# SLUA: Towards Semantic Linking of Users with Actions in Crowdsourcing

Umair ul Hassan, Sean O'Riain, Edward Curry

Digital Enterprise Research Institute
National University of Ireland,
Galway, Ireland.

{umair.ul.hassan, sean.oriain, ed.curry}@deri.org

**Abstract.** Recent advances in web technologies allow people to help solve complex problems by performing online tasks in return for money, learning, or fun. At present, human contribution is limited to the tasks defined on individual crowdsourcing platforms. Furthermore, there is a lack of tools and technologies that support matching of tasks with appropriate users, across multiple systems. A more explicit capture of the semantics of crowdsourcing tasks could enable the design and development of matchmaking services between users and tasks. The paper presents the SLUA ontology that aims to model users and tasks in crowdsourcing systems in terms of the relevant actions, capabilities, and rewards. This model describes different types of human tasks that help in solving complex problems using crowds. The paper provides examples of describing users and tasks in some real world systems, with SLUA ontology.

**Keywords:** crowdsourcing, human computation, users, tasks, ontology

## 1 Introduction

Collective intelligence systems [1] have demonstrated the use of networked humans and computers for solving complex problem, by applying techniques such as crowdsourcing [2], social computing [3] and human computation [4]. Online marketplaces like Amazon Mechanical Turk[1] provide access to large pool of human workers willing to perform variety of micro-tasks for money. Whilst other platforms focus on domain specific crowd services e.g. uTest[2] provides software testing services.

Most of the existing crowdsourcing platforms are isolated in terms of their users and tasks. People contribute towards either a few popular platforms or the systems relevant to their specific domain of knowledge. Hence human resources may be underutilized due to a lack of tools that help people in finding tasks across multiple crowdsourcing platforms. Similarly, task requesters are unable to query across multi-

[1] http://www.mturk.com
[2] http://www.utest.com

ple platforms for their tasks to find appropriate workers with required skills or knowledge.

Main objective of the SLUA (Semantically Linked Users and Actions) ontology is to define a lightweight model for describing crowdsourcing tasks and users with regard to human capabilities, actions and rewards. The scope of the ontology is limited to the micro-tasks that can be performed with in minutes. The specific aims of SLUA ontology are

- To enable interoperability and reuse among crowdsourcing platforms across the web. For example, an active user on Quora[3] for the topic on Cloud Computing might be the right person to edit a Wikipedia article on the same topic.
- To support people in finding online tasks according to their capabilities and motivation. For example, if a person is knowledgeable about the city of New York, then she can help fix problems in Wikipedia[4] articles or tag images of buildings in New York in Amazon Mechanical Turk.
- To facilitate algorithmic matching of tasks and users according to human capabilities, actions, and rewards. For example, a human computation platform might need to verify the chemical formula of a drug from a chemist with the relevant education listed on LinkedIn[5].

The main contributions of this paper are the initial description of the SLUA ontology and its mappings to other existing ontologies, and examples of various tasks described using SLUA. The rest of this paper is organized as follows. Section 2 motivates the need of ontology for describing users and tasks in crowdsourcing platforms. Section 3 highlights the requirements of ontology according to relevant concepts found in literature. Section 4 provides the description of classes and properties in SLUA ontology. Section 5 details some example usage of SLUA for semantic description of tasks and users. Section 6 discusses related work and Section 6 concludes the paper.

## 2 Motivation

Crowdsourcing platforms differ from each other in terms of the tasks that humans can perform and the characteristics of human contributors [5]. Wikipedia requires the crowd to create or edit articles by contributing textual content and references. Quora is powered by questions and answers contributed by online users. Both Wikipedia and Quora relay on the fact that people are motivated to contribute to the crowdsourcing efforts because of social good or self-serving motivations. Amazon Mechanical Turk serves as the market place of human services for performing small online tasks in exchange for money. TaskRabbit[6] allows people to outsource their small physical

[3] http://www.quora.com
[4] http://www.wikipedia.org
[5] http://www.linkedin.com
[6] http://www.taskrabbit.com

**Table 1.** Common terminology used for the concepts of tasks and users in the documentation of popular online marketplaces for crowdsourcing.

| Concept | MTurk | Mobileworks | Shorttask | CrowdFlower |
|---|---|---|---|---|
| *Task* | HIT | Task | ShortTask | Microtask |
| *User* | Worker<br>Requester | Worker<br>Developer | Solver<br>Seeker | Contributor<br>Customer |
| *Reward* | Payment | Payment | Reward | Payment |
| *Capability* | Qualification | Filter | | |

tasks to other people against small monetary price. Microtask[7] uses online gamers to solve problems typically not solvable by computers. In short, the heterogeneity of crowdsourcing systems exists at task, user, and platform levels. Some tasks require cognitive skills while other need physical abilities from humans. Some tasks reward in terms of money while others compensate through enjoyment.

The heterogeneity of crowdsourcing platforms limits interoperability of applications that access more than one crowdsourcing platform. Furthermore, development of cross platform services becomes difficult due to variations of data semantics for each platform. For instance, there is a lack of search engines for microtasks and existing general search engines fail to address this problem. Similarly existing crowdsourcing platforms do not support any application interfaces for users search based on human capabilities. Recently there has been effort to describe crowdsourcing platforms with the help of taxonomies [6]. However they do not cover the modeling of human tasks, actions, and capabilities whilst describing concepts associated with the design aspects of crowdsourcing platform. Therefore we observe that there is a need for a common language for describing human tasks, actions, rewards, and capabilities in crowdsourcing platforms, as well as their relationships. An appropriate ontology may serve the purpose therefore facilitating interoperability supporting broad range of computation services. In the next section we summarize the conceptual requirements of such ontology and assess the coverage of requirements by existing ontologies.

## 3 Ontology Requirements

In this section we analyze the requirements of ontology for human tasks in crowdsourcing. The requirements are based on the common terminology found in current crowdsourcing platforms. Table 1 shows a variety of terminology and concepts used among major crowdsourcing marketplaces. This heterogeneity of terminology creates a gap in terms of common understanding of crowdsourcing concepts among users and developers [7]. Additionally, heterogeneity is reflected in the application programming interfaces offered by crowdsourcing platform, resulting in interoperability issues in terms of the semantics of data structures and algorithms [7].

[7] http://www.microtask.com

The limitations due to the heterogeneity of crowdsourcing platforms necessitate development of domain ontology. In this work we focus on defining lightweight domain ontology for crowdsourcing platforms, specifically for microtasks.

### 3.1 Core Concepts

We define the requirements of the ontology in terms of the core concepts used by major crowdsourcing marketplaces. Existing literature in human computation and crowdsourcing has mainly described the concepts related to platforms design in the form of taxonomies [6, 8]. By comparison, our objective is to define the ontology in terms of what actions people can do for crowdsourcing systems and what human characteristics they need to perform those actions. Therefore, the following concepts constitute the main requirements of the ontology:

- **Task:** This concept is commonly used in the literature and crowdsourcing platforms to describe a unit of work to be performed by people in the crowd [2, 4]. Sometimes complex tasks are divided into smaller simple tasks to increase crowd participation [9].
- **Action:** The cognitive or psychomotor action or activity that leads towards the completion of a task [10]. A task can include one or more actions, for instance an audio transcription task includes activities of listening and writing.
- **User:** Commonly described as "worker" in crowdsourcing marketplaces due to the monetary payments earned by users [9]. However other crowdsourcing systems like Wikis and question answering systems used the concept of user to describe contributors.
- **Reward:** The concept of reward is popular in crowdsourcing marketplaces. However the existing literature considers this as a core concept related to the motivation of people in the crowd [9, 11, 12]. Although monetary rewards are common in marketplaces other motivating factors such as altruism, fun, learning, and reputation are also considered rewards.
- **Capability:** The human ability, knowledge, or skill that allows a user to perform the necessary actions for task completion [4, 13]. Availability and location of a person may include the requirements for some tasks.

### 3.2 Existing Ontologies

We have mapped the concepts described in ontology requirements with existing ontologies such as FOAF[8], SIOC[9], HRMO[10], PIMO[11], and TMO[12]. Table 2 shows how classes in the existing ontologies map with the concepts required for the ontology.

[8] http://www.foaf-project.org/
[9] http://sioc-project.org/ontology
[10] http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/ontologies/99-hrmontology
[11] http://www.semanticdesktop.org/ontologies/pimo/
[12] http://www.semanticdesktop.org/ontologies/2008/05/20/tmo/

**Table 2.** Mapping of concepts described in the ontology requirements with existing taxonomies and ontologies.

| ***Concept*** | **PIMO** | **TMO** | **HRM-O** | **FOAF** | **SIOC** |
|---|---|---|---|---|---|
| *Task* | Task | Task | | | |
| *Action* | | | | | |
| *User* | Person | | Job Seeker | Person | UserAccount |
| *Reward* | | | Compensation | | |
| *Reputation* | | | | | |
| *Money* | | | Salary | | |
| *Fun* | | | | | |
| *Altruism* | | | | | |
| *Learning* | | | | | |
| *Capability* | | | | | |
| *Location* | | | Location | | |
| *Skill* | | Skill | | | |
| *Knowledge* | | | | | |
| *Ability* | | | Ability | | |
| *Availability* | | | Interval | | |

The *Personal Information Management Ontology* (PIMO) and *Task Management Ontology* (TMO) model tasks and skills of users of semantic desktops. The mappings with other ontologies or vocabularies can be defined using *owl:equivalentClass* and *owl:equivalentProperty* in web ontology language (OWL). Standard reasoning engines can be used to carry out the mappings for the instances of mapped ontologies. The coverage gap highlighted in Table 2 underlines the need for a separate ontology for human tasks, actions, rewards, and capabilities in crowdsourcing systems.

## 4 SLUA Ontology

The *Semantically Linked Users and Actions* (SLUA) ontology contains 5 main classes and 10 sub-classes that describe users and tasks in crowdsourcing systems. In the previous section we have identified the main concepts found in literature for describing the tasks and users in crowdsourcing systems. These concepts form the set of core classes in the SLUA ontology, as shown in Figure 1. Although similar concepts are captured by other ontologies, it is the relationships, class hierarchy, and properties of these concepts that are unique to SLUA. In the rest of this section the classes and their relationships are described in more detail.

### 4.1 Main Classes

The list of classes in the SLUA ontology are

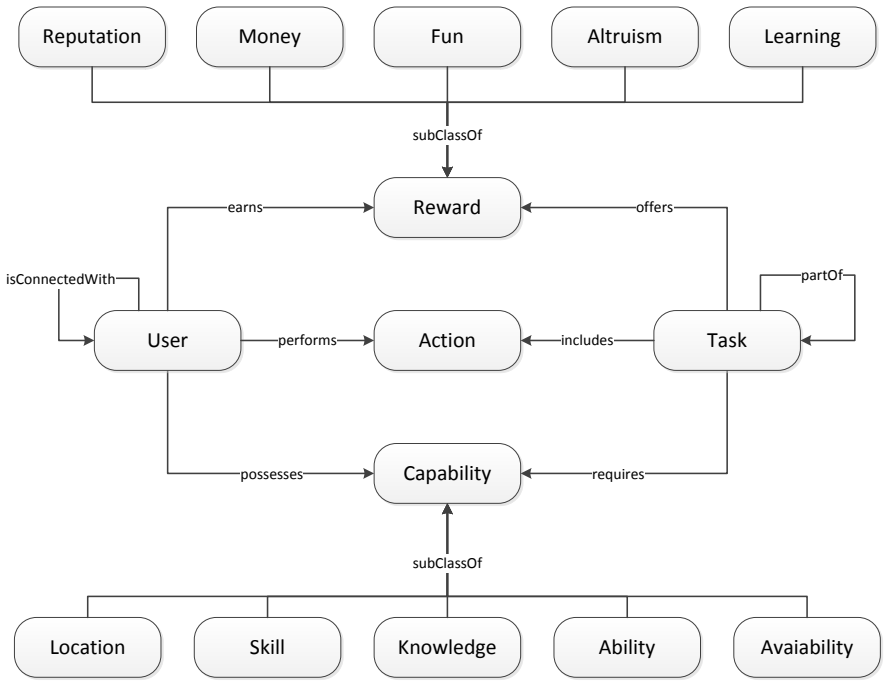


**Fig. 1.** Overview of the 5 main classes and 10 sub-classes defined in the SLUA ontology

- **Action** class represents a specific act that is performed by the members of the crowd. An action can be cognitive or physical. For example, the comparison of two images involves a cognitive action from user.
- **Task** defines the unit of work resulting in a desired outcome that is assigned to the members of the crowd. A task may require one or more actions to produce the outcome. Therefore a task at the lowest level is composed of actions. The *Task* class has composition relationship with itself because complex tasks can be broken down into small simple tasks.
- **User** is the class that describes the human contributor in crowdsourcing. The user serves as an intelligent agent that is able to perform actions for successful completion of assigned tasks.
- **Reward** is associated with a task as the incentive for the human contribution. As noted earlier currently there are five types of reward classes:
  - **Fun** class represents rewards involving entertainment value such as games.
  - **Money** class represents monetary rewards.
  - **Fame** class represents rewards that benefit people in terms of recognition such as top contributors in Wikipedia.
  - **Altruism** class represents rewards involving social good.

  - **Learning** class represents rewards resulting in personal improvement in skill or knowledge.
- **Capability** defines the human characteristics that are necessary for performing a task. For instance one system might specify a user's location capability while another system utilizes this description to assign tasks relevant to the same location. There are five sub-classes defining different capabilities:
  - **Ability** class represents the stable capacity of users to engage in a specific behavior.
  - **Knowledge** class represents a body of information accumulated by users through education or experience.
  - **Skill** class represents the proficiency of a user in performing a task. Skill is acquired through training and practice.
  - **Location** class represents the specific place where a user is or will be physically present. This type of capability enables crowd contributions that are related to a physical place.
  - **Availability** class represents the time interval or time instant during which a user can perform a task.

### 4.2 Important Properties

This sub-section describes the properties of SLUA concepts that are important for extracting meaning from classes

- **domain:** A domain definition applies to most of the classes defined above. This property can be helpful for domain specific algorithms. A common categorization system could be used to specify domains in general crowdsourcing systems. However for specific areas purpose built taxonomies defined can be more effective.
- **offers:** This property defines the relationship of *Reward* with *Task*. For example some tasks might be rewarded with money. By comparison a user who is interested in a particular reward can be described with the **earns** property.
- **requires:** A *Task* can define requirements of one or more human capabilities using this property. By contrast a *User* can be described by having similar capabilities using the **possesses** property.
- **includes:** A *Task* can define one or more actions that a *User* **performs** for generating the desired outcome of a task.
- **isPartOf:** A complex *Task* can be decomposed into small manageable tasks. Therefore this property helps is describing the composition relationship between tasks.
- **hasDeadline:** This property can be used to specify time limitations of a *Task*, which is specifically important for real-time systems employing crowds.
- **isConnectedWith:** In the context of social networks, users are connected with other users through various relations. This property captures the network structure of users to enable social network based analysis of actions and users. For example the network structure can be exploited to recommend actions to neighbor nodes in a network.


**Fig. 2.** Example of an article with a cleanup task that suggests users to add verifiable references to meet the quality standards of Wikipedia.

There are also domains specific properties that can be used to describe SLUA instance, as exemplified later in this paper.

## 5 Using SLUA

The core objective of SLUA is to provide a simple language to describe human tasks in crowdsourcing platforms to facilitate connectivity of tasks with users who can perform them. The SLUA ontology enables exchange of information on tasks, actions, users, rewards, and capabilities across crowdsourcing platforms. In the rest of this section we illustrate the use of SLUA in describing the semantics of tasks and users in different crowdsourcing systems.

### 5.1 Describing Tasks

**Collaborative Information Management.** Wikipedia is a large collection of textual articles edited collaboratively by users on the Web. Articles in Wikipedia are routinely tagged for cleanup tasks due to issues with content or style. These tasks include adding new references, revising articles, merging sections, etc. For example, Figure 2 shows the alert message for an article about the "A3 road" in the City of London. The message suggests an action is required to remedy the quality issue with the article.

The alert message of a Wikipedia article and associated task can be described in *Resource Description Framework* (RDF) format using SLUA ontology. This allows machine readable access to human actions for improving quality of content in Wikipedia. The following code gives the example of the Wikipedia cleanup task converted to an instance of the *Task* class in SLUA.

```
<http://www.wikipedia.org/wiki/A3_road/tasks/1> a
slua:Task ;
    rdfs:label "Please consider adding full citations to
the Wikipedia article";
    slua:requires [
        a slua:Location;
        slua:locatedIn
<http://live.dbpedia.org/resource/London> ];
    slua:requires [
        a slua:Knowledge;
        slua:locatedIn
<http://live.dbpedia.org/resource/Roads> ];
    slua:offers [
        a slua:Reward;
        a slua:Reputation;
        slua:amount "1 star" ];
    slua:includes [
        a slua:Action
        rdfs:label "Wiki page edit"] .
```

**Online Crowdsourcing Marketplace.** Amazon Mechanical Turk is an online marketplace where requesters submit human intelligence task (HIT) to be performed by workers (i.e. users) in return for small amounts of money. Figure 3 shows example of task requiring users to describe a video with short sentences. Using SLUA the task can be described as an instance of *Task* class. The task requires two human capabilities; the capability of *Location* having *locatedIn* property with value "United States" and the capability of *Availability* with *availableFor* property with value "60" minutes. By performing the task workers can earn *Reward* of type *Money* with *amount* property of value "$0.15".

**Cyber-Physical System.** Next generation building management systems [14] involving human-in-the-loop for performing physical actions in the environment [15]. These systems ask occupants to perform environmental actions such as closing windows for reducing energy usage. A human action in building energy management serves as

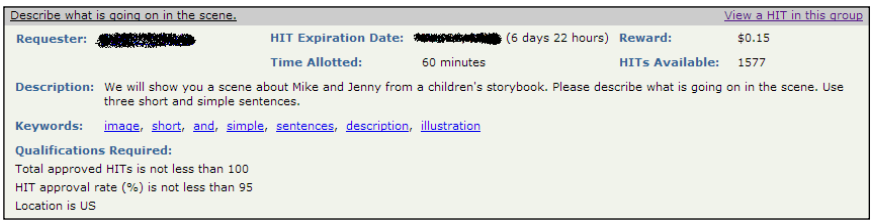


**Fig. 3.** Example of human intelligence task (HIT) on Amazon Mechanical Turk

another use case for use of SLUA ontology. In this case the window closing action can be described as *Task*, which requires location capability; the capability of *Location* having *locateNear* property with value "Room A1" and having *locationTime* property with value "10:00PM".

### 5.2 Describing Users

Similar to the description of tasks, the users of crowdsourcing platforms can be described using SLUA. Users can be described in terms of the actions they perform, the rewards they earn, and the capabilities they possess. The connection between various users can also be described to facilitate social network analysis. The following code gives an example of a Wikipedia user described with SLUA in RDF Turtle format.

```
<http://www.wikipedia.org/wiki/user/u0901> a slua:User .
    faof:name "Umair ul Hassan";
    slua:possess [
        a slua:Location;
        slua:locatedIn
<http://live.dbpedia.org/resource/London> ];
    slua:possess [
        a slua:Knowledge;
        slua:locatedIn
<http://live.dbpedia.org/resource/Roads> ];
    slua:earns [
        a slua:Reputation;
        slua:amount "4 star" ].
```

### 5.3 Leveraging Semantic Descriptions

Improving the routing of tasks to appropriate users is another objective of SLUA. In this regard the semantic descriptions of users and task can be used to perform the routing process. There are three major components of task routing system, as shown in Figure 4.

- **Task Modeling:** Uses SLUA to describe tasks. The capabilities of tasks can be discovered using methods such as cognitive task analysis [16].
- **Worker Profiling:** Uses SLUA to describe profiles of Users. The profiles can be generated using techniques such as expertise retrieval, behavior analysis, performance analysis, etc.
- **Task Routing:** Given task and user descriptions, this process involves finding suitability of user for a task. Depending on the tasks and users a variety of semantic similarity[13] approaches can be used of the purpose of matching.

[13] http://en.wikipedia.org/wiki/Semantic_similarity

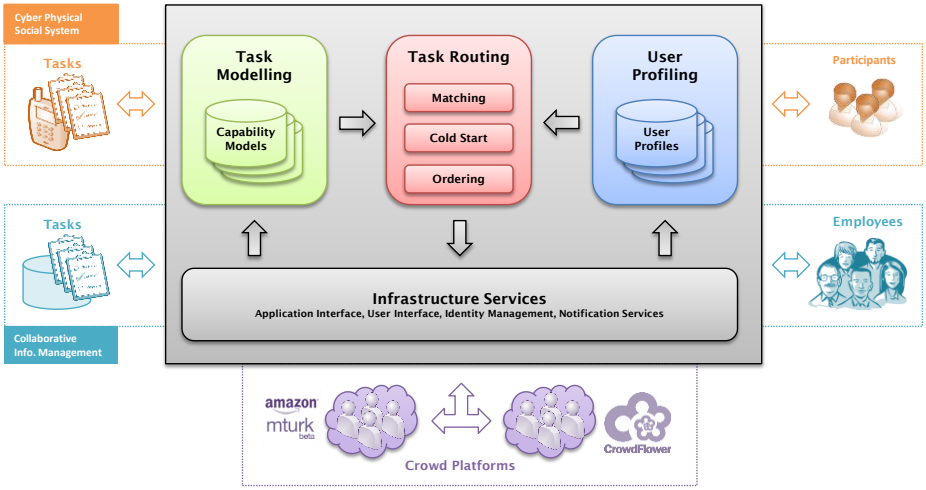


**Fig. 4.** Architecture of task routing system for heterogeneous tasks and users in crowdsourcing

## 6 Related Work

There has been considerable work on studying various dimensions of systems combining efforts of networked humans and computers. Malone et al. [1] described a framework of understanding working of a system based on collective intelligence. Doan et al. [2] discussed application of crowdsourcing to various domains. Quinn and Bederson [6] developed a taxonomy of human computation systems. Kearns [17] described tasks suited for social computing. These studies describe various aspects of human actions from a research perspective. By comparison this paper attempts to described actions and users for interoperability.

Bernstein et al. [13] called for the development of "social operating systems" for managing and allocating tasks to human resources at the global scale. Kittur et al. [9] highlight task assignment as the main research challenge for crowd work. Similarly, task routing has be defined as the fundamental aspect of human computation [4]. Ul Hassan etl al. have studied the relationship between user expertise and task routing in collaborative data quality management [18–20]. Diffallah et al. used social network profiles of users for assigning crowdsourcing tasks [21]. In this regard, SLUA provides a common language for the matching of tasks and users in crowdsourcing systems.

Existing ontologies, such as Personal Information Management Ontology [22] and Task Management Ontology [23], model some aspects of human actions and human capabilities. However these ontologies focus on task management from a desktop applications perspective. By comparison, SLUA specifies terms for crowdsourcing systems including rewards, capabilities, and actions.

## 7 Summary and Future Work

Semantically Linked Users and Actions is an initial step towards defining a light-weight ontology for describing tasks, actions, users, rewards, and capabilities in crowdsourcing platforms. This paper describes the core concepts and properties of SLUA ontology. This paper also gives example uses of SLUA to describe actions in different crowdsourcing scenarios. Future work includes the development of a prototype for exporting SLUA data from crowdsourcing platforms and developing a system that performs matchmaking between users and tasks using SLUA descriptions.

**Acknowledgement.** The work presented in this paper has been partially funded by Science Foundation Ireland Grant No. SFI/08/CE/I1380 (Lion- 2).

## References

1. Malone, T.W., Laubacher, R., Dellarocas, C.N.: Harnessing Crowds: Mapping the Genome of Collective Intelligence, http://www.ssrn.com/abstract=1381502, (2009).
2. Doan, A., Ramakrishnan, R., Halevy, A.Y.: Crowdsourcing systems on the World-Wide Web. Communications of the ACM. 54, 86–96 (2011).
3. Schuler, D.: Social computing. Communications of the ACM. 37, 28–29 (1994).
4. Law, E., Ahn, L. von: Human Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning. pp. 1–121 (2011).
5. Curry, E., Freitas, A., Riain, S.O.: The Role of Community-Driven Data Curation for Enterprises. In: Wood, D. (ed.) Linking Enterprise Data. pp. 25–47. Springer US, Boston, MA (2010).
6. Quinn, A.J., Bederson, B.B.: Human computation: a survey and taxonomy of a growing field. Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11. pp. 1403–1412. ACM Press, New York, New York, USA (2011).
7. Uschold, M., Gruninger, M.: Ontologies and semantics for seamless connectivity. ACM SIGMOD Record. 33, 58 (2004).
8. Geiger, D., Seedorf, S., Nickerson, R.C., Schader, M., Nickerson, R.: Managing the Crowd : Towards a Taxonomy of Crowdsourcing Processes. (2011).
9. Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., Horton, J.: The future of crowd work. Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13. pp. 1301–1318. ACM Press, New York, New York, USA (2013).
10. Sakamoto, Y., Tanaka, Y., Yu, L., Nickerson, J. V: The Crowdsourcing Design Space. 346–355 (2011).
11. Simperl, E., Cuel, R., Stein, M.: Incentive-Centric Semantic Web Application Engineering. Synthesis Lectures on the Semantic Web: Theory and Technology. 3, 1–117 (2013).
12. Tokarchuk, O., Cuel, R., Zamarian, M.: Analyzing Crowd Labor and Designing Incentives for Humans in the Loop. IEEE Internet Computing. 45–51 (2012).
13. Bernstein, A., Klein, M., Malone, T.W.: Programming the global brain. Communications of the ACM. 55, 41 (2012).
14. Curry, E., Hasan, S., O'Riain, S.: Enterprise Energy Management using a Linked Dataspace for Energy Intelligence. Second IFIP Conference on Sustainable Internet and ICT for Sustainability. , Pisa, Italy (2012).

15. Crowley, D.N., Curry, E., Breslin, J.G.: Closing the Loop-From Citizen Sensing to Citizen Actuation. 7th IEEE International Conference on Digital Ecosystem Technologies. (2013).
16. Schraagen, J.M., Chipman, S.F., Shalin, V.L.: Cognitive Task Analysis. Taylor & Francis (2000).
17. Kearns, M.: Experiments in social computation. Communications of the ACM. 55, 56 (2012).
18. Ul Hassan, U., O'Riain, S., Curry, E.: Towards Expertise Modelling for Routing Data Cleaning Tasks within a Community of Knowledge Workers. Proceedings of the 17th International Conference on Information Quality. , Paris, France (2012).
19. Ul Hassan, U., O'Riain, S., Curry, E.: Effects of Expertise Assessment on the Quality of Task Routing in Human Computation. Proceedings of the 2nd International Workshop on Social Media for Crowdsourcing and Human Computation. , Paris, France (2013).
20. Ul Hassan, U., O'Riain, S., Curry, E.: Leveraging Matching Dependencies for Guided User Feedback in Linked Data Applications. Proceedings of the Ninth International Workshop on Information Integration on the Web. pp. 1–6. ACM Press, New York, New York, USA (2012).
21. Difallah, D.E., Demartini, G., Cudré-Mauroux, P.: Pick-a-crowd: tell me what you like, and i'll tell you what to do. Proceedings of the 22nd international conference on World Wide Web. pp. 367–374. International World Wide Web Conferences Steering Committee (2013).
22. Sauermann, L., Elst, L. van, Dengel, A.: PIMO - a Framework for Representing Personal Information Models. Proceedings of I-Semantics' 07. pp. 270–277. JUCS, Graz, Austria (2007).
23. Ong, E., Riss, U. V., Grebner, O., Du, Y.: Semantic Task Management Framework. Proceedings of I-KNOW '08. pp. 387–394. JUCS, Graz, Austria (2008).