Ontology enhancement using crowdsourcing: a conceptual architecture

Ontology
enhancement
using
crowdsourcing

231

Chepkoech C. Kiptoo

*School of Computing and Informatics, College of Biological and Physical Sciences,
University of Nairobi, Nairobi, Kenya*

Received 5 October 2019
Revised 26 February 2020
Accepted 3 March 2020

Abstract

Purpose – This paper aims to investigate the use of crowdsourcing in the enhancement of an ontology of taxonomic knowledge. The paper proposes a conceptual architecture for the incorporation of crowdsourcing into the creation of ontologies.

Design/methodology/approach – The research adopted the design science research approach characterised by cycles of “build” and “evaluate” until a refined artefact was established.

Findings – Data from a case of a fruit fly platform demonstrates that online crowds can contribute to ontology enhancement if engaged in a structured manner that feeds into a defined ontology model.

Research limitations/implications – The research contributes an architecture to the crowdsourcing body knowledge. The research also makes a methodological contribution for the development of ontologies using crowdsourcing.

Practical implications – Creating ontologies is a demanding task and most ontologies are not exhaustive on the targeted domain knowledge. The proposed architecture provides a guiding structure for the engagement of online crowds in the creation and enhancement of domain ontologies. The research uses a case of taxonomic knowledge ontology.

Originality/value – Crowdsourcing for creation and enhancement of ontologies by non-experts is novel and presents opportunity to build and refine ontologies for different domains by engaging online crowds. The process of ontology creation is also prone to errors and engaging crowds presents opportunity for corrections and enhancements.

Keywords Crowdsourcing, Citizen science, Biodiversity informatics, Ontology development, Ontology enhancement

Paper type Research paper

1. Introduction

The creation of domain ontologies is a daunting task and one of the major reasons ontologies are not widely adopted in knowledge representation is the difficulty of creating them. This difficulty is partly due to the complexity of domain knowledge being modelled and the limitations of the tools available for ontological modelling. More recently tools have emerged that has made the creation of ontologies easier by providing more friendly interfaces and checking for knowledge consistency as axioms are added into the ontology.

© Chepkoech C. Kiptoo. Published in *International Journal of Crowd Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

International Journal of Crowd
Science
Vol. 4 No. 3, 2020
pp. 231-243
Emerald Publishing Limited
2398-7294
DOI: 10.1108/IJCS-10-2019-0028

An example of a widely used editor for building ontologies in the OWL syntax is Protégé ontology editor (Protege, 2000) together with its built-in reasoner (Horridge *et al.*, 2011).

Ontology creation using these tools requires the ontology developer to capture all facts in the ontology. In the taxonomic knowledge domain, the axioms on every taxonomic grouping are closely related and the complexity increases as refinements are made to the species level. The ontology creation exercise is, therefore, very involving and is prone to errors of associating a feature with the wrong taxonomic group, as well as the omission of important features from a group. As stated, the ontology development tools mainly check for logical consistence in the knowledge. It is thus possible to have logically correct knowledge, which is not accurate – “inaccurate logically correct knowledge”. The current tools, therefore, do not guarantee accuracy and require intense expert involvement in the checking and verification of the represented knowledge.

In this research, we seek to explore how crowdsourcing techniques can be used to enhance and hasten ontology development. In this paper, we present an architecture of how crowdsourcing could be used to enhance the creation of ontologies, and a case of taxonomic knowledge ontology is used.

The rest of the paper is organised as follows. In Section 2, background and related work is presented covering what taxonomic knowledge is and ontologies of biological knowledge. Relevant background on ontology development and crowdsourcing is also presented. Section 3 presents research approach adopted, which is the design science research (DSR). In Section 4, the case study used in the development of the architecture is presented together with an analysis of the crowdsourced data. Section 5 presents the architecture itself consisting of the components and the ontology enhancement steps. Section 6 discussion of the results and Section 7 presents the conclusion.

2. Background and related work

2.1 Taxonomic knowledge

Taxonomy, in biological sciences, is the sub-discipline that deals with the identification, naming and classification of living things according to common properties. Taxonomic knowledge, therefore, constitutes the properties of the different taxonomic groupings. The properties used are mainly the apparent morphological features of the organisms. Taxonomy is, thus, a discipline that enables other sub-disciplines of biological sciences, as other forms of biological knowledge is organised around the taxonomic groupings described in the taxonomic knowledge (Clark *et al.*, 2009; Godfray, 2002).

Although different weaknesses have been identified regarding the use of taxonomy as a premise for the organization of biological knowledge, it still remains a popular approach, and a lot of organisms and biodiversity knowledge is organised around taxonomic groupings. The taxonomic knowledge, therefore, needs to be represented in appropriate formats to support the identification, naming and classification needs of biological sciences, as well as for the general public.

Creation of such knowledge bases would best be done by taxonomists, who are experts in the classification and identification knowledge of different organisms. Experts in taxonomy are on the decline, and the few taxonomists in different institutions are already overstretched attending to their core functions of identification and classification of organisms (Dar *et al.*, 2012; Kim and Byrne, 2006; Walter and Winterton, 2007). With advances in web technologies, the possibility of involving a broader community online to create biological knowledge bases has been recognised as a potential solution (Canhos *et al.*, 2004; Hardisty and Roberts, 2013; Parr *et al.*, 2014) and this research is anchored on this proposition.

2.2 Ontologies of taxonomic knowledge

Exploration of ontologies in the representation of knowledge in the biological domain has resulted in creation of multiple ontologies. The ontologies represent different categories of biological knowledge and some of these ontologies are hosted at the OBO Foundry ontologies website (www.obofoundry.org/). Examples of ontologies that focus on organism knowledge include the vertebrate trait (VT) ontology, which contains biological traits of vertebrates (Park *et al.*, 2013); flora phenotype ontology (FLOPO), which contains knowledge on traits and phenotypes of flowering plants (Thessen *et al.*, 2015); Afrotropical bee ontology, which contains taxonomic knowledge on different species of Afrotropical bees (Gerber *et al.*, 2014) and the ontology of fruit flies of economic importance in Africa (Kiptoo *et al.*, 2016a) among others.

2.3 Ontology development

An ontology is “a formal, explicit specification of a shared conceptualization” (Gruber, 1993). Formal means the specifications are encoded in a logic-based language; explicit specification means concepts are given explicit names and definitions; shared means the knowledge in the ontology can be shared and re-used by different users who subscribe to it; conceptualization means the way people think in a given domain (Corcho *et al.*, 2006; Guarino *et al.*, 2009; Uschold and Gruninger, 2004). An ontology has to meet these properties and development of ontologies is, therefore, a demanding task requiring input from logicians, as well as the target domain experts.

Different methodologies for development of ontologies have been developed over the years (Grüniger and Fox, 1995; Uschold and King, 1995; Fernández-López, 1999; Noy and McGuinness, 2001; Horridge *et al.*, 2009; Suárez-Figueroa *et al.*, 2012). Until the late 1990s, ontology development was seen as more of a craft and researchers mostly adapted some ontology or blended methodological components from various ontologies. In the recent years, software engineering-related methodologies have emerged, providing detailed guidelines on how ontologies should be conceived, specified, developed, adapted, deployed, used and maintained (Baclawski *et al.*, 2013).

In this research, the ontology development methodology by Horridge *et al.* (2011) was adopted. The process of creating an ontology can be summarised into a number of steps, namely, define classes and subclasses; define the object properties; and define the individuals. These steps are done repetitively until the ontology is sufficient for its purpose.

Development of the fruit fly ontology used in this research followed this methodology and a description using an example is provided in Section 4.1.

2.4 Crowdsourcing

The term crowdsourcing has been defined by several researchers, and the commonality in the definitions is the use of the open call format to enlist a multitude of humans to participate in solving a problem. In this research, we adopt the definition of crowdsourcing as “tapping into the collective intelligence of the public to complete a task” (Yuen *et al.*, 2011). The crowdsourced tasks can be categorised into two, namely, “microtasks” where a task is split into small puzzles whose results are aggregated to solve the main puzzle and “megatasks” where a task is solved in its main form (Good and Su, 2013).

Crowdsourcing research challenges are currently in four dimensions, namely, contributor recruitment, task design, data aggregation and management of abuse (Doan *et al.*, 2011; Hosseini *et al.*, 2014). This research looked at the aggregation of data to generate interpretations. The task design used in the crowdsourcing for the data is reported in Kiptoo (2017).

Task design in crowdsourcing often require design of custom workflows that breaks down complex tasks into pieces that can be done independently then combined (Chilton *et al.*, 2013). Crowdsourcing tasks come in different forms, including playing games, solving puzzles, transcribing, image tagging, translation, writing, etc. In this research, the crowdsourcing is done using image tagging tasks. Image tagging using crowdsourcing has been explored by many researchers (von Ahn and Dabbish, 2004; Chen *et al.*, 2013; Mavandadi *et al.*, 2012; Qin *et al.*, 2011). The tagging in this research is done using axioms from an ontology and not generic tags like the others. The participants are guided to make keen observations, as the nature of features to be tagged are in the fine details of the image.

Techniques for the aggregation of crowdsourced data have been studied by different researchers. Hung *et al.* (2013) classified the aggregation techniques into non-iterative and iterative. Non-iterative techniques compute a single result and examples include majority decision (MD) (Kuncheva *et al.*, 2003), honeypot (HP) (Lee *et al.*, 2010) and expert label injected crowd estimation (ELICE) (Khattak and Salleb-Aouissi, 2011). Iterative techniques perform a series of multiple iterations of input given by each worker also adjusts the expertise level of the worker based on performance. Examples of iterative techniques include the expectation maximization (EM) algorithm (Ipeirotis *et al.*, 2010) and ITER (Karger *et al.*, 2011).

3. Approach

The DSR approach was adopted. DSR is premised in the pragmatic philosophical view, which believes that the main reason for doing research is to improve human life through creation of interventions. The DSR approach takes a problem in a selected domain and solves it through introduction of an artefact that alters the way things are done thereby solving the problem (Benbasat and Zmud, 1999).

Different processes for conducting research in the DSR approach are documented in literature (March and Smith, 1995; Purao, Rossi and Bush, 2002; Peffers *et al.*, 2006; Hevner, 2007; Niederman and March, 2012). The general principle behind the processes is that of creating knowledge through a cyclic process of doing something (build) and assessing the results (evaluate) as documented in Owen (1998).

In this research, the research framework presented in Hevner *et al.* (2004) is adopted. The framework presents DSR as located in the middle of two categories of activities of Rigor and Relevance as shown in Figure 1.

The rigor activities are concerned with ensuring reference to appropriate knowledge bases and that the generated knowledge is grounded on relevant theories. The relevance activities are aimed at ensuring that the resulting theories are usable in practical environments.

4. Case study

Theory creation can be done at different stages during the build and evaluate cycles of design science research (Kuechler and Vaishnavi, 2012, 2008). In this research, the architecture was developed from evaluation of a case of ontology-driven fruit fly identification through crowdsourcing.

4.1 Creation of the case ontology

The ontology of fruit fly knowledge used in this study is presented in Kiptoo *et al.* (2016b). The ontology is based on a model whose overarching principle is to represent the morphological features of any taxonomic grouping as summarised in Box 1.

Box 1. Taxonomic group feature model example
SpecificFeature = BodyPart *n* has Feature Feature
MorphologicalFeature = MorphologicalFeature *n* has SpecificFeature
Taxagroup = TaxaGroup *n* has Feature MorphologicalFeature

To represent knowledge on “a beetle with a redhead”, a number of assertions must be made as shown in Box 2.

Box 2. Beetle with a red head feature

Item	Members
Classes:	Beetle BodyPart Color
Subclasses:	Head (subclass of BodyPart) Red (Subclass of Color) redHeadBeetle (Subclass of Beetle)
Object properties:	hasColor hasMorphololigalFeature
Assertions:	redHead = Head <i>n</i> hasColor Red redHeadBeetle = Beetle <i>n</i> hasMorphololigalFeature redHead

To build an ontology that has identification knowledge up to species level requires the taxonomic groupings to be specified, all body parts to be defined as subclass of body part (BodyPart), properties to be modelled as object properties, generic features (hasColour, hasTexture, hasSmell, etc.) to be modelled as feature, morphological features to be modelled by associating BodyPart with a Feature, and then finally associating the morphological feature with the taxonomic grouping. This process may seem straightforward, but when the taxonomic groupings and features are many, one can commit multiple errors resulting in an incomplete or erroneous ontology.

4.2 Crowdsourcing using the fruit fly ontology

Using the fruit fly ontology outlined above, a crowdsourcing platform for identification of fruit flies was developed. The objective of the platform was to enable non-expert online crowds to participate in the provision of species identification services with the support of the ontology Kiptoo *et al.* (2016a, 2016b).

The crowdsourcing task was for the crowds to tag images with features from the ontology as shown in Figure 2. A sketch image showing the different body parts was provided to guide participants on the names of the fruit fly body parts used in the identification. Taxonomic knowledge is often described using anatomical body parts that may not be obvious to the participating online crowds. In this task, the participant is expected to tick the features they can see in the sample and submit. Upon submission, another sample is loaded with the same task.

The crowdsourcing activity generated a data set that was then used in assessment of crowd's capability to provide the identification services. In this research, the same data set was analysed and the results demonstrate that ontology enhancement can be performed from the ontology-driven crowd sourcing activities. The architecture for this enhancement is presented in Section 5.

4.3 Analysis of data

The crowds generated a data set of features that they observed on the images presented to them. The research used seven samples of fruit fly images already identified by experts. In

Figure 2.
Image tagging with
axioms from fruit fly
ontology

total, 30 online users participated in tagging the images with features from the ontology and close to nine thousand feature tags were made on the images. The MD aggregation technique was used through a simple count of the number of times a tag was made on a sample.

Analysis of the data together with the ontology led to the two key findings that are useful in the enhancement of the ontology, namely:

- (1) There are identification features that the crowds identified but are missing in the ontology; and
- (2) There are features that are described in the ontology but are not used by the crowd.

We describe what these two findings mean to ontology enhancement.

4.3.1 Features identified by the crowd yet missing in ontology. The crowd associated several features with some samples, and yet these features were not in the ontology. For instance, in Sample 1, the crowd tagged the feature “fore femur yellow” 24 times. A check at this feature in the taxonomic key found it present and this means that the feature was omitted in the creation of the ontology. The validity status of features that are missing from the ontology yet was tagged over 20 times by the crowd is as summarised below (Table 1).

From these results, it can be seen that the crowd can tag features that were omitted from the ontology. The important thing is to have clearly stated observable features and high-resolution images, especially when dealing with taxa that are small in size like the fruit fly case used in this research.

4.3.2 Features in ontology not observed by the crowd. The crowdsourced data showed a number of features present in the ontology yet minimally or not tagged by the crowd. The ontology had “*Ceratitis Bremii*” having the feature “Face Transverse Band Dark Yellow” associated with it yet no one tagged this feature. A sample of other features not tagged is summarised below (Table 2).

A check at the features not tagged by the crowd revealed two major possibilities; either the image is not clear enough and therefore, the feature not easily visible or the crowd participants are not able to interpret the feature. This brings out an important factor when creating a crowdsourcing platform. The task should be very clear, and the imagery used should be clear so that the features are readily visible. The names of the anatomic body parts are not obvious to the crowd and therefore, a detailed key of body parts must be provided to support the crowd in spotting the features.

Table 1.
Features spotted by
the crowd yet
missing in ontology

Sample	Feature	No. of times tagged	Validity
Ceratitis Cosyra Female	Fore femur yellow	24	The sample had the feature but was omitted during ontology creation Feature tagged by crowds but not valid. Can be attributed to ambiguity of the feature
Dacus Ciliatus Loew Male	Post Pronotal Lobe Paler	24	
Ceratitis Capitata	Scutellum Yellow With Black Spots	25	The sample had the feature but was omitted during ontology creation
Dacus Bivittatus Female	All Tibia Dark	22	The sample had the feature but was omitted during ontology creation

Table 2.
Features in ontology
but not tagged by the
crowd

Sample	Feature	Findings
Ceratitis Ditissima Female	Wing Costal Band Broad and Very Dark Without Break	The feature was in the ontology. The feature was not clear in the sample image provided and wing costal band not labelled in image key
Bactocera Olae	Wing With Reticulate Appearance	The feature was erroneously associated with the taxa in the ontology so should be removed
Several samples	Wing Sub Costal Vein Bends at 90 Degree Angle	The feature was in the ontology. The body part wing subcostal vein was not labelled in the image key. The feature was rarely tagged in various samples yet it is a main identifier of fruit flies
Ceratitis Bremii	Face Transverse Band Dark Yellow	The feature was in the ontology. The image was not clear, face transverse band not labelled in image key
Bactocera Caudata	Abdominal Tergite Separated	The feature was in the ontology. The body part abdominal tergite was not indicated in the image

5. Proposed conceptual architecture

As stated in Section 2.1, creation of ontologies involves the description of the desired domain knowledge through the declaration of axioms. In OWL ontologies, the axioms consist of classes, individuals and properties that are present in the domain of interest (Horridge *et al*, 2011). Depending on the domain knowledge and the questions the ontology is expected to answer, the ontology designer decides, which items are categorised as classes, individuals and properties.

Depending on the source of knowledge, crowdsourcing platforms can be used to engage online crowds to register their observations which upon aggregation could result in axioms, which are added to the ontology. The knowledge sources include knowledge with domain experts, knowledge recorded in different documentations and the knowledge observable in our environments. Online crowds can be engaged in creating the classes, individuals and properties based on the model of the ontology. It is up to the developer of the platform to decide how to design the crowd tasks to capture data that can be synthesised for generation of ontology axioms. The proposed architecture for crowdsourcing for ontology enhancement is as shown in Section 3.

5.1 Architecture components

The conceptual architecture has three components, namely, ontology creation, algorithms and crowdsourcing platforms, described in the sections below.

5.1.1 Ontology creation. As illustrated, the different activities – class, properties and individuals description – of the ontology creation can be achieved through the synthesis of data generated using crowdsourcing. The lines joining the algorithm and the ontology creation activities are dotted, and this means that the implementer of this architecture can target any of the ontology development activities depending on the levels at which the developer wants to engage online crowds. If the developer wants the crowd to identify classes, then crowdsourcing activities that support this are designed and once generated by the crowd, they can be added to the ontology. For example, the taxonomic ontologies, if the designer wants to use crowdsourcing to identify the colours present in different body parts

of a sample, a crowdsourcing task to ask for colours the crowd can see. The common colours can then be added as subclasses to the colour class.

5.1.2 Algorithms. Algorithms consist mainly of the tools used to analyse the crowdsourced data for purposes of adding axioms to the ontology or scheduling more crowd activities when needed. As shown in Figure 3, the inputs to the algorithms component are the stable version of the ontology and the crowdsourced data.

5.1.3 Crowdsourcing platform. The crowdsourcing platform component contains the tools designed to enable engagement of online crowds through the performance of clearly defined tasks. The platforms must be designed in ways that the data generated can be analysed to produce axioms for the ontology. A key input to the crowdsourcing platform is the stable version of the ontology. Axioms from this ontology is used to solicit for input from online crowds. This input can then be aggregated to generate more axioms for the enhancement of the ontology. In the fruit fly example, using the stable version of the ontology led to identification of omitted and incorrect axioms.

5.2 Ontology enhancement steps

The ontology enhancement steps can be summarised as follows:

- Use the stable version of the ontology.
- Create crowdsourcing platform that can generate a data set for ontology enhancement.
- Analyse the data set for omissions and generate potential axioms.
- Analyse the data set for committed errors and generate potential axioms for deletion.
- Confirm validity of potential axioms.
- Enhance the ontology accordingly.

Figure 3.
Conceptual
architecture for
ontology
enhancement using
crowdsourcing

6. Discussion of results

From the analysis above, we see that online crowds can be engaged to verify the axioms in the ontology. Creation of ontologies is a repetitive exercise and therefore, prone to errors. Engaging online crowds can aid in the correction of assertions made in the ontology. It is also possible to structure the ontology creation problem in a way that the ontology expert creates the high-level structures based on a model then the task of populating the ontology with repetitive features can be delegated to online crowds. In this case of taxonomic knowledge, the experts can focus on defining the features, then the role of associating those features with samples can be delegated to online crowds. It is, however, important to note that the quality of the crowdsourcing environment and samples used are key in getting high-quality results. The crowdsourcing platform must ensure the crowd is adequately guided so that they can effectively tag the features. The samples used must also be high-quality images with features clearly visible.

7. Conclusion and future work

In this paper, an architecture for ontology enhancement using crowdsourcing is presented. The results of the case used in the study demonstrate that this architecture can be adopted for online crowds' participation in ontology enhancement. The architecture is an addition to the existing crowdsourcing models in literature and forms basis for interrogation and enhancement by other researchers. The architecture is also an artefact that can be adopted by application developers when building applications for ontology enhancement.

In future evaluation of the architecture can be conducted with respect to aggregation algorithms used and design of crowd tasks. This research used a simple aggregation algorithm – MD – in the analysis of the crowd sourced data. More sophisticated aggregation algorithms could be employed so that different weights can be assigned to the contributors thus attaining trustable thresholds even with few participants. The crowdsourcing platform can also be improved to support the use of multiple images for the same sample thereby allowing observation of features from different dimensions.

References

- Baclawski, K., Bennett, M., Dean, M., Denny, M., Gruninger, M., Hashemi, A., Longstreth, T., (2013), *Communique. Towards Ontology Evaluation across the Life Cycle*, in Neuhaus, F. and Vizedom, A. (Eds), Ontology Summit, p. 14.
- Benbasat, I. and Zmud, R.W. (1999), "Empirical research in information systems: the practice of relevance", *MIS Quarterly*, Vol. 23 No. 1, pp. 3-16.
- Canhos, V.P., de Souza, S., Giovanni, R.D. and Canhos, D.A.L. (2004), "Global biodiversity informatics: setting the scene for a 'new world' of ecological forecasting", *Biodiversity Informatics*, Vol. 1, doi: [10.17161/bi.v1i0.3](https://doi.org/10.17161/bi.v1i0.3).
- Chen, M., Zheng, A. and Weinberger, K. (2013), "Fast image tagging", *Proceedings of the 30th International Conference on Machine Learning*, pp. 1274-1282.
- Chilton, L.B., Little, G., Edge, D., Weld, D.S. and Landay, J.A. (2013), "Cascade: crowdsourcing taxonomy creation", *Conference on Human Factors in Computing Systems – Proceedings*, pp. 1999-2008.
- Clark, B.R., Godfray, H.C.J., Kitching, I.J., Mayo, S.J. and Scoble, M.J. (2009), "Taxonomy as an eScience", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 367 No. 1890, pp. 953-966.

-
- Dar, G.H., Khuroo, A., Reddy, C.S. and Malik, A. (2012), "Impediment to taxonomy and its impact on biodiversity science: an Indian perspective", *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, Vol. 82 No. 2, pp. 235-240.
- Corcho, O., Fernández-López, M. and Gómez-Pérez, A. (2016), "Ontological engineering: principles, methods, tools and languages", in Calero, C., Ruiz, F. and Piattini, M. (Eds), *Ontologies for Software Engineering and Software Technology*, Springer Berlin Heidelberg, Heidelberg, pp. 1-48.
- Doan, A., Ramakrishnan, R. and Halevy, A.Y. (2011), "Crowdsourcing systems on the World-Wide Web", *Communications of the Acm*, Vol. 54 No. 4, doi: [10.1145/1924421.1924442](https://doi.org/10.1145/1924421.1924442).
- Fernández-López, M. (1999), "Overview of methodologies for building ontologies", *Proceedings of the IJCAI99 Workshop on Ontologies and Problem Solving Methods Lessons Learned and Future Trends*, CEUR Publications, available at: <http://oa.upm.es/5480/>
- Gerber, A., Eardley, C. and Morar, N. (2014), "An ontology-based taxonomic key for Afrotropical bees", in Garbacz, P. and Kutz, O. (Eds), *Proceedings of Frontiers in Artificial Intelligence and Applications (FOIS)*, IOS Press, pp. 277-288.
- Godfray, H.C.J. (2002), "Challenges for taxonomy", *Nature*, Vol. 417 No. 6884, pp. 17-19.
- Good, B.M. and Su, A.I. (2013), "Crowdsourcing for bioinformatics", *Bioinformatics (Oxford, England)*, Vol. 29 No. 16, pp. 1925-1933.
- Gruber, T.R. (1993), "A translation approach to portable ontology specifications", *Knowledge Acquisition*, Vol. 5 No. 2, pp. 199-220.
- Grüniger, M. and Fox, M.S. (1995), "Methodology for the design and evaluation of ontologies", *Workshop on Basic and Ontological Issues in Knowledge Sharing*, Citeseer, Montreal.
- Guarino, N., Oberle, D. and Staab, S. (2009), "What is an Ontology?", *Handbook on Ontologies*, Springer, pp. 1-17.
- Hardisty, A. and Roberts, D. (2013), "A decadal view of biodiversity informatics: challenges and priorities", *BMC Ecology*, Vol. 13 No. 1, p. 16.
- Hevner, A.R. (2007), "A three cycle view of design science research", *Scandinavian Journal of Information Systems*, Vol. 19 No. 2, pp. 4.
- Hevner, A.R., March, S., Park, J. and Ram, S. (2004), "Design science research in information systems", *MIS Quarterly*, Vol. 28 No. 1, pp. 75-105.
- Horridge, M., Jupp, S., Moulton, G., Rector, A., Stevens, R. and Wroe, C. (2009), *A Practical Guide to Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1. 2*, The University of Manchester.
- Horridge, M. Knublauch, H. Rector, A. Stevens, R. Wroe, C. Jupp, S. Moulton, G. *et al.* (2011), "A practical guide to building OWL ontologies using protégé 4 and CO-ODE tools edition 1.3", available at: http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf
- Hosseini, M., Phalp, K., Taylor, J. and Ali, R. (2014), "The four pillars of crowdsourcing: a reference model", *Proceedings – International Conference on Research Challenges in Information Science*, IEEE Computer Society, doi:[10.1109/RCIS.2014.6861072](https://doi.org/10.1109/RCIS.2014.6861072).
- Hung, N.Q.V., Tam, N.T., Tran, L.N. and Aberer, K. (2013), "An evaluation of aggregation techniques in crowdsourcing", *Web Information Systems Engineering – WISE*, 2013, Springer, pp. 1-15.
- Ipeirotis, P.G., Provost, F. and Wang, J. (2010), "Quality management on Amazon mechanical turk", *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 64-67.
- Karger, D.R., Oh, S. and Shah, D. (2011), "Iterative learning for reliable crowdsourcing systems", *Advances in Neural Information Processing Systems*, pp. 1953-1961.
- Khattak, F. and Salheb-Aouissi, A. (2011), "Quality control of crowd labeling through expert evaluation", *Proceedings of the NIPS 2nd Workshop on Computational Social Science and the Wisdom of Crowds*, pp. 1-5.

- Kim, K.C. and Byrne, L.B. (2006), "Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science", *Ecological Research*, Vol. 21 No. 6, pp. 794-810.
- Kiptoo, C.C., Gerber, A. and Van der Merwe, A. (2016b), "Towards citizen-expert knowledge exchange for biodiversity informatics: a conceptual architecture", *The African Journal of Information and Communication (AJIC)*.
- Kiptoo, C.C. (2017), "An ontology and crowd computing model for expert-citizen knowledge transfer in biodiversity management. Study leaders: Auroa G., van de Merwe A", PhD Thesis., University of Pretoria, available at: <https://repository.up.ac.za/handle/2263/62776>
- Kiptoo, C.C. Gerber, A. and van der Merwe, A. (2016a), "The ontological modelling of fruit fly control and management knowledge", *Fruit Fly Research and Development in Africa – Towards a Sustainable Management Strategy to Improve Horticulture*, available at: https://doi.org/10.1007/978-3-319-43226-7_11
- Kuechler, W.L. and Vaishnavi, V.K. (2012), "A framework for theory development in design science research: multiple perspectives", *Journal of the Association for Information Systems*, Vol. 13 No. 6, pp. 395-423.
- Kuechler, W.L.J. and Vaishnavi, V.K. (2008), "On theory development in design science research: anatomy of a research project", *European Journal of Information Systems*, Vol. 17 No. 5, pp. 489-504.
- Kuncheva, L.I., Whitaker, C.J., Shipp, C.A. and Duin, R.P.W. (2003), "Limits on the majority vote accuracy in classifier fusion", *Pattern Analysis and Applications*, Vol. 6 No. 1, pp. 22-31.
- Lee, K., Caverlee, J. and Webb, S. (2010), "The social honeypot project: protecting online communities from spammers", *Proceedings of the 19th International Conference on World Wide Web*, pp. 1139-1140.
- March, S.T. and Smith, G.F. (1995), "Design and natural science research on information technology", *Decision Support Systems*, Vol. 15 No. 4, pp. 251-266.
- Mavandadi, S., Dimitrov, S., Feng, S., Yu, F., Sikora, U., Yaglidere, O., Padmanabhan, S., *et al.* (2012), "Distributed medical image analysis and diagnosis through crowd- sourced games: a malaria case study", *PLoS One*, Vol. 7 No. 5, p. e37245.
- Niederman, F. and March, S.T. (2012), "Design science and the accumulation of knowledge in the information systems discipline", *ACM Transactions on Management Information Systems*, Vol. 3 No. 1, pp. 1-15.
- Noy, N.F. and McGuinness, D.L. (2001), "Ontology development 101: a guide to creating your first ontology", Stanford Knowledge Systems Laboratory, Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880.
- Owen, C. (1998), "Design research: building the knowledge base", *Design Studies*, Vol. 19 No. 1, pp. 9-20.
- Park, C.A., Bello, S.M., Smith, C.L., Hu, Z.-L., Munzenmaier, D.H., Nigam, R., Smith, J.R., *et al.* (2013), "The vertebrate trait ontology: a controlled vocabulary for the annotation of trait data across species", *Journal of Biomedical Semantics*, Vol. 4 No. 1, p. 13.
- Parr, C.S., Wilson, N., Leary, P., Schulz, K.S., Lans, K., Walley, L., Hammock, J.A., *et al.* (2014), "The encyclopedia of life v2: providing global access to knowledge about life on earth", *Biodiversity Data Journal*, Vol. 2, p. e1079.
- Peffers, K., Tuunanen, T., Gengler, C.E., Rossi, M., Hui, W., Virtanen, V. and Bragge, J. (2006), "The design science research process: a model for producing and presenting information systems research", *Proceedings of Design Research in Information Systems and Technology DESRIST'06*, Vol. 24, pp. 83-106.
- Protege, T. (2000), "Protege 2000", available at: <http://protege.stanford.edu/>
- Purao, S., Rossi, M. and Bush, A. (2002), "Towards an understanding of the use of problem and design spaces during object-oriented system development", *Information and Organization*, Vol. 12 No. 4, pp. 249-281.

-
- Qin, C., Bao, X., Roy Choudhury, R. and Nelakuditi, S. (2011), "TagSense: a smartphone-based approach to automatic image tagging", *MobiSys'11 – Compilation Proceedings of the 9th International Conference on Mobile Systems, Applications and Services and Co-Located Workshops*, pp. 1-14.
- Suárez-Figueroa, M.C., Gómez-Pérez, A. and Fernández-López, M. (2012), "The neon methodology for ontology engineering", *Ontology Engineering in a Networked World*, Springer, pp. 9-34.
- Thessen, A.E., Bunker, D.E., Buttigieg, P.L., Cooper, L.D., Dahdul, W.M., Domisch, S., Franz, N.M., *et al.* (2015), "Emerging semantics to link phenotype and environment", *PeerJ*, Vol. 3, p. e1470.
- Uschold, M. and King, M. (1995), "Towards a methodology for building ontologies", *Workshop on Basic and Ontological Issues in Knowledge Sharing*, Vol. 80, pp. 275-280.
- Uschold, M. and Gruninger, M. (2004). "Ontologies and semantics for seamless connectivity" *ACM SIGMOD Record*, Vol. 33 4, pp. 58-64.
- von Ahn, L. and Dabbish, L. (2004), "Labeling images with a computer game", *Proceedings of the 2004 Conference on Human Factors in Computing Systems – CHI '04*, pp. 319-326.
- Walter, D.E. and Winterton, S. (2007), "Keys and the crisis in taxonomy: extinction or reinvention?", *Annual Review of Entomology*, Vol. 52 No. 1, pp. 193-208.
- Yuen, M., King, I. and Leung, K. (2011), "A survey of crowdsourcing systems", *IEEE International Conference on Social Computing*, pp. 766-773.

Further reading

Hevner, A. and Chatterjee, S. (2010), "Design science research in information systems", available at: http://link.springer.com/10.1007/978-1-4419-5653-8_2 (accessed 4 July 2016).

Corresponding author

Chepkoech C. Kiptoo can be contacted at: cckiptoo@gmail.com