Crowd Work CV: Recognition for Micro Work

Cristina Sarasua(✉) and Matthias Thimm

Institute for Web Science and Technologies, University of Koblenz-Landau, Koblenz-landau, Germany
{csarasua,thimm}@uni-koblenz.de

Abstract. With an increasing micro-labor supply and a larger available workforce, new microtask platforms have emerged providing an extensive list of marketplaces where microtasks are offered by requesters and completed by crowd workers. The current microtask crowdsourcing infrastructure does not offer the possibility to be recognised for already accomplished and offered work in different microtask platforms. This lack of information leads to uninformed decisions in selection processes, which have been acknowledged as a promising way to improve the quality of crowd work. To overcome this limitation, we propose Crowd Work CV, an RDF-based data model that, similarly to traditional Curriculum Vitae, captures crowd workers' interests, qualifications and work history, as well as requesters' information. Crowd Work CV enables the representation of crowdsourcing agents' identities and promotes their work experience across the different microtask marketplaces.

Keywords: Microtask crowdsourcing · CV · Ontology · Crowd worker · Requester · Marketplace

1 Introduction

One of the challenges in human computation systems is to involve the humans who, as intelligent and independent beings with a particular knowledge, are crucial to solve problems that machines can hardly solve alone. Crowdsourcing alleviates this challenge, as it provides a mechanism to distribute a task among a potentially large group of people who subscribe to an open call on the Web [10]. A promising strategy to improve the quality of crowd work, which is particularly relevant for knowledge-intensive crowdsourced tasks, is to find the most suitable worker(s) for a microtask (or vice versa), as Kittur and colleagues highlighted [8]. With the current increasing order of magnitude (in available micro work and workforce), and the evidence of the crowd being diverse in terms of background [11], personality [6], and motivation [5], analysing different aspects of the agents involved in crowdsourcing in order to improve microtask assignment accordingly becomes meaningful. However, the realisation of such process is hindered by the current microtask crowdsourcing infrastructure, which is highly focused on independent marketplaces. Even if many of them have adopted some common patterns (e. g. consider majority voting as aggregation method), each

L.M. Aiello and D. McFarland (Eds.): SocInfo 2014 Workshops, LNCS 8852, pp. 429–437, 2015.
DOI: 10.1007/978-3-319-15168-7_52

of them acts as a data silo. When crowd workers are registered and work in several marketplaces, the work they perform is registered in the marketplace they worked at and only visible there. If a requester is interested in knowing further information on the achievements and proven skills of a worker (e. g. through obtained qualifications) in other marketplaces, this information is not accessible programmatically, even though the data exists and it is visible to the worker. The same applies to requester information. This lack of data interoperability between marketplaces has a negative impact in the process of finding the best combination of workers and microtasks and may result in uninformed decisions. In Section 2 we describe a motivational scenario in more detail.

To overcome this limitation and as a solution to what we proposed in our previous work [12], we introduce Crowd Work CV, an RDF-based data model to represent someone's crowd work life, equivalently to what traditional Curriculum Vitae reflect. Crowd Work CV enables the aggregation of valuable information about crowd workers and requesters, which may be exchangeable if the data owner—the agent represented in the Crowd Work CV—decides to do it. The approach is conceived to boost transparency among crowdsourcing agents, which has a positive effect in crowdsourcing environments, too [4]. The contributions of this work are:

- The definition of a conceptual model to represent Crowd Work Curriculum Vitae information (see Section 3)
- The implementation of the data model into an OWL vocabulary (see Section 3)

The description of the Crowd Work CV data management system is out of the scope of this paper.

2 Motivational Scenario

Let us imagine Alice, who has registered in several marketplaces[1]. She is being assessed as a candidate crowd worker for a group of microtasks published at ClixSense about sentiment analysis of Spanish Web sites. The requester who published the microtasks trusts experienced crowd workers more than unexperienced crowd workers. 1) Alice registered at *ClixSense* but did not work there yet. 2) Alice worked on text translation at *Neobux*, where she obtained a Spanish qualification that a requester defined. 3) At *GetPaid* Alice successfully completed several microtasks which are equivalent to those for which she is going to be assessed, because CrowdFlower distributed the group of microtasks over several marketplaces (ClixSense and GetPaid). 4) At *MTurk* Alice worked with very good performance on other microtasks, which required her to analyse the sentiment of Tweets—similar purpose, with different type of data.

At ClixSense, Alice will be poorly evaluated because her ClixSense work history is empty. Other candidate crowd workers who have worked on Web site

[1] ClixSense http://www.clixsense.com, GetPaid http://www.getpaid.com, Neobux http://www.neobux.com, MTurk http://www.mturk.com.

sentiment analysis microtasks with a much lower performance than what Alice did at GetPaid will be better considered. Alice has a language qualification and she has proven to be capable of solving the type of job being analysed, and even other related microtasks dealing with a similar problem. Still, due to a lack of shared information, the requester will not consider her work experience. This has drawbacks for both parties: the requester is not taking advantage of a potentially good worker for the task at hand, and the crowd worker is missing an opportunity to work on something she might be interested in because of its similarity to previously completed crowd work.

3 Modelling the Crowd Work CV

With Crowd Work CV, we aim at adopting the procedure of reporting work experience from the traditional workplace, where there are plenty of guidelines about the information that should be included in CVs. We identify 5 requirements for a CV in microtask crowdsourcing: First, **domain independence**, i. e. a clear separation between the domain knowledge (e. g. media, fashion, biology) and the management of crowd work history needs to be ensured in order to enable the reusability of crowd work activity reports. Second, **marketplace independence**; the model needs to guarantee a certain level of generality, representing well-established processes instead of particular isolated characteristics provided by one particular marketplace. Third, **semantic and syntactic interoperability**; an agreement on vocabulary should be ensured. The semantics should be explicitly defined and shared separately from the data using a common (in our case) Web-based syntax. Fourth, **extensibility**, because the appearance of new features in marketplaces, or the definition of new workflows in crowd work should not interfere in the already specified model and existing crowd work CV descriptions. Fifth, **compatibility with traditional CV information** defined in standard systems like Europass and LinkedIn[2].

3.1 The Crowd Work CV Ontology

The Crowd Work CV ontology describes crowdsourcing agents (i. e. crowd workers and requesters), their **interests**, obtained **qualifications** and **work history**. The ontology is available online, and written in OWL[3]. We followed the ontology engineering methodology proposed by Noy and McGuiness[9] and considered reusing related ontologies. We decided to reuse some classes and properties of FOAF[4] for the description of agents and SIOC[5] for the description of user

[2] Europass http://europass.cedefop.europa.eu/en/documents/curriculum-vitae and LinkedIn https://www.linkedin.com/.

[3] Implementation of the Crowd Work CV data model: https://github.com/criscod/CrowdWorkCV/tree/master/ontology.

[4] FOAF vocabulary http://xmlns.com/foaf/spec/.

[5] SIOC Core Ontology Specification http://www.w3.org/Submission/sioc-spec/.

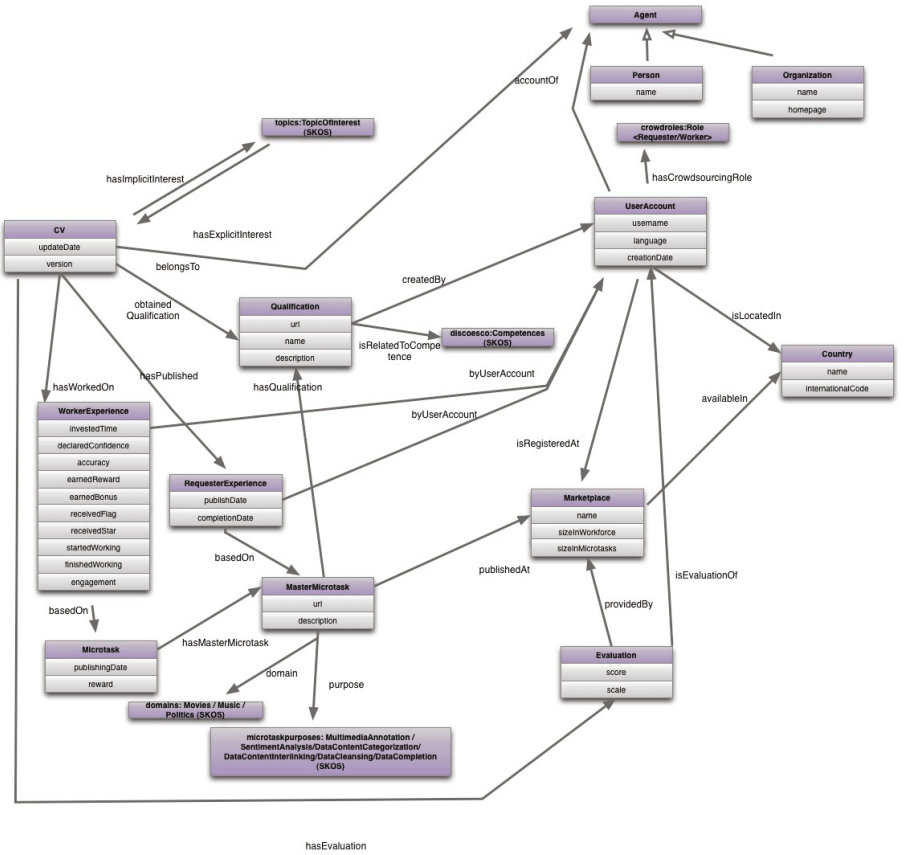

Fig. 1. Overview of the Crowd Work CV ontology. With Crowd CV it is possible to describe agents, their user accounts, CVs, qualifications, work experiences, microtasks and their master microtasks, marketplaces.

accounts, because their definition fits directly our needs and information annotated with such vocabularies on the Web becomes reusable. While the Crowd Work CV elements share some commonalities with the ResumeRDF[6] ontology, for modularity reasons, we decided to define our own ontology elements (which are more oriented to crowdsourcing) and align the CV concept to this ontology. We list the most relevant elements in the Crowd Work CV ontology and describe their purpose. Figure 1 shows the graphical representation of the ontology.

CV is the core class of the ontology. It aggregates all the information that is used to report the crowd work life of an `Agent`, which (from FOAF) can be either a `Person` or an `Organisation` (subclasses). A `CV` may refer

[6] ResumeRDF http://www.w3.org/wiki/ResumeRDFOntology.

to the interests of its owner, which might have been explicitly stated by the owner (`hasExplicitInterest`), or might have been inferred by the interaction in the marketplace (`hasImplicitInterest`). When we think of crowd workers, a `CV` may be related to `obtained qualifications`, which are related to competencies. We propose the use of the SKOS vocabulary for competences included in the Europass[7], but any taxonomy about crowdsourcing-oriented skills can be connected in the same way. For each piece of work accomplished, the `CV` connects the relation `hasWorkerExperience` to a new `WorkerExperience`, which consists of information about the way the crowd worker solved the microtasks (e. g. the time the worker invested, whether the requester gave flags or stars in such work, and the engagement of the worker in the complete group of microtasks). When we think of requesters, a CV is related to the `RequesterExperience`, which refers to the work they offer. The CV may have an `Evaluation` associated, which reflects usually a global evaluation connected to a particular `UserAccount` (e. g. the global reputation of a worker in a marketplace). We align our `CV` class to the CV class in the ResumeRDF vocabulary (with `owl:equivalentClass`).

UserAccount represents the account that an `Agent` may have in a marketplace. It is a SIOC class, to which we associate a role defined in SKOS (`Requester` or `Worker`). Besides the geographical information related to a `UserAccount`, what is relevant for the `CV` is the relation between the `UserAccount` and the `Marketplace` where the account belongs to. A crowd worker may have several accounts (one per marketplace), which are described with a `username`, the `language`(s) spoken by the owner and its `creationDate`.

Qualification refers to the achievement that determines whether an agent (usually a crowd worker) has the required knowledge on a particular topic. Qualifications—which are obtained through qualification tests—can be specified as requirements of microtasks (`hasQualification`), to restrict the set of crowd workers who may accomplish the microtasks. Requesters can write their own tests or reuse the questions provided by marketplaces. In the Crowd Work CV, qualifications may be defined with a textual `description`, a `URL` with a deployed example and a `name`. This class may be extended in the future if categories of qualifications are defined (e. g. language qualifications could be a subclass).

MasterMicrotask is a set of `Microtasks` grouped by the same structure, description and configurations. Usually microtasks are generated applying a template (for the UI and other crowdsourcing settings). Templates are combined with data and the `Marketplaces` convert these into specific microtasks. We have collected a set of common microtask purposes (e. g. from the task templates published by CrowdFlower) and defined a SKOS vocabulary with these. The taxonomy can be further extended. In the same way, we have included in a SKOS vocabulary some examples of possible domains.

[7] ESCO https://ec.europa.eu/esco/download/-/Download/skos.

New domains and purposes that marketplaces or requesters may define can also be included.

Microtask represents the particular instances of `MasterMicrotasks`. The specific unit of work that crowd workers need to solve. `WorkerExperience` is related to `Microtask`, since the information associated to the WorkerExperience (e. g. accuracy, invested time) is based on (`basedOn`) the results obtained in the microtasks. A `Microtask` is related to the `MasterMicrotask` which from it originated (after combining a template with data).

Marketplace represents the crowdsourcing platform containing a Website where microtasks are offered and accomplished. Such platforms provide support for both requesters (for creating the microtasks, defining basic restrictions on who to accept in their microtasks, monitoring the evolution of the work, and obtaining the crowdsourced results) and crowd workers (for browsing available microtasks, acquiring qualifications, submit their work, monitoring their activity in the marketplace and sending their feedback). `Marketplaces` may be described by a `name` and their `sizeInMicrotasks` and `sizeInWorkforce` to have some statistical information about them.

Evaluation reflects the assessment of an `Agent` in a `Marketplace`. It is generally described by a `score` within a `scale`, but it could easily be extended, for example with new intermediate properties that following the criteria suggested by Turkopticon express that a `Requester` is evaluated by its communicativity, generosity, fairness and promptness[8]

Figure 2 shows an excerpt of the serialisation of the motivational scenario. The complete data can be found at the GitHub repository[9].

```
ex:rex a cwcv:RequesterExperience;
    cwcv:byUserAccount ex:acc1;

ex:cv1 cwcv:hasRequesterExperience
ex:rex;

ex:mm1 cwcv:publishedAt ex:ClickSense;
    cwcv:publishedAt ex:GetPaid;
    cwcv:purpose
microtaskpurposes:SentimentAnalysis

ex:cv2 cwcv:obtainedQualification
          ex:q1;
ex:q1 a cwcv:Qualification;
    cwcv:name "Spanish A1";
    cwcv:isRelatedToCompetence
       disco:Capability1;
```

```
ex:mm2 a cwcv:MasterMicrotask;
    cwcv:hasQualification ex:q1;
    cwcv:publishedAt ex:Neobux;

ex:cv2 cwcv:hasWorkerExperience ex:mex1;
ex:mex1 a cwcv:WorkerExperience;
    ex:mex cwcv:basedOn ex:m1;
    cwcv:byUserAccount ex:acc2;
    cwcv:accuracy "0.9";
ex:m1 a cwcv:Microtask;
    cwcv:hasMasterMicrotask ex:mm1;
```

Fig. 2. Crowd Work CV data to describe the work accomplished in marketplaces. For each work done or published an experience is created.

[8] Turkopticon's evaluation criteria http://turkopticon.ucsd.edu/help.

[9] Example of generated Crowd Work CV: https://github.com/criscod/CrowdWorkCV/tree/master/ontology.

3.2 Ontology Verification

In order to ensure that we are following best practices in ontology engineering, we validated our ontology with the OOPS! pitfall scanner[10], which considers a list of 40 common pitfalls in ontology specifications. Except for the imported concepts and properties from other ontologies, we ensured that we do not have important nor critical piftalls.

We also verified the fulfillment of the aforementioned Crowd Work CV requirements: the main elements of the Crowd Work CV ontology refer to **domain-independent** objects in crowdsourcing systems (e. g. microtasks, user accounts and marketplaces). The SKOS vocabularies connected to the core of the Crowd Work CV ontology, which express the purpose of microtasks or the domain, are responsible for bringing the specific knowledge domain into the CV data. Along the same lines, the ontology elements are general enough to be used in **different marketplaces**. For instance, the overall evaluation of a worker in a marketplace or the qualifications do not refer to particular evaluation schemes that MTurk or Clickworker have—which might be different from other marketplaces. The **semantic and syntactic interoperability** of the Crowd Work CV data is achieved with the use of the OWL ontology language. Furthermore, the Crowd Work CV ontology can be **extended** by defining subclasses (e. g. subclasses of qualifications), subproperties, or adding new relations between existing and new ontology concepts. The SKOS vocabularies can also be easily extended in order to have for example, a broader catalogue of microtask purposes. The Crowd Work CV ontology is compliant with existing **standard traditional CV information**, describing the particular instances of work experience, the educational achievements (in our case qualifications) and related skills and competences. More details on the comparison can be found in the GitHub repository[11].

4 Related Work

Several authors have proposed new methods for matching crowd workers and tasks in crowdsourcing environments. Khazankin and colleagues [7] defined a framework for selecting suitable crowd workers to solve a task based on skill requirements attached to tasks, the availability workers report they have, and the skills workers have. Goel and colleagues [2] introduced a method for assigning tasks to workers, which analyses both skills and costs. Difallah and colleagues [1] implemented in a Facebook App a recommendation strategy that pushes suitable tasks to users based on information extracted from their Facebook profiles and previously accomplished HITs, following various assignment strategies (i. e. category-based, text-based, and graph-based). These approaches do not offer a shareable and reusable description of worker expertise that could be used across-platforms. Ul Hassan and colleagues [3] proposed the SLUA ontology for

[10] OOPS! http://oeg-lia3.dia.fi.upm.es/oops/index-content.jsp.

[11] Comparison https://github.com/criscod/CrowdWorkCV/blob/master/ontology/EuropassLinkedIncomparison.txt.

matching users and actions in crowdsourcing scenarios. While the authors raised the problem of lacking interoperability between platforms aligned to our initial proposition [12], their approach has a different focus: they describe tasks, users, rewards and capabilities primarily for routing. In contrast, our goal is to gather more information and be able to share it as a means to recognition for work. We in addition consider microtasks, marketplaces, qualifications and requesters' information. Moreover, our data leads to a workflow for building CV summaries out of large sets of RDF triples. ResumeRDF[12] is an RDFS vocabulary to express information of Curriculum Vitae, including personal details, attended courses, skills and work experience. Celino[13] proposed the Human Computation ontology, which enables the annotation of crowdsourced data and is mapped to the Provenance Ontology. These data models share some common concepts with ours but do not cover all the crowdsourcing-specific domain required in a Crowd Work CV.

5 Conclusions and Future Work

Because microtask crowdsourcing builds a social system, with humans who invest time and money with a purpose, we need to define methods that satisfy the needs and expectations of all involved agents. We have presented Crowd Work CV, an approach for modelling and sharing knowledge about crowd work experience across different marketplaces, which could facilitate a fruitful requester-crowd worker interaction in microtask marketplaces and weave relations of trust. Our approach would considerably enrich the way reputation and credentials are managed in the current crowd workplace. The Crowd Work CV would also encourage job specialisation policies in microtask crowdsourcing.

Future work will focus on the development of the infrastructure of the Crowd Work CV data management system. An interesting area we would like to investigate is the automatic generation of Crowd Work CV summaries out of large sets of Crowd Work CV RDF triples.

The research leading to these results has received funding from the Seventh Framework Programme of the European Union, technological development and demonstration under grant agreement no. 611242 Sense4Us

References

1. Difallah, D.E., Demartini, G., Cudré-Mauroux, P.: Pick-a-crowd: tell me what you like, and I'll tell you what to do. In: Proceedings of the 22nd International Conference on World Wide Web (WWW 2013) (2013)
2. Gagan Goel, A.K., Singla, A.: Matching workers expertise with tasks: incentives in heterogeneous crowdsourcing markets. In: NIPS 2013 Workshop on Crowdsourcing: Theory, Algorithms and Applications (2013)

[12] ResumeRDF http://rdfs.org/resume-rdf/.
[13] Human Computation Ontology http://swa.cefriel.it/ontologies/hc.html.

3. ul Hassan, U., O'Riain, S., Curry, E.: Slua: Towards semantic linking of users with actions in crowdsourcing. In: CrowdSem (2013)
4. Huang, S., Fu, W.: Don't hide in the crowd! increasing social transparency between peer workers improves crowdsourcing outcomes. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (2013)
5. Kaufmann, N., Schulze, T.: Worker motivation in crowdsourcing and human computation. In: Proceedings of the AAAI Workshop on Human Computation (HCOMP) (2011)
6. Kazai, G., Kamps, J., Milic-Frayling, N.: Worker types and personality traits in crowdsourcing relevance labels. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management (2011)
7. Khazankin, R., Psaier, H., Schall, D., Dustdar, S.: QoS-based task scheduling in crowdsourcing environments. In: Proceedings of the 9th International Conference on Service-Oriented Computing (2011)
8. Kittur, A., Nickerson, J.V., Bernstein, M.S., Gerber, E.M., Aaron, S., Zimmerman, J., Lease, M., Horton, J.J.: The future of crowd work. In: 16th ACM Conference on Computer Supported Cooperative Work (CSCW 2013) (2013)
9. Noy, N.F., McGuinness, D.L., et al.: Ontology development 101: a guide to creating your first ontology. Tech. Rep. 2 (2001)
10. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2011)
11. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers?: shifting demographics in mechanical turk. In: CHI 2010 Extended Abstracts on Human Factors in Computing Systems (2010)
12. Sarasua, C., Thimm, M.: Microtask available, send us your cv! In: International Workshop on Crowd Work and Human Computation (CrowdWork 2013) (2013)