scharffe, F., de Bruijn, J., Martín-Recuerda, F., Manov, D., Ehrig, M.: D4. 2.2 state-of-the-art survey on ontology merging and aligning V2. In: EU-IST Integrated Project IST-2003-506826 SEKT, p. 79 (2005)
6. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. *IEEE Trans. Knowl. Data Eng.* **25**(1), 158–176 (2013)
7. Noy, N.F., Musen, M.A.: SMART: automated support for ontology merging and alignment. In: Proceedings of the 12th Workshop on Knowledge Acquisition, Modelling, and Management (KAW 1999), Banf, Canada (1999)
8. Noy, N.F., Musen, M.A.: Anchor-PROMPT: using non-local context for semantic matching. In: Proceedings of the Workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI), pp. 63–70 (2001)
9. Noy, N.F., Musen, M.A.: The PROMPT suite: interactive tools for ontology merging and mapping. *Int. J. Hum. Comput. Stud.* **59**(6), 983–1024 (2003)
10. Chalupsky, H.: Ontomorph: a translation system for symbolic knowledge. In: KR, pp. 471–482 (2000)
11. Ichise, R., Takeda, H., Honiden, S.: Rule induction for concept hierarchy alignment. In: Workshop on Ontology Learning (2001)
12. Kalfoglou, Y., Hu, B.: CROSI Mapping System (CMS) - result of the 2005 ontology alignment contest. In: Ashpole, B., Ehrig, M., Euzenat, J., Stuckenschmidt, H. (eds.) Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies, pp. 77–85 (2005)
13. Stumme, G., Maedche, A.: FCA-merge: bottom-up merging of ontologies. *IJCAI* **1**, 225–230 (2001)
14. McGuinness, D.L., Fikes, R., Rice, J., Wilder, S.: The chimaera ontology environment. In: AAAI/IAAI 2000, pp. 1123–1124 (2000)
15. Otero-Cerdeira, L., Rodríguez-Martínez, F.J., Gómez-Rodríguez, A.: Ontology matching: a literature review. *Expert Syst. Appl.* **42**(2), 949–971 (2015)
16. Sinha, B., Chandra, S.: Semantic web query on e-governance data for crop ontology model of Indian agriculture domain. In: Dutta, B., Madalli, D.P. (eds.) International Conference on Knowledge Modelling and Knowledge Management (ICKM), Bangalore (Bengaluru), pp. 56–66 (2013a)
17. Sinha, B., Chandra, S.: Semantic web ontology model for Indian agriculture domain. In: Dutta, B., Madalli, D.P. (eds.) International Conference on Knowledge Modelling and Knowledge Management (ICKM), Bangalore (Bengaluru), pp. 101–111 (2013b)
18. Chatterjee, N., Kaushik, N.: A practical approach for term and relationship extraction for automatic ontology creation from agricultural text. In: International Conference on Information Technology (ICIT), Bhubaneshwar, pp. 241–247 (2016)