Semantic Web and Human Computation
The status of an emerging field

Sabou, Marta; Aroyo, Lora; Bontcheva, Kalina; Bozzon, Alessandro; Qarout, Rehab K.

DOI
10.3233/SW-180292

Publication date
2018

Document Version
Accepted author manuscript

Published in
Semantic Web: interoperability, usability, applicability

Citation (APA)
Sabou, M., Aroyo, L., Bontcheva, K., Bozzon, A., & Qarout, R. K. (2018). Semantic Web and Human Computation: The status of an emerging field. *Semantic Web: interoperability, usability, applicability*, *9*(3), 291-302. https://doi.org/10.3233/SW-180292

Semantic Web 0 (0) 1
IOS Press

Semantic Web and Human Computation: the Status of an Emerging Field

Marta Sabou [a,*], Lora Aroyo [b] Kalina Bontcheva [c] Alessandro Bozzon [d] and Rehab K. Qarout [c]
[a] *Faculty of Informatics, Technical University of Vienna, Vienna, Austria*
E-mail: marta.sabou@ifs.tuwien.ac.at
[b] *Computer Science Department, Vrije Universiteit Amsterdam, The Netherlands*
E-mail: lmaroyo@gmail.com
[c] *Department of Computer Science, University of Sheffield, Sheffield, United Kingdom*
E-mails: k.bontcheva@sheffield.ac.uk, rkqarout1@sheffield.ac.uk
[d] *Software Technology Department, Delft University of Technology, Delft, The Netherlands*
E-mail: a.bozzon@tudelft.nl

Abstract.
This editorial paper introduces a special issue that solicited papers at the intersection of Semantic Web and Human Computation research. Research in that inter-disciplinary space dates back a decade, and has been acknowledged as a research line of its own by a seminal research manifesto published in 2015. But where do we stand in 2018? How did this research line evolve during the last decade? How do the papers in this special issue align with the main lines of work of the community? In this editorial we inspect and reflect on the evolution of research at the intersection of Semantic Web and Human Computation. We use a methodology based on *Systematic Mapping Studies* to collect quantitative bibliographic data which we analyze through the lens of research topics envisioned by the research manifesto to characterize the evolution of research in this area, thus providing a context for introducing the papers of this special issue. We found evidences of a thriving research field; while steadily maturing, the field offers a number of open research opportunities for work where Semantic Web best practices and techniques are applied to support and improve the state-of-the-art in Human Computation, but also for work that exploits the strength of both areas to address scientifically and societally relevant issues.

Keywords: Semantic Web, Human Computation, Crowdsourcing

1. Introduction

In 2015, a research manifesto [1] proposed a roadmap for research at the intersection of the Semantic Web and Crowdsourcing research areas, advocating the existence of ample synergies between these two research fields that need to be exploited. The manifesto and the general enthusiasm for this line of research motivated us to organize a Special Issue as an outlet for publishing papers at the intersection of Semantic Web research and the broader area of Human Computation and Crowdsourcing (HC&C).

The goal of this editorial is to convey a picture of how this line of research has evolved over the past decade, and especially during the three years since the publication of the manifesto. This is performed in two ways. On the one hand, we aim to provide a broad and quantitative view of the field by performing an analysis of the scientific literature in this area published in the last decade (2008-2018) in Section 3. On the other hand, we briefly present the four papers published in this special issue and position them in the broader context of research in Section 4. We conclude in Section 5 with lessons learned from our analysis and discuss outstanding open challenges that could be pursued in this exciting area of research.

*Corresponding author. E-mail: marta.sabou@ifs.tuwien.ac.at.

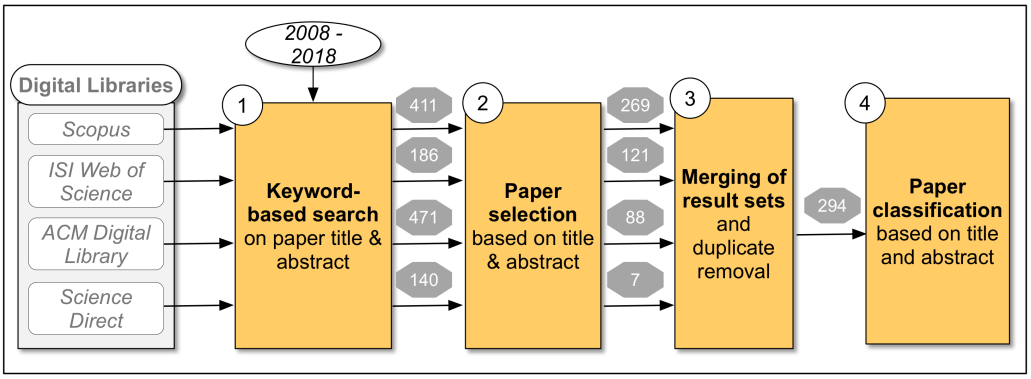


Fig. 1. Main steps of the Systematic Mapping Study and their outputs in paper numbers.

2. Synergies of Semantic Web and Human Computation Research

There exist several synergies between the fields of Semantic Web and Human Computation that open up a number of avenues for research [1].

Stemming from its original motivation of extending the Web with a layer of semantic representation [2, 3], the Semantic Web (SW) aims to solve a set of complex problems that computers cannot yet fully master. Examples include the creation of conceptual models (e.g., ontologies), the semantic annotation of various media types, or entity linking across Linked Open Datasets and Knowledge Graphs. As a result, the large-scale deployment of Semantic Web technologies often depends on the availability of significant human contribution. Such contributions are traditionally provided by experts – e.g. ontology engineers to build ontologies, or annotators to create the semantic data or to link between the instances of various data sets.

Human Computation (HC) methods leverage human processing power to solve problems that are still difficult to solve by using solely computers [4], and therefore are well-suited to support Semantic Web research especially in those areas that still require human contributions. For example, HC methods could be used to create training data for advanced algorithms or as means to evaluate the output of such algorithms. However, in order to increase the accuracy and efficiency of data interpretation at scale, increasingly algorithms (machines) and human contributions are brought together in a natural symbiosis [5]. Such synergy is often performed as iterative interactions, also known as the Human-in-the-Loop paradigm. In this paradigm the user has the ability to influence the outcome of the machine process by providing feedback on different opinions, perspectives and points of views. Additionally, this paradigm contributes to increasing the explainability and transparency of Artificial Intelligence results.

While HC methods could theoretically involve only small numbers of contributors, *crowdsourcing* approaches, leverage the “wisdom of the crowd" by engaging a high number of online contributors to accomplish tasks that cannot yet be automated, often replacing a traditional workforce such as employees or domain experts [6]. As such, crowdsourcing methods not only support the creation of research relevant data, but more importantly they can also help to solve the bottleneck of knowledge experts and annotators needed for the large-scale deployment of Semantic Web and Linked Data technologies.

The potential benefits at the intersection of Semantic Web and Human Computation fields were already discussed in 2015 [1], where two main possible research branches where identified and documented.

On the one hand, HC&C offers promising techniques to solve typical Semantic Web tasks. We refer to this branch as *HC&C for Semantic Web (shortly, HC4SW)*. Two scenarios were envisioned in [1] as typical for the HC4SW research line, as follows:

- *Ontology Engineering and Knowledge Base Curation*: it concerns the acquisition of knowledge structures (e.g., ontologies, knowledge bases, knowledge graphs) through a number of tasks such as defining classes and their hierarchies, identifying relations, extending ontologies with instances, labels, documentation and metadata.
- *Validation and Enhancement of Knowledge*: it covers tasks that aim to improve the quality of semantic data sources by “analyzing, verifying,

Fig. 2. Overlap of relevant paper sets collected from the four digital libraries.

correcting or extending" [1] selected aspects of knowledge structures.

On the other hand, *Semantic Web technologies could support HC&C* research (*SW4HC*) in one of the following ways:

- *Knowledge Representation*: using ontologies to provide semantic representations of the data and knowledge in HC&C systems.
- *Data Integration*: formally represented knowledge could enable easier data integration, especially with data sets that could augment and extend the data of the HC&C systems.
- *Automatic Reasoning*: semantics can be used to perform a range of automated reasoning tasks, e.g., for automating the verification of collected data or for generating automatic feedback to the human contributors.

As this special issue marks a decade of research at the intersection of the Semantic Web and Human Computation research areas, in the next section we investigate in a quantitative study how the research in this area evolved over time.

3. Insights into a Decade of Research

To provide a broader view of the interaction between the research areas of Semantic Web on the one hand, and Human Computation on the other, we performed a bibliographic analysis of research published in the last decade (2008-2018).

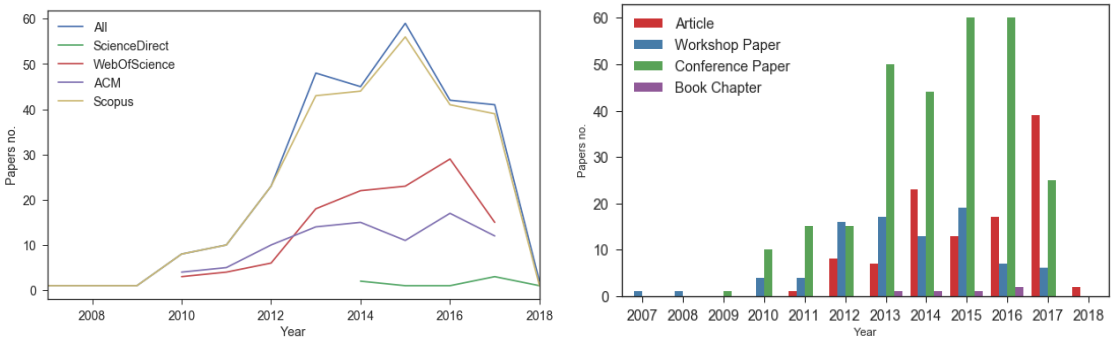
We address four major digital libraries: ACM Digital Library (ACM), Scopus, Science Direct (SciDir) and ISI Web of Science (WebScie). The literature search was based on a methodology inspired from *Systematic Literature Studies*, which are broadly adopted in social science and in software engineering [7]. More precisely, we followed a variant of this method, namely a *Systematic Mapping Study* [8], which is more adequate in endeavours for addressing broader research questions, such as mapping (the evolution of) topics in a research area. As our study is not an in-depth survey, we only focused on the first stages of the Mapping Study method concerned with finding and selecting relevant papers. The concluding stages of the methodology focus on detailed data collection, but were not performed as they were beyond the scope of this study.

Our aim was to complement the manifesto of Sarasua at el. (2015) by providing quantitative insights into how the research topics envisioned by the manifesto actually evolved. Therefore, our research questions are related to the volume, evolution and main lines of research addressed by the community in the last decade. Accordingly, we devised a search query which identified all papers for which either the title or the abstract (or both) contained a combination or terms from the two research areas. As keywords representative for the Semantic Web research area we chose: *semantic web, ontolog*, linked data, knowledge base, knowledge graph*. Terms for HC&C included: *crowdsourc*, human computation, human-in-the-loop*. The search query took the following form:

("semantic web" OR ontolog OR "linked data" OR "knowledge base" OR "knowledge graph") AND (crowdsourc* OR "human computation" OR human-in-the-loop)*

Our methodology for collecting relevant papers is depicted in Figure 1 and included the following steps:

1. **Keyword-based search** on the four digital libraries returned a total of 1208 papers, distributed over the main digital libraries as shown in Figure 1.
2. **Paper Selection.** We manually filtered each result set and determined whether the returned papers were relevant for our search by judging from their title, keywords and abstract. The selection was performed by two researchers to reduce bias. This resulted in 488 relevant papers.



(a) paper numbers (b) paper types

Fig. 3. Research evolution over time in terms of (a) paper numbers published in digital libraries and (b) paper types.

3. **Merging of result sets.** The individual result sets from Step 2 were merged in order to remove duplicates and lead to 294 papers. Figure 2 depicts a Venn Diagram with the intersection of relevant paper sets returned by the four digital libraries. Scopus had the best coverage of the research area of interest, but all other libraries contributed papers that were not found with Scopus or any other library. This result shows that search in several digital libraries is justified to obtain a high recall of the relevant literature.
4. **Paper classification**. Several classification stages followed each focusing on classifying papers according to different criteria such as (1) the type of paper (workshop paper, conference paper, journal); (2) the research community where the paper was published (see details in Section 3.2) as well as (3) the research topic addressed by the paper in terms of the scenarios defined by [1] - see details in Section 3.3.

3.1. Evolution over Time

Figure 3(a) shows the number of published papers per years and per digital library as well as the merged data ("All"). According to the merged data a peak of this research was reached in 2015, while Web Of Science and ACM show this peak for 2016 instead. There is a decline in 2017, but this could be still due to delays in indexing 2017 events.

Figure 3(b), on the other hand, shows the number of different paper types per years. Besides confirming the peak in terms of paper volumes in 2015 and 2016, this figure provides an additional insight on how the community is moving from publishing initial ideas in workshop and conference papers towards publishing mature research in journal articles in 2017 and 2018. This hints to the research field undergoing a process of becoming more mature.

3.2. Community Analysis

An interesting side effect of our methodology of performing a broad search, is that we have the possibility to also investigate the main research communities that publish research combining Semantic Web and HC&C. We considered the following communities:

- **Bio** for venues related to bioinformatics and medical information systems.
- **CS** for computer science and (management) information systems venues.
- **Eng** for venues related to software engineering and data engineering.
- **HCI** for human computer interaction and human computation venues.
- **NLP** for venues related to natural language processing and text processing.
- **SW** for Semantic Web venues.
- **WWW** for world wide web research venues.

Figure 4 shows a broad spectrum of research communities that publish the research of interest. Indeed, 30% of all papers we retrieved were published in Semantic Web venues. Semantic Web venues represent the cradle for the start of this research line and constitute the core publication venue till 2014-2015, after which this research seems to spread into other communities, in general computer science venues, as well as more specialized fields such as bioinformatics, NLP or data and software engineering. Interestingly, this re-

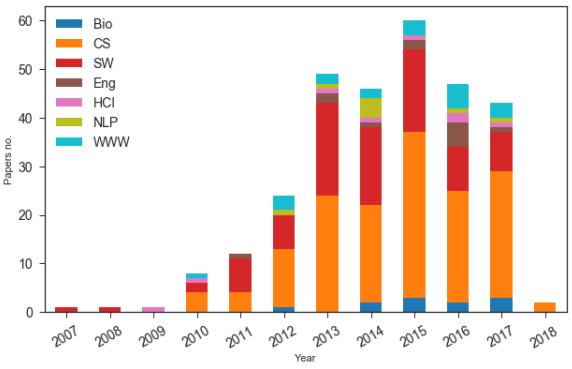


Fig. 4. Distribution of papers across research communities.

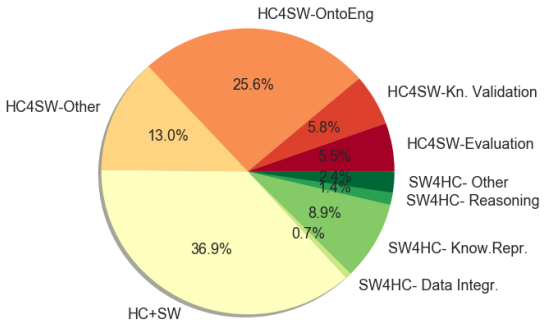


Fig. 5. Distribution of papers in terms of the main topic addressed.

search line is weakly represented in venues related to human computation and human computer interaction. Only 2.7% of the papers from our collection were published in HCI venues in the last decade.

3.3. Topic Analysis

As envisioned by [1], the identified research papers fall primarily into two categories: the largest subset of the papers (146) showcase the use of HC&C as solution (parts) for typical Semantic Web tasks (HC4SW) while 41 papers investigate how Semantic Web techniques could support some aspect of HC&C systems (SW4HC). We have identified also a third category of papers (107), which combine both Semantic Web and HC&C techniques in order to support a third task from a different research area or application domain (HC+SW). Figure 5 shows the distribution of papers into these three major research categories, as well as their subtopics, while in the next section we briefly discuss each category in turn.

3.3.1. Human Computation for Semantic Web (HC4SW)

A first group of papers investigates how HC&C techniques can be used to solve a variety of tasks relevant for the Semantic Web. Within this category, Sarasua et al. [1] distinguished approaches that collect new data through HC&C to build ontologies and knowledge bases (*HC4SW-OntoEng*). We found a total of 75 papers in this category, which cover context-aware knowledge acquisition on mobile devices [9], socio-technical systems that support communities, such as the Paleoclimate community, to develop and extend a community ontology in a collaborative effort [10]. Lou et al. focus on the crowdsourced-acquisition of more complex knowledge structures, namely sanctioning rules in a use case related to the International Classification of Diseases (ICD-11) medical standard [11].

Sixteen papers use crowds to validate and enhance existing knowledge (*HC4SW-Kn.Validation*). For example, ul Hassan et al. [12] focus on the quality assessment of Linked Data and propose a method for selecting suitable crowd-workers for this task. Mortensen et al. show how crowdsourcing can be successfully used to verify large-scale medical ontologies such as SNOMED CT [13].

We observed the emergence of a new category of 15 papers from the area of Semantic Web research, where crowdsourcing is used as a means to support Semantic Web research evaluation (*HC4SW-Evaluation*). Other communities, such as the NLP community, routinely use crowdsourcing for key stages in the scientific process and especially for evaluating the results produced by newly developed algorithms [14]. Some examples of papers which use crowdsourcing for evaluating the results of new Semantic Web approaches or algorithms are as follows. Potoniec et al. propose an algorithm that extracts `SubClassOf` axioms from Linked Data sources and verify the correctness of the extracted axioms through crowdsourcing [15]. Kliegr et al. evaluate their entity typing algorithm on a crowdsourced gold-standard data set of 2000 entities aligned with their corresponding types from the DBpedia ontology [16]. Note that here we only report on papers that made it clear already in their abstract that crowdsourcing is used for evaluation purposes, but we expect that this category of papers is much larger as it includes also papers that do not mention their evaluation approach in their abstract and were therefore not retrieved by our keyword-based search approach.

We could not categorize 40 papers in either of these three categories (*HC4SW-Other*). Examples are works

on vertical topics relevant for a variety of scenarios such as capturing disagreement with the CrowdTruth framework [17] or works that cover both ontology creation and knowledge validation, such as the uComp Protégé Plugin [18].

3.3.2. Semantic Web for Human Computation (SW4HC)

While Semantic Web research benefits significantly from Human Computation research, there is also a trend of applying Semantic Web techniques to support HC&C, as done by 41 papers in our study.

As predicted by Sarasua et al. in 2015, firstly, 24 papers report on benefiting from the knowledge representation capabilities of Semantic Web technologies (*SW4HC-Know.Repr.*). For example, ontologies of tasks allow improved participant selection in mobile crowdsourcing settings [19] and semantic descriptions of workflows facilitate the crowdsourcing of a constitution [20]. Another line of work focuses on describing the workers, their CVs and skills [21–23].

To a lesser extent we found papers where ontologies supported data integration, for example, in the healthcare [24] and multimedia processing [25] domains (*SW4HC-DataIntegr.*). Automated reasoning on formally represented knowledge is harnessed (*SW4HC-Reasoning*) in order to optimize the collection of missing values with crowdsourcing [26] or to validate the quality of data collected through crowdsourcing [27].

In some papers, Linked Data technology enabled openly publishing data collected through crowdsourcing [28] or data from crowd-sourced experiments in an effort to support reproducibility of research [29]. This use of Linked Data was already foreseen by Sarasua et al. in 2015 [1] and our literature search found concrete realisations of this line of work.

3.3.3. Combining Semantic Web and Human Computation (SW+HC)

Going beyond Sarasua et al.'s manifesto [1], our search also retrieved a substantial number of papers in which the two research areas were used in combination to solve a problem from another research area or to create more complex solutions that address scenarios from a variety of application domain. The papers in this category showcased the combined use of the two technologies in very diverse settings ranging from citizen science to security, as shown by a few representative examples we mention next.

In the area of *citizen science* a semantic wiki is used to collect community provided annotations of the zebrafish gene [30]. *Crisis management* and post-disaster recovery is also a frequently addressed topics such as in [31].

In the area is *Smart Cities* there are initiatives that create Linked Data based collection of citizen complaints [32] or collect and integrate urban data to support city planning [33]. Mobile crowdsourcing is the basis of several papers which deal with *geo-data*, for example, the verification and extension of geo data collections such as the OpenStreetMap dataset [34] or enabling collaborative ontology construction in crowdsourced cartographic projects [35]. Hu et al. describe the combination of these two technologies for recommendation based systems that support *personal health management* [36].

In the area of *cultural heritage*, crowdsourcing was instrumental in the semantic annotation of visual artworks [37, 38]. Even in the area of *security* there are examples of how these two technologies can be combined, for example, in order to enable the creation of rules for detecting malicious software [39].

3.3.4. Limitations and Threats to Validity

The analysis presented in this section is meant to give an *indicative* insight into the evolution of research. We are aware of the following limitations.

Related to the *recall* of all relevant papers (i.e., the coverage of the study data set) this could be further improved by (1) selecting more keywords for our queries; (2) querying additional bibliographic sources or (3) adding relevant papers known to the authors but which were not retrieved for any number of reasons. For instance, there could be papers not indexed by the digital libraries; papers that do not mention the search-query keywords in their title/abstract; or simply papers that were omitted during paper selection.

We are also aware that the *precision* of the paper categorization process in the various topics could have been affected by the fact that it was performed based on paper abstracts only. This categorization could be more precise if papers were read in detail, but this step was outside the scope of our study. In fact this step can only be performed for a smaller set of study papers selected for a focused topic, but here we aimed to capture the breadth of research even if at some expense of the precision.

The categorization of the papers was sometimes hampered by the fact that the distinctions between research categories were not clear-cut. Also, we could have looked at the papers from a different perspective than the scenarios defined by Sarasua et al. [1].

All these aspects should be considered by any follow up studies which aim to create precise, in-depth surveys of (selected aspects) of this research area.

4. Special Issue Papers

This special issue attracted a total of 10 submissions from which four papers [40–43] were accepted for publication as summarized in the next sections.

4.1. Detecting Linked Data Quality Issues via Crowdsourcing: A DBpedia Study

This paper focuses on the problem of verifying the quality of Linked Data, in particular data from DBpedia [40]. As such it is illustrative of the scenario, where HC&C is used for knowledge validation and enhancement (*HC4SW-Kn.Validation*).

The authors observe that several of the quality issues frequent in DBpedia, which cannot be reliably detected automatically, can be identified with human involvement. The study focuses on verifying four types of quality issues frequent in DBpedia triples, related to (1) incorrect object values in a triple, (2,3) incorrect data types or language tags and (4) incorrect links.

The paper investigates three main research questions, referring to (1) whether and to what extent these error types can be detected by crowds; (2) how do crowds with diverse skill sets (e.g., experts vs. layman) perform on these tasks and (3) what are optimal workflow designs that combine crowds with these different skill sets in order to maximize accuracy. To investigate their research questions, the authors employ two different crowdsourcing genre: expert contests on the one hand and traditional micro-task crowdsourcing on Amazon Mechanical Turk (AMT) on the other. The Find-Fix-Verify workflow is used in both genre.

The paper provides several interesting lessons. Firstly, by contrasting the HC-based results with state-of-the-art quality assessment tools, it is shown that the majority of errors can only be detected with HC techniques. This provides a good example of a task that currently cannot be reliably automated. Secondly, experiments confirmed that expert and laymen crowds can reliably detect the error types under investigation, each crowd having their own strengths. Thirdly, experiments show that workflows combining and exploring the synergies of crowds with complementary aptitudes (i.e., experts vs. layman crowds) lead to more effective results than when using these crowds in isolation.

4.2. Using Microtasks to Crowdsource DBpedia Entity Classification: A study in Workflow Design

The paper addresses the problem of how human computation could be used to support the typical Semantic Web task of *entity typing* in knowledge bases, with a focus on DBpedia (*HC4SW-OntoEng*) [41]. Knowledge bases such as DBpedia are becoming an important asset for scientists and practitioners, but suffer from a number of flaws that could be traced back to missing or factually wrong information.

The authors investigate how the contribution from workers operating in microwork platforms could be organised to select the entity type (e.g. company, device, food) from a tree of hierarchically organised classes. As a real-world hierarchy could easily contain thousands of classes, there exists a fundamental trade-off between the precision that could be obtained by automatic systems, and the cost of engaging experts.

The paper contributes an analysis of the main design dimension that affect the design of human-enhanced workflows that include both automated and crowdsourced components, and reports on their performance in terms of precision (in terms of correctness of entity typing) and cost (in terms of amount of required manual work). Workflows include three main steps: 1) a prediction step, where a list of candidate classes for a given entity is generated (automatically, or from the crowd); 2) an error detection step, where the output is manually checked, and 3) an error correction step. The authors focus on three types of workflows, where the main variations affect the prediction step.

Experiments were conducted on 120 untyped DBpedia entities, and have demonstrated the intrinsic complexity of the entity typing problem. Even when humans are involved, three main issues seem to affect the classification precision: 1) the (lack-of) domain-specific expertise of crowd workers; 2) the unbalanced structure of the type hierarchy; and 3) the ambiguity of some entities. Results clearly indicate the need for further investigation, in terms of both workflow design and optimization strategies.

4.3. An Extended Study of Content and Crowdsourcing-related Performance Factors in Named Entity Annotation

This paper addresses an important problem related to named entity recognition (NER) performed on noisy social media microposts, e.g., tweets (*HC4SW-OntoEng*) [42]. The basic assumption of the authors is

that some types of social media microposts are more amenable to crowdsourcing than others.

In order to prove their hypothesis the authors study the impact of the micropost content on the accuracy of human annotations. For this, experiments were performed using a game with a purpose for NER called Wordsmith which sourced workers from the CrowdFlower crowdsourcing platform. Four datasets of microposts were used in these experiments (Ritter Corpus 2010, Finin Corpus 2008, MSM 2013 Corpus and Wordsmith Corpus 2014), i.e. two experiments per dataset evaluating a total of 7665 tweets.

Two research questions and two hypotheses guided these experiments. On the one hand, the authors investigated what is the effect of micropost features on the accuracy and speed of entity annotation performed by non-expert crowd workers. Authors measured the number and type of entities recognized, as well as the length and sentiment of the post. On the other hand, the authors also investigated whether crowd workers prefer some NER tasks over others. Specifically, they measured the number of skipped annotations, the precision of the annotation, the time spent and the overall user interface interaction.

The experimental investigations confirmed that features such as micropost length, number and type of mentioned entities are good indicators of how well crowds will perform NER on posts: shorter posts with less entities are more often correctly annotated than longer posts with more entities, while crowd-workers perform better at identifying entities of type person and location in comparison to identifying organizations or miscellaneous entities. This work on better characterizing which posts are amenable for processing with HC paves the way to building hybrid human-machine NER workflows where each post is assigned to either the human or machine component of the system based on its characteristics.

4.4. Empirical Methodology for Crowdsourcing Ground Truth

This paper focuses on the problem of gathering ground truth data for information extraction tasks through Human Computation, and specifically the problem of evaluating the quality of such corpora when (1) there is ambiguity in the data, which typically triggers a multitude of opinions in its interpretation, and (2) when experts are scarce to perform the annotation task [43] (*HC4SW-Other*).

The authors present an empirically derived methodology for gathering of ground truth data, validated in a number of domains and annotation tasks (e.g. medical relation extraction, Twitter event identification, news event extraction and sound interpretation). Interesting in this approach is the use of the CrowdTruth quality metrics. These show that (1) measuring inter-annotator disagreement and (2) a sufficiently high number of crowd workers are both essential for acquiring a robust high quality ground truth, as opposed to mainstream approaches, such as majority vote when employing just a few annotators.

The paper offers two major conclusions. First, it shows that ambiguity-aware quality metrics perform better than consensus-enforcing metrics, and at least as well as domain experts. Second, experimental results prove that this methodology for aggregating crowdsourcing annotations works both for open and closed task types.

5. Conclusions

Based on our investigation of a decade of papers at the intersection of Semantic Web and Human Computation, as well as the papers in this special issue, we draw the following conclusions on the evolution of this inter-disciplinary research area.

5.1. Overall Trends

- *A maturing field*: while there is some evidence of a decline in the number of papers published in 2016/2017, the overall maturity of the work increases as paper types move from primarily workshop and conference papers, to journal articles. Proof to this is also the number of 10 papers submitted to this special issue.
- *Expanding to other research communities*: the Semantic Web venues were the cradle of this research, hosting 30% of all papers. We observe however an increasing number of papers published in venues of other research communities, especially those that benefit from the combination of the Semantic Web and Human Computation approaches. Research is published in general computer science venues, as well as in venues of specialized communities, such as NLP, Bioinformatics, or data and software engineering. Surprisingly, this line of research is weakly represented in venues related to Human Computation and Human Computer Interaction.

- *An asymmetric relation between the two research fields* was identified, with Human Computation research being more strongly adopted in the Semantic Web community than the other way around. Indeed, from the collected research papers, far more papers investigate the use of HC for Semantic Web research (HC4SW) than using Semantic Web for enabling Human Computation tasks (SW4HC). The SW4HC papers primarily focused on exploring the use of semantics for knowledge representation, while the use of these technologies to support data integration and reasoning was only addressed to a limited extent. We believe this to be a promising avenue for future research. For instance, recent HC work focusing on the analysis of task properties (e.g. complexity [44] and clarity [45]) and on task recommendation [46] could benefit from the adoption of Semantic Web approaches for knowledge representation and named entity linking. We also identified initial work on using Linked Data to publish research results, in order to support research reproducibility [28, 29, 47, 48] which we hope will be adopted on a larger scale by the community.
- *The emergence of a combined use of Semantic Web and Human Computation.* Our search found a large number of papers which do not necessarily use one of the research areas to support the other, but rather use these two areas in combination (i.e., as parts of the same larger system or approach) to support a task or application from another research community.

5.2. Trends in the Papers of this Special Issue

In line with the general trend of research, the papers in this special issue cover work on the use of Human Computation for addressing Semantic Web tasks (HC4SW), mostly within the topic of ontology engineering [41, 42] or knowledge validation [40]. The fourth paper investigates an issue that is relevant for HC4SW work in general, namely, an ambiguity-metric based approach to collect ground truth data [43].

In terms of the research challenges defined by Sarasua et al. in their research manifesto [1], this issue's papers advance the state of knowledge on the following challenges:

- *Task and Workflow Design*: Acosta et. al [40] experiment with several workflows that explore the complementary aptitudes of different crowds harnessed with diverse HC genres (microtasks and games with a purpose). In [41] various workflow designs are proposed for combining human and machine computation in the context of solving the problem of entity typing.
- *Using Multiple Crowdsourcing Genres*. HC genres all have their strengths and weaknesses which open up opportunities for their combined use. For example, in [40] several workflows are described which combine diverse HC genres (i.e., gamification and micro-task crowdsourcing) to reach a better performance than appraoches relying on a single genre. An example of a scenario where GWAP players are sourced from CrowdFlower is provided in [42].
- *Managing Hybrid Workflows* which combine algorithmic and human computation techniques is also a popular topic. Bu et al. [41] study the performance of several workflow designs which combine human and machine components. The work presented in [42], paves the way towards creating machine-human workflows in the area of NER on noisy social media data.
- *Quality of Contributions*. Collecting ground-truth information is essential for evaluating and ensuring the quality of contributions collected in HC systems. This is a core issue that is of interest to approaches focusing on both the ontology engineering or knowledge validation research axes. Dumitrache et al. bring a contribution to this topic by showing that the CrowdTruth ambiguity-aware quality metrics perform better than more traditional consensus-enforcing metrics (as used for example in the other papers of this special issue), and at least as well as domain experts [43]. Their proposed methodology for aggregating crowdsourcing annotations while being aware of disagreements between workers, was shown to be effective for both open and closed task types.

5.3. Open Challenges and Future Work

Our search revealed a high number of very diverse papers at the intersection of Semantic Web and Human Computation research, yet no focused surveys of this area. Therefore, this line of research could benefit from a (series) of in-depth surveys covering, for example, one of the three research branches identified (HC4SW, SW4HC and SW+HC). One expected benefit of these in-depth surveys is that they could further refine and extend the current set of topics and sce-

narios envisioned for this line of work by Sarasua et al. [1]. For instance, we identified emerging clusters of papers around the topics such of using HC as support for evaluating Semantic Web research (*HC4SW-Evaluation*) or relying on Linked Data as a technology for openly publishing research data.

In the area of using Human Computation for Semantic Web research (HC4SW), there are a few trending topics both in the overall paper corpus we collected and in the special issue papers. For example, research on workflow design has considered workflows that combine different HC genre [40, 49] as well as hybrid human-machine workflows [41, 42]. The latter type of workflows dovetails with recent efforts to construct Human-in-the-Loop systems and still raises several open research issues as discussed in [41]. There are also interesting efforts to exploring novel interfaces for HC based knowledge acquisition, such as chatbots [9] and aiming to collect more complex knowledge structures (e.g., rules) [11].

Last, but not least, to lower the overhead in adopting and using HC in SW, there is a need for reusable tools and user interfaces for common Semantic Web tasks (e.g. ontology learning, entity linking), and vice versa - tools, ideally integrated with major crowdsourcing platforms, that help researchers utilize ontologies and semantic annotations, as part of defining the Human Computation tasks and projects (as part of the SW4HC branch). One such example from the area of Natural Language Processing is the open-source GATE Crowdsourcing plugin [50], which offers infrastructural support for mapping documents to crowdsourcing units and back automatically, as well as automatically generating reusable crowdsourcing interfaces for NLP classification and selection tasks. Initial work in this direction within the Semantic Web area has been done as part of the uComp Protégé plugin [18] for supporting a range of ontology engineering tasks.

We also found that the adoption of Semantic Web technologies to support Human Computation systems is currently limited and is focused on the formal knowledge representation capabilities of these technologies, but falls short of exploring more advanced capabilities made possible by semantics such as data integration and automated reasoning.

We conclude that, while this special issue reports on important advances on a number of fundamental research challenges, there are ample so far unexplored opportunities for future work in the context in this maturing, diverse and multi-disciplinary research area.

Acknowledgements

This work was partially supported by the FFG funded *CitySPIN* project (project number 861213); by a UK EPSRC grant No. EP/I004327/1; and by the Amsterdam Institute for Advanced Metropolitan Solutions, with the *AMS Social Bot* grant.

References

[1] C. Sarasua, E. Simperl, N. Noy, A. Bernstein and J.M. Leimeister, Crowdsourcing and the Semantic Web: a Research Manifesto, *Human Computation* **2**(1) (2015), 3–17. doi:10.15346/hc.v2i1.2.

[2] T. Berners-Lee, J. Hendler, O. Lassila et al., The Semantic Web, *Scientific American* **284**(5) (2001), 28–37.

[3] B. Glimm and H. Stuckenschmidt, 15 Years of Semantic Web: An Incomplete Survey, *KI-Künstliche Intelligenz* **30**(2) (2016), 117–130. doi:10.1007/s13218-016-0424-1.

[4] A.J. Quinn and B.B. Bederson, Human Computation: A Survey and Taxonomy of a Growing Field, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'11, ACM, New York, NY, USA, 2011, pp. 1403–1412. ISBN 978-1-4503-0228-9. doi:10.1145/1978942.1979148.

[5] G. Demartini, D.E. Difallah, U. Gadiraju and M. Catasta, An Introduction to Hybrid Human-Machine Information Systems, *Foundations and Trends in Web Science* **7**(1) (2017), 1–87. doi:10.1561/1800000025.

[6] J. Howe, The Rise of Crowdsourcing, *Wired Magazine* **14**(6) (2006). http://www.wired.com/wired/archive/14.06/crowds.html.

[7] B.A. Kitchenham and S. Charters, Guidelines for performing Systematic Literature Reviews in Software Engineering, Technical Report, Version 2.3, 2007.

[8] B.A. Kitchenham, D. Budgen and O. Pearl Brereton, Using Mapping Studies As the Basis for Further Research - A Participant-observer Case Study, *Information and Software Technology* **53**(6) (2011), 638–651. doi:10.1016/j.infsof.2010.12.011.

[9] L. Bradeško, M. Witbrock, J. Starc, Z. Herga, M. Grobelnik and D. Mladenić, Curious Cat–Mobile, Context-Aware Conversational Crowdsourcing Knowledge Acquisition, *ACM Trans. Inf. Syst.* **35**(4) (2017), 33–13346, ISSN 1046-8188. doi:10.1145/3086686.

[10] Y. Gil, D. Garijo, V. Ratnakar, D. Khider, J. Emile-Geay and N. McKay, A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations, in: *The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II*, 2017, pp. 231–246. doi:10.1007/978-3-319-68204-4_24.

[11] Y. Lou, S.W. Tu, C. Nyulas, T. Tudorache, R.J.G. Chalmers and M.A. Musen, Use of Ontology Structure and Bayesian Models to Aid the Crowdsourcing of ICD-11 Sanctioning Rules, *J. of Biomedical Informatics* **68**(C) (2017), 20–34, ISSN 1532-0464. doi:10.1016/j.jbi.2017.02.004.

[12] U. ul Hassan, A. Zaveri, E. Marx, E. Curry and J. Lehmann, ACRyLIQ: Leveraging DBpedia for Adaptive Crowdsourcing in Linked Data Quality Assessment, in: *Knowledge Engineering and Knowledge Management*, E. Blomqvist, P. Ciancarini, F. Poggi and F. Vitali, eds, Springer International Publishing, Cham, 2016, pp. 681–696. ISBN 978-3-319-49004-5. doi:10.1007/978-3-319-49004-5_44.

[13] J.M. Mortensen, E.P. Minty, M. Januszyk, T.E. Sweeney, A.L. Rector, N.F. Noy and M.A. Musen, Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT, *JAMIA* **22**(3) (2015), 640–648. doi:10.1136/amiajnl-2014-002901.

[14] M. Sabou, K. Bontcheva and A. Scharl, Crowdsourcing Research Opportunities: Lessons from Natural Language Processing, in: *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW'12, ACM, New York, NY, USA, 2012, pp. 17–1178. ISBN 978-1-4503-1242-4. doi:10.1145/2362456.2362479.

[15] J. Potoniec, P. Jakubowski and A. Lawrynowicz, Swift Linked Data Miner, *Web Semant.* **46**(C) (2017), 31–50, ISSN 1570-8268. doi:10.1016/j.websem.2017.08.001.

[16] T. Kliegr and O. Zamazal, LHD 2.0: A text mining approach to typing entities in knowledge graphs, *J. Web Sem.* **39** (2016), 47–61. doi:10.1016/j.websem.2016.05.001.

[17] O. Inel, K. Khamkham, T. Cristea, A. Dumitrache, A. Rutjes, J. van der Ploeg, L. Romaszko, L. Aroyo and R.-J. Sips, CrowdTruth: Machine-Human Computation Framework for Harnessing Disagreement in Gathering Annotated Data, in: *The Semantic Web – ISWC 2014*, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandecic, P. Groth, N. Noy, K. Janowicz and C. Goble, eds, Springer International Publishing, Cham, 2014, pp. 486–504. doi:10.1007/978-3-319-11915-131.

[18] G. Wohlgenannt, M. Sabou and F. Hanika, Crowd-based ontology engineering with the uComp Protégé plugin, *Semantic Web* **7**(4) (2016), 379–398. doi:10.3233/SW-150181.

[19] J. Wang, Y. Wang, L. Wang and Y. He, GP-selector: a generic participant selection framework for mobile crowdsourcing systems, *World Wide Web* (2017), ISSN 1573-1413. doi:10.1007/s11280-017-0480-y.

[20] N. Luz, M. Poblet, N. Silva and P. Novais, Defining Human-Machine Micro-Task Workflows for Constitution Making, in: *Outlooks and Insights on Group Decision and Negotiation*, B. Kami'nski, G.E. Kersten and T. Szapiro, eds, Springer International Publishing, Cham, 2015, pp. 333–344. doi:10.1007/978-3-319-19515-5-26.

[21] A. Bozzon, M. Brambilla, S. Ceri, M. Silvestri and G. Vesci, Choosing the Right Crowd: Expert Finding in Social Networks, in: *Proceedings of the 16th International Conference on Extending Database Technology*, EDBT'13, ACM, New York, NY, USA, 2013, pp. 637–648. ISBN 978-1-4503-1597-5. doi:10.1145/2452376.2452451.

[22] K.E. Maarry, W.-T. Balke, H. Cho, S.-w. Hwang and Y. Baba, Skill Ontology-Based Model for Quality Assurance in Crowdsourcing, in: *Database Systems for Advanced Applications*, W.-S. Han, M.L. Lee, A. Muliantara, N.A. Sanjaya, B. Thalheim and S. Zhou, eds, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 376–387. doi:10.1007/978-3-662-43984-5-29.

[23] C. Sarasua and M. Thimm, Crowd Work CV: Recognition for Micro Work, in: *Social Informatics*, L.M. Aiello and D. McFarland, eds, Springer International Publishing, Cham, 2015, pp. 429–437. doi:10.1007/978-3-319-15168-7_52.

[24] M. Sohn, S. Jeong, J. Kim and H.J. Lee, Crowdsourced healthcare knowledge creation using patients' health experience-ontologies, *Soft. Computing* **21**(18) (2017), 5207–5221. doi:10.1007/s00500-017-2529-3.

[25] A. Bozzon, P. Fraternali, L. Galli and R. Karam, Modeling CrowdSourcing Scenarios in Socially-Enabled Human Computation Applications, *Journal on Data Semantics* **3**(3) (2014), 169–188, ISSN 1861-2040. doi:10.1007/s13740-013-0032-2.

[26] H.-Z. Wang, Z.-X. Qi, R.-X. Shi, J.-Z. Li and H. Gao, COSSET+: Crowdsourced Missing Value Imputation Optimized by Knowledge Base, *Journal of Computer Science and Technology* **32**(5) (2017), 845–857. doi:10.1007/s11390-017-1768-1.

[27] A. Kaufmann, J. Peters-Anders, S. Yurtsever and L. Petronzio, Automated Semantic Validation of Crowdsourced Local Information – The Case of the Web Application "Climate Twins", in: *Environmental Software Systems. Fostering Information Sharing*, J. Hrebicek, G. Schimak, M. Kubasek and A.E. Rizzoli, eds, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. doi:10.1007/978-3-642-41151-9_3.

[28] V.W. Anelli, A. Cali, T. Di Noia, M. Palmonari and A. Ragone, Exposing Open Street Map in the Linked Data Cloud, in: *Trends in Applied Knowledge-Based Systems and Data Science*, H. Fujita, M. Ali, A. Selamat, J. Sasaki and M. Kurematsu, eds, Springer International Publishing, Cham, 2016, pp. 344–355. doi:10.1007/978-3-319-42007-3_29.

[29] O. Feyisetan, M. Luczak-Roesch, E. Simperl, R. Tinati and N. Shadbolt, Towards Hybrid NER: A Study of Content and Crowdsourcing-Related Performance Factors, in: *The Semantic Web. Latest Advances and New Domains*, F. Gandon, M. Sabou, H. Sack, C. d'Amato, P. Cudre-Mauroux and A. Zimmermann, eds, Springer International Publishing, Cham, 2015, pp. 525–540. doi:10.1007/978-3-319-18818-8_32.

[30] M. Singh, D. Bhartiya, J. Maini, M. Sharma, A.R. Singh, S. Kadarkaraisamy, R. Rana, A. Sabharwal, S. Nanda, A. Ramachandran, A. Mittal, S. Kapoor, P. Sehgal, Z. Asad, K. Kaushik, S.K. Vellarikkal, D. Jagga, M. Muthuswami, R.K. Chauhan, E. Leonard, R. Priyadarshini, M. Halimani, S. Malhotra, A. Patowary, H. Vishwakarma, P.R. Joshi, V. Bhardwaj, A. Bhaumik, B. Bhatt, A. Jha, A. Kumar, P. Budakoti, M.K. Lalwani, R. Meli, S. Jalali, K. Joshi, K. Pal, H. Dhiman, S.V. Laddha, V. Jadhav, N. Singh, V. Pandey, C. Sachidanandan, S.C. Ekker, E.W. Klee, V. Scaria and S. Sivasubbu, The Zebrafish GenomeWiki: a crowdsourcing approach to connect the long tail for zebrafish gene annotation, *Database: the journal of biological databases and curation* **2014** (2014). doi:10.1093/database/bau011.

[31] O. Mejri, S. Menoni, K. Matias and N. Aminoltaheri, Crisis information to support spatial planning in post disaster recovery, *International Journal of Disaster Risk Reduction* **22** (2017), 46–61, ISSN 2212-4209. doi:10.1016/j.ijdrr.2017.02.007.

[32] S. Egami, T. Kawamura, K. Kozaki and A. Ohsuga, Construction of Linked Urban Problem Data with Causal Relations Using Crowdsourcing, in: *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2017, pp. 814–819. doi:10.1109/IIAI-AAI.2017.189.

[33] A. Psyllidis, A. Bozzon, S. Bocconi and C. Titos Bolivar, A Platform for Urban Analytics and Semantic Data Integration in City Planning, in: *Computer-aided architectural design fu-*

tures: New technologies and the future of the built environment: 16th International Conference, CAAD Futures 2015, Sao Paulo, Brazil, July 8-10, 2015. Selected papers, G. Celani, D. Sperling and J. Franco, eds, Spinger, 2015, pp. 21–36. ISBN 978-3-662-47385-6. doi:10.1007/978-3-662-47386-3_2.

[34] R. Karam and M. Melchiori, A Human-Enhanced Framework for Assessing Open Geo-spatial Data, in: *Advances in Conceptual Modeling*, J. Parsons and D. Chiu, eds, Springer International Publishing, Cham, 2014, pp. 97–106. doi:10.1007/978-3-319-14139-8_12.

[35] A. Ballatore and P. Mooney, Conceptualising the Geographic World: The Dimensions of Negotiation in Crowdsourced Cartography, *Int. J. Geogr. Inf. Sci.* **29**(12) (2015), 2310–2327, ISSN 1365-8816. doi:10.1080/13658816.2015.1076825.

[36] H. Hu, A. Elkus and L. Kerschberg, A Personal Health Recommender System incorporating personal health records, modular ontologies, and crowd-sourced data, in: *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 1027–1033. doi:10.1109/ASONAM.2016.7752367.

[37] J. Oosterman, A. Nottamkandath, C. Dijkshoorn, A. Bozzon, G.-J. Houben and L. Aroyo, Crowdsourcing Knowledge-intensive Tasks in Cultural Heritage, in: *Proceedings of the 2014 ACM Conference on Web Science*, WebSci'14, ACM, New York, NY, USA, 2014, pp. 267–268. ISBN 978-1-4503-2622-3. doi:10.1145/2615569.2615644.

[38] J. Oosterman, J. Yang, A. Bozzon, L. Aroyo and G.-J. Houben, On the impact of knowledge extraction and aggregation on crowdsourced annotation of visual artworks, *Computer Networks* **90** (2015), 133–149, ISSN 1389-1286. doi:https://doi.org/10.1016/j.comnet.2015.07.008.

[39] A.C. d. Marchi, A. Gregio and R. Bonacin, Enhancing the Creation of Detection Rules for Malicious Software through Ontologies and Crowdsourcing, in: *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2017, pp. 290–295. doi:10.1109/WETICE.2017.31.

[40] M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, F. Flöck and J. Lehmann, Detecting Linked Data Quality Issues via Crowdsourcing: A DBpedia Study, *Semantic Web Journal* (2018). http://www.semantic-web-journal.net/system/files/swj1293.pdf.

[41] Q. Bu, E. Simperl, S. Zerr and Y. Li, Using microtasks to crowdsource DBpedia entity classification: A study in workflow design, *Semantic Web Journal* (2018). http://www.semantic-web-journal.net/system/files/swj1408.pdf.

[42] O. Feyisetan, E. Simperl, M. Luczak-Roesch, R. Tinati and N. Shadbolt, An Extended Study of Content and Crowdsourcing-related Performance Factors in Named Entity Annotation, *Semantic Web Journal* (2018). http://www.semantic-web-journal.net/system/files/swj1535.pdf.

[43] A. Dumitrache, O. Inel, B. Timmermans, C. Ortiz, R.-J. Sips, L. Aroyo and C. Welty, Empirical Methodology for Crowdsourcing Ground Truth, *Semantic Web Journal* (2018). http://www.semantic-web-journal.net/system/files/swj1739.pdf.

[44] J. Yang, J. Redi, G. Demartini and A. Bozzon, Modeling Task Complexity in Crowdsourcing, in: *Fourth AAAI Conference on Human Computation and Crowdsourcing*, 2016.

[45] U. Gadiraju, J. Yang and A. Bozzon, Clarity is a Worthwhile Quality: On the Role of Task Clarity in Microtask Crowdsourcing, in: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, HT'17, ACM, New York, NY, USA, 2017, pp. 5–14. ISBN 978-1-4503-4708-2. doi:10.1145/3078714.3078715.

[46] Z. Sun, J. Yang, J. Zhang and A. Bozzon, Exploiting both Vertical and Horizontal Dimensions of Feature Hierarchy for Effective Recommendation, in: *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[47] S. Mesbah, A. Bozzon, C. Lofi and G.-J. Houben, Describing Data Processing Pipelines in Scientific Publications for Big Data Injection, in: *Proceedings of the 1st Workshop on Scholarly Web Mining*, SWM'17, ACM, New York, NY, USA, 2017, pp. 1–8. ISBN 978-1-4503-5240-6. doi:10.1145/3057148.3057149.

[48] S. Mesbah, K. Fragkeskos, C. Lofi, A. Bozzon and G.-J. Houben, Semantic Annotation of Data Processing Pipelines in Scientific Publications, in: *The Semantic Web*, E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler and O. Hartig, eds, Springer International Publishing, Cham, 2017, pp. 321–336. ISBN 978-3-319-58068-5. doi:10.1007/978-3-319-58068-5_20.

[49] M. Sabou, A. Scharl and M. Föls, Crowdsourced Knowledge Acquisition: Towards Hybrid-Genre Workflows, *Int. J. Semant. Web Inf. Syst.* **9**(3) (2013), 14–41, ISSN 1552-6283. doi:10.4018/ijswis.2013070102.

[50] K. Bontcheva, I. Roberts, L. Derczynski and D. Rout, The GATE Crowdsourcing Plugin: Crowdsourcing Annotated Corpora Made Easy, in: *Proceedings of Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Association for Computational Linguistics, 2014, pp. 97–100.