Supporting Taxonomy Development and Evolution by Means of Crowdsourcing

Binh Vu and Matthias Hemmje
FernUniversität in Hagen, Hagen, Germany

Keywords: Taxonomy, Taxonomy Development, Taxonomy Management, Crowdsourcing, Knowledge, Knowledge Management, Information Overload, Ecosystem Portal.

Abstract: Information overload continues to be a challenge. By dividing the material into many different small subsets, classification based on a taxonomy makes data exploration and retrieval faster and more accurate. Instead of having to know the exact keywords that describe the knowledge resource, users can browse and search for them by selecting the categories that the resource is most likely to belong. Nevertheless, developing taxonomies is not an easy task. It requires the authors to have a certain amount of knowledge in the domain. Furthermore, the workload will increase as any new taxonomy needs to be frequently updated to remain relevant and useful. To combat these problems, this paper proposes another approach to crowdsource taxonomy development and evolution. We describe in this paper the concept of this approach along with different types of evaluations targeting on the one hand to demonstrate the feasibility of the approach and the usability of the initial prototype as well as on the other hand the quality and effectiveness of the chosen method.

1 INTRODUCTION AND MOTIVATION

Today's internet is a big source of information available in the form of content and also more explicit forms of knowledge resources. Every day there is a huge amount of such content and knowledge resource data moving on the internet. Most companies in the U.S. in 2005 have at least 100 TB of such data stored. They estimate that by 2020, 40 Zettabytes (43 trillion Gigabytes) will be created, an increase of 300 times from 2005 (The Four V's of Big Data, 2005). Such data not only needs to be indexed, but also the index terms should be unique and descriptive. Otherwise, an indexer would have to classify documents, which are from the same topic, to various categories, despite the fact that these categories may have the same or very similar meaning. This would make searching and comparing results afterward more difficult. A taxonomy, in this case, can be a source of a unique and well-controlled vocabulary. It is a hierarchy of agreed-on terms, which later can be used for indexing or classifying documents. This means, with the support of taxonomy, classification consistency can be achieved (Vu et al., 2018).

The development of a new taxonomy is usually done by knowledge workers and domain experts. While providing many benefits and advantages, it also has problems. New approaches involving the crowd in the development and management of taxonomies were introduced to overcome these constraints using the "wisdom of the crowd" (Karampinas & Triantafillou, 2012).

By using "the power of the crowd", one can achieve definitions of taxonomy terms and relations that no person or organization alone can achieve. One example of crowdsourcing is the knowledge resource Wikipedia, which is considered as one of the world's largest crowdsourcing projects. It was initially an English-language encyclopedia. Today, Wikipedia has more than 40 million articles in 301 different languages. All of them were written by the crowd through a model of content editing by means of web-based applications, called a wiki (Wikipedia, n.d.).

With crowdsourcing, human resources only need to work when they want, when they need to, as much as they need to and for whomever they like, and to choose the activities that they will do. This makes them happier compared to traditional types of employing human resources. Moreover, people's goods can be shared to lower their expenses and avoid waste due to collaborative consumption (Andro, 2018).

Vu, B. and Hemmje, M.
Supporting Taxonomy Development and Evolution by Means of Crowdsourcing.
DOI: 10.5220/0008348003510358
In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019)*, pages 351-358
ISBN: 978-989-758-382-7

From a company's point of view, by applying crowdsourcing, they ideally receive work results in much higher quality, quantity, and at a lower cost in less time. The work is done for free as the crowd workers hope to be compensated by the crowds, e.g., in a crowdcontest or other forms of social incentives not necessarily completely excluding later payment. In addition, a company would ideally benefit from a large number of proposals while having only a few individuals to compensate for a much lower overall cost than that of traditional human resource employments (Andro, 2018).

In this paper, we want to introduce existing approaches to develop and manage taxonomy, as well as point out their challenges. To overcome these problems, we purpose another method of applying crowdsourcing in taxonomy development, evolution, and management.

2 PROBLEM STATEMENT

The combination of taxonomy development and crowdsourcing poses additional challenges. A part of these challenges was mentioned in the authors' previous publication (Vu et al., 2018)

Developing a taxonomy involves many people, such as IT staff, corporate knowledge workers, departmental publishers, etc (Kon & Hoey, 2005). The more people are working together, the more problems it can potentially generate. On another hand, working alone can get us surrounded by information and knowledge resources that only support one point of view and forget other alternatives.

Furthermore, things always change. To reflect, e.g., the changing needs in knowledge domain concept and resource modeling, taxonomies need to be maintained frequently. Without maintenance and governance, and especially a tool to manage version and ownership, taxonomies can be drifting away from business and organizational information needs (Lambe, 2007).

Storing and processing taxonomy representations potentially requires a lot of computational and storage resources. Therefore, we need to consider how to organize the taxonomy in the database in such a way that it requires less space and is fast to retrieve.

One primary problem of crowdsourcing is how to motivate the crowd. Each individual engaged in crowdsourcing has their motivation. The motivation to participate in crowdsourcing is not very different from the motivation to participate in blogging, creating open-source software, etc. (Brabham, 2013). Some do it for fun and recognition. Some do it for financial reward. The problem is not every organization has the ability to provide all these incentives to the crowd.

The next problem is the quality of results produced by the crowd. Although one requirement for "the wisdom of crowds" is diversity, there is always unskilled, unrelated, insufficient people in the crowd. Compared to experts, cheap (sometimes, free) labor is likely to produce less quality work. Therefore, we need to either lower the complexity of the task or find a skilled crowd, which is not always easy (Eskenazi et al., 2013)

Crowdsourcing is difficult to manage. Not only it needs more resources for management but also bears challenges in security and privacy. It is hard to keep a project secret when it involves many people working on it from everywhere. Furthermore, collaboration generates personal data, which need to be handled carefully. All of this adds more problems to the management process, which was already difficult (Bar & Maheswaran, 2014).

3 STATE OF THE ART

In this chapter, we provide an overview of a selection of important fundamental concepts that are related to knowledge resource management, taxonomy management, and crowdsourcing. Furthermore, relevant approaches using social tagging and applying crowdsourcing in forming a term corpus or creating hierarchical relationships between terms will be mentioned.

3.1 Knowledge Management

Knowledge Management (KM), like most complex things, has many different definitions. Depending on the nature of the scientific area, the definition of KM might have a different meaning. Nevertheless, what Devenport and Prusak wrote in their book "Working Knowledge: How Organizations Manage What They Know" was agreed and cited the most: "Knowledge management draws from existing resources that your organization may already have in place good information systems management, organizational change management, and human resources management practices. If you've got a good library, a textual database system, or even effective education programs, your company is probably already doing something that might be called knowledge management" (Davenport & Prusak, 1998).

The Content and Knowledge Management Ecosystem Portal (KM-EP), has been developed to provide powerful web-based tools for managing knowledge resources and content (Vu, 2015). Figure 1 presents KM-EP's architecture, which consists of five subsystems:

- Information Retrieval Subsystem (IRS) indexes contents and lets the user search for them in a quick manner.
- Learning Management Subsystem (LMS) helps, e.g., a course creator, who is not an expert of the KM-EP and the underlying Learning Management System - Moodle, to create and manage courses.
- Content and Knowledge Management Subsystem (CKMS) manages contents and knowledge resources. It allows users to create, edit, remove and rate different type of contents in the ecosystem.
- User Management Subsystem (UMS) manages users, groups of users, authentication, and access control for all subsystems.
- Storage Management Subsystem (SMS) preserves the integrity of the digital file and its metadata for the lifetime of an asset. (Vu et al., 2018)

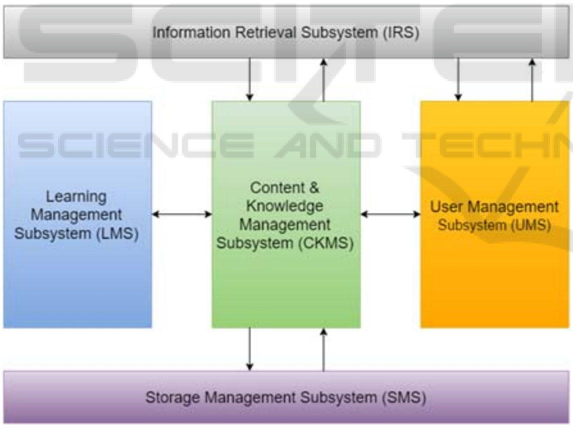

Figure 1: KM-EP architecture (Vu, 2015).

In the context of the research and development work presented in this paper, an additional taxonomy management system was developed as part of the KM-EP. The system allows domain experts to create new taxonomies. These taxonomies later will be used for supporting the classification, searching, and browsing of content and knowledge resources in the KM-EP. Nevertheless, without the support of crowdsourcing, only a given group of people had access rights to modify existing taxonomies. Normal users could not access the system and therefore, were not able to create their own taxonomy. Furthermore, there is no option to add information about the authors of a taxonomy, and additional properties such as descriptions or keywords.

3.2 Taxonomy Development

The term "taxonomy" has a very broad meaning and is being used in many areas, from psychology, biology to computer science.

In her book "The Accidental Taxonomist", Hedden see taxonomy in a broad sense as "any means of organizing concepts of knowledge" and in a broader sense as "a knowledge organization system or knowledge organization structure" (Hedden, 2010). The term "knowledge organization systems" was mentioned in 2000 by Hodge as a synonym for taxonomy. There are various types of knowledge organization systems, which include (1) term lists, such as authority files, glossaries, dictionaries, and gazetteers, (2) classifications and categories, such as subject headings, classification schemes, taxonomies, and categorization schemes and (3) relationship lists, such as thesauri, semantic networks, and ontologies (Hodge, 2000).

The development of a new taxonomy is usually done using the Delphi method, which is a technique to obtain the most reliable consensus of opinions of a group of experts through a series of intensive questionnaires interspersed with controlled opinion feedback (Dalkey & Helmer, 1963). The traditional method of using knowledge workers and experts for reviewing, while providing many benefits and advantages, has its own problems. One example is that experts are not always available. They have other jobs to do, and rounds of reviewing take too much time from their schedule. Another example is that people tend to ignore disagreements. The result is, e.g., a poor design decision which is ignored during reevaluation and not getting fixed.

3.3 Crowdsourcing and Crowdknowledge

"Under the right circumstances, groups are remarkably intelligent and are often smarter than the smartest people in them. Groups do not need to be dominated by exceptionally intelligent people in order to be smart" wrote James Surowiecki in his book, The Wisdom of Crowds. By putting together a big enough and diverse enough group of people, we can produce decisions better than experts. Therefore, chasing the expert for answers is a mistake. Group's decisions will, over time, be intellectually superior to the isolated individual, no matter how smart or well-informed he is (Surowiecki, 2005).

According to Estellés-Arolas and Guevara, there are 40 definitions for the concepts of crowdsourcing that come from 32 distinct articles published from 2006 to 2011 (Estellés-Arolas & Guevara, 2012). The term was created by Jeff Howe in his article "The Rise of Crowdsourcing" in 2006. It a combination of "crowd" and "outsourcing" and can be described as "the act of taking work once performed within an organization and outsourcing it to the general public through an open call for participants" (Ridge, 2014).

Crowdvoting is, e.g., one type of crowdsourcing. Its objective is to know the opinions of the crowd regarding specific issues or products. Here, people are giving their opinions and vote on a certain topic (Simon, Pechuan, & Estelles-Miguel, 2015) (Jimenez-Crespo, 2017) (Kitchens & Crane, 2014) (Turban, King, Lee, Liang, & Turban, 2015).

3.4 Related Works

The concept of crowdsourcing is fairly new. Nevertheless, the idea of crowdsourcing taxonomy development and evolution was already applied in scientific publications. There are two steps involving in the process of developing a new taxonomy: forming a term corpus and creating hierarchical relationships between terms. Crowdsourcing can be used in either one of these steps or in both of them.

The work of forming a term corpus using crowdsourcing in the first step was introduced by the mean of social tagging and folksonomy. Popular tagging systems, which were mentioned the most in scientific publications, are social bookmarking website Delicious and photo-sharing site Flickr. They have features that allow the user to add tags to existing contents, in contrast to stricter systems like libraries where a book will have exactly one proper call number based on content (Heymann & Garcia-Molina, 2006). These tags together form a folksonomy and can be used as terms for the developing taxonomy.

Nevertheless, folksonomy has its disadvantages. There is no control of synonymy and homonymy, there are many formats for dates and a lot of typing and orthographic errors. Tags can also contain words from different languages or even compound words consisting of more than two words or a mixture of languages. Combining all tags from a system, we can find many words that have the same meaning or same words but in different forms, e.g., "bag" vs "bags", "computer science" vs "computer_science" and "computer-science" (Peters & Stock, 2007).

Besides the approach using folksonomy, there are other methods to create a term corpus for taxonomy without using crowdsourcing, such as extracting words with top term frequency - inverse document frequency score (Brooks & Montanez, 2006) or get words or phrases in the top-ranked documents that commonly co-occur with each other across many of the passages (Sanderson & Croft, 1999).

From the terms' corpus created in the first step, the creators form hierarchical relationships between terms and get the final result as a new taxonomy in the second step. One method is to apply an algorithm to grow deeper, bushier tree by merging saplings created by different users, called SAP (Plangprasopchok, Lerman, & Getoor, 2010).

Another method was introduced by Heymann and Garcia-Molina (Heymann & Garcia-Molina, 2006). Their idea is to convert tag into tag vectors and calculate the similarity between tags using the cosine similarity between tag vectors. The end product is a tag similarity graph where each tag is represented by a vertex, and two vertices are connected by an edge if the similarity of the nodes they represent is above some set threshold.

Liu et al. computes a generality score for each tag, then use agglomerative hierarchical clustering approach to generate the concept hierarchy (Liu, Fang, & Zhang, 2010). Their algorithm has the same principle as Heymann's. Tags are sorted by their score in descending order. In this case, it is the generality score. Then the algorithm tries to find the parent node in the taxonomy tree for each tag. If it cannot be found, the tag is added as a child of the root. Agglomerative Hierarchical Clustering (AHC) is frequently used to build the hierarchy of tags. It relies on how similar / distant two nodes are in building a hierarchy. Li et al. proposed an enhance AHC framework by skipping the error-prone step of calculating each tag's generality and integrating a topic model to capture thematic correlations among tags (Li, et al., 2012).

An interesting approach was introduced by Karampinas and Triantafillou (Karampinas & Triantafillou, 2012). Rather than calculating the similarity score between two tags, they use the crowd to annotate parent-children relationships between tags. An algorithm, called "CrowdTaxonomy", was introduced to grow the taxonomy tree based on the crowd's annotations. The algorithm is called on every vote. This method includes the crowd in both steps of the taxonomy development process. The crowd is used to form a terms corpus by the mean of social tagging, and they vote to annotate pair between two terms. Hierarchical relationships are built based on their annotations. The work of Karampinas and Triantafillou showed that the crowd could provide high-quality input in terms of completeness and correctness that leads to the emerging of good quality taxonomies.

Nevertheless, the mentioned approaches are only possible if the system already has a large number of tags that can be used to form the term corpus. It is difficult to develop a new taxonomy if there are no tagged contents in the system. Furthermore, the existing tags also needed to cover the topic of the taxonomy that is being developed. If there are mistakes or missing terms, the system administrators need to correct them themselves, which is missing the point of replacing the work of experts in the development of taxonomy.

4 CROWDSOURCING TAXONOMY DEVELOPMENT

In this paper, we want to apply another method to crowdsource taxonomy development and evolution with the support of the KM-EP. From all the taxonomies that were created from a seed taxonomy, the one that has the highest score (highest rated) will be chosen. It will replace the current seed to be used for classification, searching or navigation in the system. Furthermore, it will act as the seed for further expansion of the taxonomy's evolution tree in the next round. Figure 2 presents an example of the evolution of a taxonomy following our approach.

The example from the figure below describes the case where we have a simple taxonomy A as the seed. From this taxonomy, we have different taxonomies (A1, A2, A3, and A4) that were created by different users using crowdsourcing. These taxonomies were rated by the system's users (crowdvoting). Taxonomy A1 had the highest score and became the next official base. Taxonomy A1 was used for content classification, searching, and navigation in the system until next round. In the next round, users cloned taxonomy A1 and updated it as they find suitable. The created taxonomies got rated again by other users, and the taxonomy that had the highest score (A1C) became the next seed. The process repeats as long as the administrators allow it.

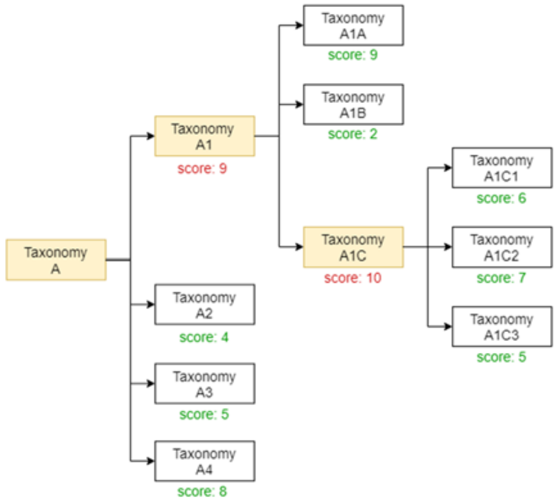

Figure 2: An Example of the Taxonomy Evolution with Support of Crowdsourcing.

To support the evolution process, a version control system based on Git (Git, n.d.) was implemented. This system allows the user to save the current state of the taxonomy, check history with detail about taken snapshots. This is a great method to keep track of taxonomy builds. The crowd is able to identify which version is currently in development, what are the changes etc. Furthermore, the user can reset a taxonomy to a previous state or replace a taxonomy with one of its clones. This is a crucial feature for debugging error, which always happens in the development of taxonomy. Caching mechanisms were also added to increase the processing speed and reduce the computational resource needed. Combining with fast and efficient algorithms, thousands of terms can be retrieved in a matter of milliseconds.

5 EVALUATION

The newly developed Taxonomy Manager prototype was deployed in several R&D projects as part of the KM-EP. The goals in evaluating this prototype are first to evaluate the feasibility, usability, and efficiency of the user experience of the implemented prototype based on user's direct feedback that was collected by means of questionnaires after initially working with the prototype. Secondly, the goal was to evaluate the introduced approach of crowdsourcing taxonomy from a more effectiveness point of view. We want to test if this proof-of-concept can be used successfully in reality.

In the concept of the EU-funded R&D project RAGE (RAGE, n.d.), the first goal of evaluating the feasibility, usability, and efficiency was achieved. Teams from different work packages of the project were working together to create and develop a new taxonomy about the domain of Applied Gaming using the implemented Taxonomy Manager. To evaluate the aspect related to the usage of the Taxonomy Manager, an evaluation questionnaire was created by the authors. Members of the consortium and external game developers as well were contacted by project members to participate in the evaluation of the taxonomy manager. The questionnaire was combined of questions that are related to the usability, usefulness and user interface of the prototype, quality of the tutorial, experience of the participants, quality of the system's features, such as version control, export, and import. The result of this evaluation was published in 2018 by the authors (Vu, et al., 2018). Figure 3 provides an overview of the detailed results obtained from the evaluation categories, with 0 is the lowest score, and 7 is the highest.

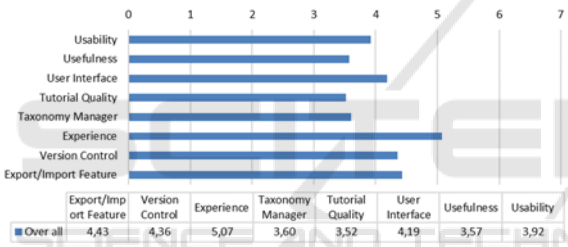

Figure 3: Mean scores of all evaluation categories.

Overall, the scores show that participants, in general, appreciate using the taxonomy manager and its' features, but also that it needed some improvements in the tutorial. Nevertheless, this evaluation does not say anything about the usefulness of the prototype in the sense of the quality of the taxonomy that was created by the users.

Therefore, to achieve the next goal of evaluating the qualitative effectiveness of the tool in terms of the quality of the work on the taxonomy, a second evaluation is now planned. The general concept of this evaluation is to let experts and the crowd do the same task then compare the result and see if the crowd is really doing a similar good or even better job than the experts. In this second evaluation that is presented as a concept in this paper, we chose to use IAB's Quality Assurance Guidelines (QAG) Taxonomy as the initial expert taxonomy. The Interactive Advertising Bureau (IAB) is one of the most influential organizations in the online advertising business and, currently, brings together more than 650 leading companies in the industry that control 86% of the U.S. market. Today IAB has become a standard for content classification, especially in fields with strong ties to the digital economy and new social media (Filippis, 2018). The Quality Assurance Guidelines Taxonomy was created in 2011 by IAB Networks and Exchanges Committee as part of the Quality Assurance Guidelines (QAG) Program. This taxonomy has 2 tiers. The first tier is made of 24 categories, and the second tier has 361 sub-categories. Table 1 presents category "Automotive" as part of the IAB's Quality Assurance Guidelines Taxonomy.

Table 1: Category "Automotive" and its sub-categories.

Automotive	
Auto Parts	Hybrid
Auto Repair	Luxury
Buying/Selling Cars	Minivan
Car Culture	Motorcycles
Certified Pre-Owned	Off-Road Vehicles
Convertible	Performance Vehicles
Coupe	Pickup
Crossover	Road-Side Assistance
Diesel	Sedan
Electric Vehicle	Trucks & Accessories
Hatchback	Vintage Cars
	Wagon

Since then, IAB's Taxonomy and Mapping Working Group have been working on the QAG Taxonomy with the goal of to create an enhanced and more powerful taxonomy, enabling content creators to more accurately and consistently describe content, facilitating more relevant advertising and providing a higher quality and more granular foundation for data analysis (Flood & Agnew, 2017). As a result, a new version of the QAG Taxonomy was introduced in 2017 called "IAB Tech Lab Content Taxonomy Version 2.0". The new Content Taxonomy has more than 400 new site content classifications across 29 tier 1 categories (Content Taxonomy, n.d.). We choose to evaluate this taxonomy as the new version of the expert taxonomy.

For the second evaluation, an experiment guideline for the participants was prepared. In this guideline, information related to the experiment, such as the introduction of taxonomy, goals of the evaluation, introduction of IAB and the QAG taxonomy, was described. Furthermore, the tasks and an example of how they need to be done were also presented. Finally, information on how to report the result was given in the guideline.

In this second evaluation, deliberate participants from the crowd will be given two tasks. In the first task, all the changes that the experts made to upgrade the initial expert taxonomy (Quality Assurance Guidelines Taxonomy) to the new version of the expert taxonomy (IAB Tech Lab Content Taxonomy Version 2.0), such as adding new terms, renaming terms, deleting terms and moving terms into groups are given as a task to each member of the crowd. We ask them to redo each change at a position they find comfort in the initial taxonomy and see if they can re-create the same new taxonomy as the experts did. The purpose of this task is to evaluate the crowd's qualitative performance against the expert taxonomy evolution. This can be considered as a benchmark against the global "expert-based truth".

After the experiment, the result will be collected and analyzed. We compare the similarity between the taxonomy created by the crowd and the new version of the expert taxonomy (Karampinas & Triantafillou, 2012). Let S_1 be the set of all parent-children pairs in the expert taxonomy and S_2 be the set of all parent-children pairs in the crowd taxonomy. We have:

$$Precision = \frac{|S_1 \cap S_2|}{|S_2|} \tag{1}$$

$$Recall = \frac{|S_1 \cap S_2|}{|S_1|} \tag{2}$$

$$F = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

Precision measures the exactness of the crowd taxonomy evolution, and recall is a measure of the completeness of the crowd taxonomy evolution while F-score measures the accuracy of the evaluation. The result will lead to the answer to the question "are the crowd's taxonomy evolution actions as good as those of the experts?".

In the second task, we show both initial taxonomy and the new taxonomy of the expert to each member of the crowd and ask them to answer some questions, such as "what has been changed", "do you agree", "if not, what would you change and why". From the result, we take the taxonomies that the crowd made in the last question and give it to another group of user and experts along with the initial taxonomy and new taxonomy created by IAB. We let the group vote for the best taxonomy and see if it is the one created by the crowd or the experts. In this task, we hope to be able to show which group provided a better taxonomy and validate if the crowd is truly better than the experts.

It is worth to mention that the second evaluation that we described above has not been completed and is considered as future work.

6 CONCLUSIONS AND OUTLOOK

In this paper, we have described the concept of knowledge as well as knowledge management and the content and knowledge management ecosystem portal KM-EP. Furthermore, we presented crowdsourcing and new approach of applying crowdsourcing in the development of taxonomy.

In result, a taxonomy management system was implemented as a component of the KM-EP. The new component allows the crowd to create and manage taxonomy and its structure. The Delphi method was replaced by crowdsourcing and crowdvoting, where users have the ability to vote for each taxonomy. With the support of version control, taxonomy evolution will be faster, more efficient and agile.

Finally, an evaluation was conducted in the scope of the EU-funded project RAGE. This evaluation validates if the implemented prototype fulfils all the requirements and how it performs. The outcome proved the importance, usefulness and usability of the implemented taxonomy management system.

Another evaluation aiming at a qualitative comparison of expert-based taxonomy evolution with crowd-based taxonomy solution was described and planned. Due to time limitation, only a small set of the crowd can be organized, but a big enough and diverse enough crowd can be gathered in the near future for a better evaluation result. This might be done by using the user-base of the RAGE KM-EP, which is growing by the success of the project.

REFERENCES

Andro, M. (2018). *Digital Libraries and Crowdsourcing (Computer Engineering: Digital Tools and Uses Set).* Wiley-ISTE.

Bar, A. R., & Maheswaran, M. (2014). *Confidentiality and Integrity in Crowdsourcing Systems.* Springer.

Brabham, D. C. (2013). *Crowdsourcing.* The MIT Press.

Brooks, C. H., & Montanez, N. (2006). Improved Annotation of the Blogosphere via Autotagging and Hierarchical Clustering. *Proceedings of the 15th international conference on World Wide Web* (pp. 625-632). ACM.

Content Taxonomy. (n.d.). Retrieved from https://www.iab .com/guidelines/taxonomy/

Dalkey, N., & Helmer, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*, 458-467.

Davenport, T. H., & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know.* Harvard Business Press.

Eskenazi, M., Levow, G.-A., Meng, H., Parent, G., & Suendermann, D. (2013). *Crowdsourcing for Speech Processing: Applications to Data Collection, Transcription and Assessment.* Wiley.

Estellés-Arolas, E., & Guevara, F. G. (2012). Towards an Integrated Crowdsourcing Definition. *Journal of Information Science*.

Filippis, L. D. (2018, March 19). *Updated version of the IAB model in the Deep Categorization API*. Retrieved from Meaning Cloud: https://www.meaningcloud.com/ blog/updated-iab-model-meaningcloud

Flood, K., & Agnew, N. (2017, November 30). *IAB Tech Lab Announces Final Content Taxonomy v2 Ready for Adoption*. Retrieved from IAB Tech Lab: https://iabtechlab.com/blog/iab-tech-lab-announces-final-content-taxonomy-v2-ready-for-adoption/

Git. (n.d.). Retrieved from https://git-scm.com/

Hedden, H. (2010). *The Accidental Taxonomist.* Information Today Inc.

Heymann, P., & Garcia-Molina, H. (2006). *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems.* InfoLab Technical Report.

Hodge, G. (2000). Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. *Council on Library and Information Resources.*

Jimenez-Crespo, M. A. (2017). *Crowdsourcing and Online Collaborative Translations: Expanding the Limits of Translation Studies.* John Benjamins Publishing Co.

Karampinas, D., & Triantafillou, P. (2012). Crowdsourcing Taxonomies. *The Semantic Web*, 545-559.

Kitchens, F., & Crane, C. (2014). Customersourcing: to Pay or be Paid. *27th Bled eConference.* Bled, Slovenia.

Kon, H., & Hoey, M. (2005). Leveraging collective knowledge. *The 14th ACM international conference on Information and knowledge management* (pp. 560-567). Bremen, Germany: ACM New York, NY, USA.

Lambe, P. (2007). *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness.* Chandos Publishing.

Li, X., Wang, H., Yin, G., Wang, T., Yang, C., Yu, Y., & Tang, D. (2012). Inducing Taxonomy from Tags: An Agglomerative Hierarchical Clustering Framework. *International Conference on Advanced Data Mining and Applications* (pp. 66-77). Springer.

Liu, K., Fang, B., & Zhang, W. (2010). Ontology Emergence from Folksonomies. *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1109-1118). Toronto, ON, Canada: ACM New York, NY, USA.

Peters, I., & Stock, W. G. (2007). Folksonomy and Information Retrieval. *Proceedings of the American Society for Information Science and Technology*, (pp. 1-28).

Plangprasopchok, A., Lerman, K., & Getoor, L. (2010). Growing a Tree in the Forest Constructing Folksonomies by Integrating Structured Metadata. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 949-958). ACM.

RAGE. (n.d.). Retrieved from http://rageproject.eu/

Ridge, M. (2014). *Crowdsourcing our Cultural Heritage.* Ashgate Publishing, Ltd.

Sanderson, M., & Croft, W. B. (1999). Deriving Concept Hierarchies from Text. *Proceeding SIGIR '99 Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 206-213). ACM.

Simon, F. J., Pechuan, I. G., & Estelles-Miguel, S. (2015). *Advances in Crowdsourcing.* Springer International Publishing AG.

Surowiecki, J. (2005). *The Wisdom of Crowds.* Anchor.

The Four V's of Big Data. (2005). Retrieved from IBM Big Data & Analytics Hub: http://www.ibmbigdatahub. com/infographic/four-vs-big-data

Turban, E., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2015). *Electronic Commerce: A Managerial and Social Networks Perspective.* Springer.

Vu, B. (2015). Realizing an Applied Gaming Ecosystem - Extending an Education Portal Suite towards an Ecosystem Portal. *TU Darmstadt*.

Vu, B., Mertens, J., Gaisbachgrabner, K., Fuchs, M., & Hemmje, M. (2018). Supporting Taxonomy Management and Evolution in a Web-based Knowledge Management System. *HCI 2018.* Belfast, UK.

Wikipedia. (n.d.). Retrieved from https://en.wikipedia.org/ wiki/Wikipedia