Process activity ontology learning from event logs through gamification

SAREH SADEGHIANASL, ARTHUR H.M. TER HOFSTEDE, MOE THANDAR WYNN, SELEN TURKAY, AND TRINA MYERS.
Queensland University of Technology, Brisbane, Australia
E-mail: {s.sadeghianasl, a.terhofstede, m.wynn, selen.turkay, trina.myers}@qut.edu.au
Corresponding author: S. Sadeghianasl (e-mail: s.sadeghianasl@qut.edu.au).

ABSTRACT The quality of event log data is a constraining factor in achieving reliable insights in process mining. Particular quality problems are posed by activity labels which are meant to be representative of organisational activities, but may take different manifestations (e.g. as a result of manual entry synonyms may be introduced). Ideally, such problems are remedied by domain experts, but they are time-poor and data cleaning is a time-consuming and tedious task. Ontologies provide a means to formalise domain knowledge and their use can provide a scalable solution to fixing activity label similarity problems, as they can be extended and reused over time. Existing approaches to activity label quality improvement use manually-generated ontologies or ontologies that are too general (e.g. WordNet). Limited attention has been paid to facilitating the development of purposeful ontologies in the field of process mining. This paper is concerned with the creation of activity ontologies by domain experts. For the first time in the field of process mining, their participation is facilitated and motivated through the application of techniques from crowdsourcing and gamification. Evaluation of our approach to the construction of activity ontologies by 35 participants shows that they found the method engaging and that its application results in high-quality ontologies.

INDEX TERMS Process mining, activity labels, ontology, data quality, gamification, crowdsourcing.

I. INTRODUCTION

Process mining is the analysis of event logs, historical process data, which generates actionable insights to improve processes by business owners [1]. The use of poor-quality event logs leads to unreliable analysis results and insights (i.e., garbage in – garbage out) [2]. Of critical importance are the activities that are performed in a process and their correct identification requires a deep insight into the domain involved. However, data quality can be compromised. For example, the same activity in a process might be referred to by different labels, a case of so-called synonymous labels [3]. Also, different abstraction levels of activity labels (i.e., too detailed vs too general) are quite common [4].

To use domain knowledge in improving the quality of activity labels, the preliminary step is to represent knowledge in a format that a computer can process. An ontology is one of the prominent methods to formally specify concepts and relationships in a particular domain. Ontology learning is the task of extracting ontological elements (e.g. concepts, and relations) from data [5]. The main goal of ontology learning is to share and reuse the knowledge of the domain effectively [6]. While automatic approaches to ontology learning from general data sources have been proposed (e.g. [7]–[11]), there has been limited attention paid to ontology learning from process data. As argued by Mendling et al. [12], learning ontologies from process-related data sources is one of the challenges that semantic process mining is facing.

The quality of the ontology learned from data is shown to be proportional to the human intervention level [13]. Domain experts are well-positioned to create activity ontologies from process data; however, it is a challenge to engage them in the tedious task of ontology learning as they are usually in-demand, expensive, and time-poor. Thus, the research question of this paper is how can we create activity ontologies from process data in an engaging manner?

In this paper, the Process Activity Ontology (PAO) is created through a novel method, which uses crowdsourcing and gamification techniques. Crowdsourcing [14] provides a diversity of knowledge of a large group of people without any limitation to their time and location. In this approach, some strategies are used to further estimate the level of domain expertise of participants. Gamification is the use of game elements in non-gaming contexts to enhance user experience [15]. To the best of our knowledge, ontology learning has not been conducted through the use of a gamified system. The concepts in the PAO are activity labels and the relations
II. RELATED WORK

This section reviews the existing literature on the topics of activity label quality improvement in process mining, ontology learning in general, ontology learning in process mining, and gamification.

The improvement of the quality of activity labels has been studied in the process mining literature [17]–[20]. Some approaches are at the process model level [17], [18], [21], and some others are at the event log level [19], [20], which mainly focus on detecting and repairing activity labels with the same syntax but different semantics (i.e., homonymous labels [3]). Activity quality improvement has been conducted through the use of domain knowledge [22]–[25], which can be sourced from an existing domain ontology [18], or it can be provided by humans (either a few individuals [22], [23], or a crowd [24], [25]). Koschmieder et al. [18] propose a semi-automatic approach to revise activity labels in process models using part-of-speech tagging algorithms and existing or manually created ontologies. The final revisions of the labels are selected by human users of the designed tool. Other methods [24], [25] identify synonymous labels through crowdsourcing and argue that such a task is not as straightforward as most automatic approaches suppose. In our earlier work [26], a gamified crowdsourced method is proposed which detects and repairs synonymous activity labels in an event log. While existing approaches focus on synonymous labels, the approach proposed in this paper identifies not only synonymy but also other types of semantic relations that are hypernymy (i.e., the super-class sub-class relation), holonymy (i.e., the whole-part relation), and antonymy between activity labels through the use of a crowdsourced gamified system.

Ontology learning is the task of extracting ontological elements from input data [5] that has been well-studied in the literature (refer to [5], [13], [27] for some example review papers). Based on a review by Hazman et al. [27], approaches to ontology learning can be classified based on the input data and the method. The input data to an ontology learning system might be unstructured (e.g., textbooks) [8], semi-structured (e.g., HTML, web pages) [28], or structured (e.g., relational databases) [29]. The method used for ontology learning can be automatic [7], [8] or it can involve human intervention. Asim et al. [13] review the automatic ontology learning approaches and classify the techniques they use into three main categories: (1) linguistic methods [9], based on language characteristics such as part-of-speech-tagging; (2) statistical methods [10], based on probability and frequency measures; and (3) logical methods [11], based on inductive logic programming techniques to derive hypotheses.

Ontology learning has been conducted manually by a limited number of domain experts with the aid of tools like Protégé [30] or OntoEdit [31] in some works [32], [33]. Another set of approaches [34]–[36] use human input to either prepare the input for automatic algorithms or to verify and amend their outputs. The idea of using the knowledge of the crowd (i.e., crowdsourcing) for ontology learning has also been studied in the literature [37]–[39]. Ontologies have been used to improve gamification design [40], [41]. However, to the best of our knowledge, no gamified system has hitherto been designed to crowdsource ontology learning.

Within the field of process mining, there has been limited attention to the use and creation of ontologies to improve analysis outcomes. There are some works that manually or semi-automatically (with the aid of tools) integrate existing domain ontologies into their methods to, for example, extract event logs from relational databases [42], transform activities of an event log to a higher level of abstraction [43], or revise activity labels in process models [18]. Some approaches [44], [45] generate a glossary of activity labels found in a collection of process models in a domain. The glossary automatically created by Peters et al. [44] consists of activity labels as complete phrases, their element type (event or function in the Event-driven Process Chain (EPC) process modelling language), and their control flow information (e.g., co-occurrence, exclusiveness, and strict order). The glossary manually generated by Becker et al. [45] includes grammar checked activity labels as well as their breakdown terms. The semantic relations between the terms that are synonymy, antonymy, homonymy (i.e., the relation between activity labels with the same syntax but different semantics [3]), and word formation are also manually extracted from a thesaurus. The former approach [44] is automated but lacks information about the semantic relations between labels, while the latter [45] is manual and may thus require serious efforts, which threatens its applicability.

In another automatic method, Coskuncay et al. [46] create a process ontology from process models. Their ontology includes different elements of a business process, such as activities, resources, and tools. However, the created ontology leaves the question of semantic relations between activities unanswered because such information cannot be acquired from a process model, only from domain experts. There are also a few automatic methods that use natural language processing techniques to build a process ontology from textual descriptions available in organisational archives [47], [48]. The structure of their ontologies is similar to the one created by Coskuncay et al. [46] but adds further information, for example, the whole-part relation between organisational roles. These approaches are only applicable when the rele-
IV. THE PROCESS ACTIVITY ONTOLOGY

The activities performed in a process and their semantic relations are modelled in the PAO, which employs the semantic web technologies. This section provides a background to the semantic web concepts (Section IV-A), and describes the structure (Section IV-B) and the semantic relations (Section IV-C) of the PAO.

A. SEMANTIC WEB BACKGROUND

The Semantic web aims to improve the World Wild Web by representing web data in a machine-readable format [66]. The knowledge representation in the semantic web is performed in a number of steps (a.k.a. the “layer cake” of the semantic web [67]), each constructing a layer based on another [68]. The layers sorted upwards in the layered cake are the Uniform Resource Identifiers (URIs; which name things uniformly across the web), the Extensible Markup Language (XML; which structures web documents in a user-defined format), the Resource Description Framework (RDF; which models objects and their semantic relations as triples – subject-predicate-object), the Resource Description Framework Schema (RDFS; which provides primitives, e.g. class, property, sub-class, domain and range, to structure web objects into hierarchies), and the Web Ontology Language (OWL; which is based on the Description Logic (DL) language and has richer semantic constructs than RDFS, e.g. disjointness and cardinality of classes).

The PAO will be represented in OWL [62], which is developed by the Web Ontology Working Group of the World Wide Web Consortium (W3C). It is a standard ontology language that is supported by well-established software packages such as Protégé [30]. Inference (a.k.a. reasoning) is the task of deriving new data from existing data and is performed by an inference engine (a.k.a. reasoner or classifier). Reasoning over an ontology in a DL language, e.g. OWL-DL (an OWL sub-language [69]), can be carried out by reasoners such as HermiT [70], FACT++ [71], and RacerPro [72].

Reasoning in DL-based languages is based on the Open World Assumption (OWA). In OWA, the reasoner does not classify the statements that are not proven to be true as false statements; they are rather assumed to be missing from the ontology [73]. For example, if the statement ‘activity a and activity b are synonymous’ is added to the PAO, the query ‘is activity b synonymous to activity c?’ will return ‘unknown’. In the opposite Closed World Assumption (CWA; used in relational databases), where the available information is the whole truth [73], the answer would be ‘false’. The OWA is an appropriate assumption here because the activities and relations that are missing from an incomplete PAO do exist in the real world. For example, activity b and activity c may be synonyms in reality but their relation might be simply missing from the PAO at a certain point of time.

An OWL ontology is a collection of basic pieces of knowledge called triples. A triple is an expression consisting of a subject, a predicate, and an object, e.g. ‘Human eats vegetable’, or ‘garden is green’. OWL triples are built on
top of RDF triples. For example, the triple ‘Human rdf:type owl:Class’ declares an OWL class representing the concept ‘Human’.

A set of triples construct an axiom in OWL. Axioms denote concepts as classes, relationships as object properties, attributes as data properties, and objects as individuals of classes. OWL allows to define a hierarchy of classes (with the root class “Thing”), where the sub-classes represent more specific concepts than the super-class(es). For example, the OWL class axiom ‘Class (Human)’ consists of one triple ‘Human rdf:type owl:Class’. Other types of axioms in OWL include disjoint, domain and range, and closure axioms.

**B. THE PROCESS ACTIVITY ONTOLOGY STRUCTURE**

Figure 1(a) depicts a snippet of an event log related to a loan application process. Each row describes an event (identified by an eventId), which is the record of the execution of a well-defined step in a process. The events are categorised into a number of cases (identified by caseId), each denoting one execution of the process (e.g. a specific loan application). An event is described by an activity (e.g. ‘Submit documents’) performed at a particular point of time (represented by a timestamp) as part of a case (e.g. c1). There may be some additional attributes for an event, such as the resource (e.g. ‘User1’) who is originating the activity, and some data attributes (e.g. ‘Amt’ which is the loan amount requested in an application).

The PAO represents activities performed in a particular process (e.g. the loan application process) and their semantic relations. Figure 1(b) shows the general structure of the PAO consisting of two high-level classes: Activity (representing all activities that can be performed in a process) and DataAttribute. The Activity class has four data properties, which are timestamp, caseId, resource, and suitability (that is 0 by default and is used to identify the dominant activity label amongst a number of synonyms), and the DataAttribute class has two data properties, which are name (denoting the name of a data attribute, e.g. ‘Amt’), and value (denoting the value of a data attribute, e.g. ‘10000’). The object property hasData is defined from the class Activity to the class DataAttribute. The lower level classes of the PAO depend on the process it represents. The sub-classes of the Activity class are the distinct activity labels recorded in one or more event log(s) related to a process. For example, Figure 1(c) depicts the lower level classes of the PAO related to the loan application process recorded in the event log illustrated in Figure 1(a). The OWL statement:

Class (Check_loan_amount)
SubClassOf(Check_loan_amount Activity)

defines the class Check_loan_amount representing the activity ‘Check loan amount’ in Figure 1(a). Check_loan_amount is a sub-class of Activity because it is a type of activity that is performed in the loan application process. The individuals of a certain activity class (e.g. ‘Check_loan_amount’) in the PAO are the events (e.g. e4 and e13) where the activity is performed in the event log(s). In this formulation, an event is equated with an instance of the execution of an activity. The relations between activity classes in Figure 1(c) are described in Section IV-C.

**C. SEMANTIC RELATIONS OF ACTIVITIES**

The PAO models four possible relationships between activities of a process, although it can be extended in the future to include more relations. The relationships are synonymy, hypernymy, holonymy, and antonymy and are discussed in the following sections.

1) Synonymy relation

Two activity labels are synonyms if they are referring to the same task in a process, e.g. ‘Begin check’ and ‘Start check’ in Figure 1(a). Synonymous labels are most common when an event log stems from multiple systems that name the same task differently [3].

A synonymy relationship between two activity labels is represented by the equivalence axiom between their corresponding classes in the PAO. For instance, the OWL statement:

EquivalentClasses (Begin_check Start_check)

declares that the two classes Begin_check and Start_check are referring to the same concept (i.e., activity) in the loan application process. As a result, the DL reasoner will automatically infer that any individual (i.e., event) that is a member of the class Begin_check also belongs to the class Start_check and vice versa. The synonymy relation is depicted by the green edges in Figure 1(c).

2) Hypernymy relation

Hypernymy is a semantic relation between a general activity called hypernym and a more specific activity called hyponym. For example, ‘Pay’ is a hypernym and ‘Pay online’ is a hyponym activity. Also, two activities that share the same hypernym activity are called cohyponyms, e.g. ‘Pay online’ and ‘Pay by cash’ that share the same hypernym ‘Pay’. Such relationships are more common between labels of event logs that are derived from systems that allow free-text data entry. As a result, the performers of the activities record them with different levels of details in data [74].

The hypernymy relationship between activities is modelled as the sub-class axiom in the PAO. For example, the OWL statements:

SubClassOf (Pay Activity)
SubClassOf (Pay_online Pay)
SubClassOf (Pay_by_cash Pay)

mean that the class Pay is the super-class (i.e., hypernym) and the classes Pay_online and Pay_by_cash its sub-classes (hyponyms). These statements also imply that the two classes

---

1 Please note that spaces in OWL are replaced by underscores.
Sadeghianasl et al.: Learning ontologies from event logs

Pay_online and Pay_by_cash that share the same super-class are cohyponyms. The DL reasoner will classify all members of a sub-class (e.g. Pay_online and Pay_by_cash) as members of its super-class (e.g. Pay). The hypernymy relation is depicted by the blue arrows (with the arrow towards the hypernym) in Figure 1(c).

3) Holonymy relation
Holonymy denotes the relation between a whole activity called holonym and a partial activity called meronym. For example, ‘Submit documents’ is a holonym and ‘Submit form A’ is a meronym activity, where form A is one of the documents that is submitted as part of a loan application. Two activities that are parts of the same whole activity (i.e., holonym) are called coneronyms, e.g. ‘Submit form A’ and ‘Submit form B’. Part-of relations, similar to kind-of relations, are more typical between activity labels of event logs recorded through free-text inputs.

The holonymy relationship between activities is defined by an object property called hasPart and a number of class restrictions that use existential quantifier someValuesFrom and universal quantifier allValuesFrom. For instance, assuming that forms A and B are the only documents that need to be submitted as part of an application, the OWL statements:

\[
\text{Class (Submit_documents Activity)} \\
\text{restriction (hasPart someValuesFrom Submit_form_A)} \\
\text{restriction (hasPart someValuesFrom Submit_form_B)} \\
\text{restriction (hasPart allValuesFrom (Submit_form_A or Submit_form_B))}
\]

mean that Submit_documents is the class of any activity that, among other things, has some Submit_form_A parts and also some Submit_form_B parts and has only Submit_form_A parts.
or Submit_form_B parts (Figure 1). The third restriction is a closure axiom without which, due to OWA, it could be possible that the Submit_documents activity would have some parts other than Submit_form_A and Submit_form_B. The holonomy relation is depicted by the orange arrows (with the arrow towards the meronym) labelled with ‘hasPart’ in Figure 1(c).

4) Antonymy relation

Two activity labels are antonyms if they represent opposite tasks, e.g. ‘Begin check’ and ‘End check’. Antonymous labels can be performed at any stage of a process; for example, they are common after decision points where an application, for instance, is either accepted (e.g. ‘Notify acceptance’) or rejected (e.g. ‘Notify rejection’).

The antonymy relationship is represented by disjoint axioms in the PAO. For example, the OWL statement:

\[
\text{DisjointClasses (~Begin\_check End\_check) }
\]

means that the classes Begin_check and End_check refer to completely different concepts (Figure 1). The DL reasoner will conclude that the members of the class Begin_check can never belong to the class End_check and vice versa. OWL classes can generally overlap (i.e., an individual of a class can also belong to another class) unless they are explicitly specified to be disjoint from one another. Furthermore, because the classes Begin_check and Start_check are already declared as equivalent, the reasoner concludes that Start_check and End_check classes are also disjoint. The antonymy relation is depicted by the red edges in Figure 1(c).

V. ONTOLOGY LEARNING APPROACH

Incorporating domain knowledge in process mining analysis leads to more reliable outcomes. Figure 2 depicts a novel gamified approach to ontology learning. The approach takes one or more event log(s) related to a process as the input and generates a PAO (or adds to an existing one) as the output. If more event logs related to the process become available, they can be used as the input for another round of ontology learning through the gamified system to enhance the existing PAO. The approach collects knowledge from multiple domain experts (i.e., through crowdsourcing), and uses a number of motivational drives [63] to engage them in the task of ontology learning. To further assess the level of domain expertise in participants, two strategies are used: the participants are asked to declare how familiar they are with the task of ontology learning. To enhance the level of domain expertise in participants, two strategies are used: the participants are asked to declare how familiar they are with the task of ontology learning. To further assess the level of domain expertise in participants, two strategies are used: the participants are asked to declare how familiar they are with

and used to build or enhance the PAO. The input event logs will be used to add individuals (events) to the PAO. The three phases are described in the following sections, where the set of all activities in the input event log(s) is referred to as the activity universe and is denoted by \( \mathcal{A} \).

A. GAME SETUP

The input event log(s) may contain a large number of distinct activity labels. In order to make the approach feasible, candidate label pairs that may have some sort of semantic relation are identified from the event log(s). For this purpose, the notion of context similarity as defined in previous work [75] is used. This notion is a weighted average of a number of similarities between activities: string, control flow, resource, data, and time. The idea is that activities that are semantically related (i.e., are synonyms, antonyms, or at different levels of abstraction) are performed at the same stage of the process (control flow similarity), by the same people or roles (resource similarity), at similar times, e.g. day of the week or month (time similarity), or they have similar data attributes (data similarity). Even in the case of antonyms, and this may be less intuitive, some sort of context similarity may exist. For example, ‘accept’ and ‘reject’ activities related to a particular decision are both performed at the same stage of a process.

Formally, given \( \mathcal{A} \), the activity universe, let \( S_{avg}(a, b) \) be the context similarity between activities \( a \) and \( b \) (note that \( S_{avg}(a, b) = S_{avg}(b, a) \)). Assume a total order \( \prec \subseteq \mathcal{A} \times \mathcal{A} \). The particular choice of order, e.g. lexicographical, does not matter, but its existence allows us to avoid having to examine symmetrical label pairs. If we know the relation between \( a \) and \( b \), we know the relation between \( b \) and \( a \) as well. As candidate questions, we consider activity pairs with a high enough context similarity, i.e. greater or equal to a given threshold \( \theta_s \). This set \( C \) is defined as:

\[
C = \{(a, b) \in \mathcal{A} \times \mathcal{A} \mid S_{avg}(a, b) \geq \theta_s \land a \prec b \}.
\]

The actual questions, recorded in \( Q \), are selected from \( C \). For practical reasons, their number is restricted and limited to those with the highest context similarity. The set should not contain redundant questions, i.e. questions to which the answer can be derived from other questions. Therefore, we exclude questions concerning \( a \) and \( b \) if we also have questions relating \( a \) and \( b \) resp. to \( c \). Formally, let \( n \) be the desired number of questions, then \( Q \subseteq C \) such that all of the following conditions hold:

\[
|Q| = n,
\]

\[
\exists c \in C \land Q \exists q \in Q[S_{avg}(c) > S_{avg}(q)],
\]

\[
\exists a, b, c \in \mathcal{A}[\{(a, b) \in Q \land ((a, c) \in Q \lor (c, a) \in Q) \land ((b, c) \in Q \lor (c, b) \in Q)\}].
\]
between two activities performed in the domain of the event log(s). The answer to a control question is known and is used to determine a participant’s level of domain expertise. The idea of using control questions is similar to the idea of using control words to assess the correctness of interpreted words from scanned old print materials by the reCAPTCHA system [76].

B. GAME PLAY
A gamified system called The Quality Guardian Rosebud is designed to motivate domain experts to participate in the ontology learning task. The game supports a number of motivation drives defined in the Octalysis gamification design framework [63]. Some drives are primarily reinforced in the design (i.e., primary drives), and some are maintained as side motivations (i.e., secondary drives). The three primary drives are as follows:

1) **Epic meaning**: People with this drive enjoy doing tasks for reasons beyond themselves, e.g. helping other less fortunate people. The game supports the epic meaning desire by describing the reasons for building the ontology, for example, helping to advance knowledge or to improve data quality.

2) **Development and accomplishment**: This is the internal desire of people to progress, achieve goals, overcome challenges, and acquire new knowledge. Collectable points and badges are designed to encourage participants to set in-game goals. Progress bars show the progress towards the goals (Figures 3 and 5). Different types of semantic relations between activities with some examples are presented in the game to support gaining knowledge.

3) **Social influence and relatedness**: This is the desire to obtain social acceptance, collaborate, and compete with others. The game supports this desire by ranking participants on a leaderboard (Figure 3) and showing a summary of the responses of other participants. These aim to give participants the feeling that they are involved in a social task.

In addition to the primary drives, two secondary drives are incorporated into the game design; they are as follows:

1) **Unpredictability and curiosity**: This drive motivates people to experience tasks that involve uncertainty or chance. In the game, participants are asked to spin two fortune wheels to find out what their next question is (Figure 4). This feeling of curiosity aims to encourage them to proceed to the next question.

2) **Creativity and feedback**: This is the desire for creativity to bring imagination to life. As participants determine the semantic relations between label pairs, a graph, where nodes are labels and edges are relations, is built for them in parallel (Figures 4 and 5). To support the creativity of participants, the game provides them with means to reshape, filter, and edit the graph.

Participants initially view the main screen of the game, which conveys the message of “Jointly building a knowledge base” to encourage them to contribute their knowledge. Next, participants see a demo video of the game (skippable), which describes activity label quality issues in event logs and that they can be resolved through the use of an activity ontology. Also, the demo teaches participants how to play the game, where nodes are labels and edges are relations, is built for them in parallel (Figures 4 and 5). To support the creativity of participants, the game provides them with means to reshape, filter, and edit the graph.

When people are given a choice, they tend to enjoy the task more than when they are given a single option [77]. To provide participants with a sense of choice, the input label pairs are categorised into multiple *topics* in the game. The categorisation can be based on different context perspectives, e.g. control flow (i.e., the label pairs that are performed in the

---

2The set of control questions may be created by a domain expert or it may be available as a domain-specific resource. This prerequisite is acknowledged to potentially impact the applicability of the approach to specific domains.

3Available at https://guardian.qistudio.io.
same stage of a process are in the same topic), or resource (i.e., the label pairs that are performed by similar roles are in the same topic). Topics can be viewed and selected on the user profile page (Figure 3). Each topic consists of a number of levels. Participants start with the first level, which has the lowest number of questions (the easiest level), and as they proceed to the next levels, they answer more questions. The increasing level of difficulty aims to support participants’ desire for competence and accomplishment [78].

A game level consists of a number of questions, each asking about the semantic relation between two specific activity labels. On the main board of the game (Figure 4), there are two fortune wheels, which participants need to spin to reveal the next question. The participants can select one of the options synonym, antonym, a hypernymy relation (i.e., hypernym, hyponym, or cohyponym), a holonymy relation (i.e., holonym, meronym, or comeronym), or no relation. They can view the description of these relations along with some examples. In case of selecting any of the three relations: synonym, cohyponym, and comeronym, participants are asked to provide an extra label. If they choose the synonym relation, they are asked to suggest a repair for the identified synonyms. They can either choose one of the labels as the repair, or suggest a new label through an input field. In case of the cohyponym and comeronym relations, the common hypernym or holonym is asked, respectively.

Three hints are offered for each question: (1) the absolute and relative frequency of the two labels in the input event log(s), (2) their context similarity based on their control flow, resource, time, and data information in the event log(s) using the similarity measures defined in previous work [75], and (3) the summary of the responses of other participants, that is, the semantic relations offered by other participants. The third hint aims to give participants the feeling that there are other participants, thus inspiring a sense of social presence.

During the game, the recorded responses of participants are demonstrated to them as a graph (Figure 4). The nodes denote labels’ ID and the edges represent the semantic relations between labels. As illustrated in Figure 4, different types of semantic relations are shown by different colours and shapes of edges. The synonymy and antonymy relations are depicted by green and red lines, respectively. A yellow arrow shows the hypernymy relation (with the arrow towards the hypernym), and a blue arrow with a diamond head shows the holonymy relation (with the diamond towards the holonym). The width of an edge between two nodes is larger if more participants have identified the corresponding relation between the two labels. For example, in Figure 4, the edge between nodes 53 and 54 is thicker than the edge between 54 and 55 because more participants think that the label with ID 53 is the meronym of the label with ID 54. Filtering options are also provided to allow focusing on relations of interest. The graph provides participants with an overall view of their responses, which can be revisited and changed if required.

Upon completing a level, participants earn a number of stars and diamonds. The stars measure participants’ game experience and are granted based on the number of questions answered in the level. The diamonds reflect the participants’ domain knowledge and are granted based on the correctness of the answers to the control questions embedded in each level, and the string score (defined in our previous work [26] as the repair score). The string score is computed for the new labels suggested by participants as the repair of synonym labels or as the shared hypernym of cohyponym labels or holonym of comeronym labels. The score considers the length of a string and whether it is free from special characters and numbers. The diamonds are used to prioritise answers accordingly in the post-game phase. To further inspire participants’ feelings of competence, they are awarded different badges, e.g. Phenom Contributor (Figure 3), once they gather a certain number of stars and diamonds.

Providing ways for people to compare their scores to others may make them feel more competitive [79]. A leaderboard is embedded on participants’ profile page (Figure 3), which enables them to compare the number of their stars and diamonds to their peers’ and aims to fulfil their need for social relatedness. As another feedback mechanism, the participants receive summative feedback when they complete all levels in a topic, as depicted in Figure 5. The feedback includes the total number of stars and diamonds earned in the topic and a graph that represents the semantic relations identified between label pairs in the topic. This feedback re-illustrates what participants have accomplished on a topic of the game to satisfy their needs for competence.

C. POST-GAME

In the post-game phase, the PAO is built using collected answers from participants. The method is to approve responses of higher weight, where the weight of an answer is denoted by the participant’s level of domain knowledge, represented by the number of diamonds gained in the game, and the participant’s domain expertise level, which they declared at the sign-up. In crowdsourcing studies, this idea is referred to as vote weighting [80].

Let \( \Lambda \subseteq Q \times \Phi \times \mathcal{R} \times \mathcal{P} \) be the set of all game responses, where \( Q \subseteq A \times A \) is the label question bank generated in the game setup phase (Section V-A), each question consisting of a pair of labels, \( \Phi = \{ \text{synonym}, \text{antonym}, \text{hypernym}, \text{hyponym}, \text{cohyponym}, \text{holonym}, \text{meronym}, \text{comeronym}, \text{no relation}\} \) is the set of all semantic relations, \( \mathcal{R} \subseteq \Sigma^* \) is the set of all extra labels suggested by participants (in case of synonym, cohyponym, and comeronym relations) where \( \Sigma \) is the set of all characters, and \( \mathcal{P} \) is the set of participants. An answer \( \lambda = (\lambda_q, \lambda_p, \lambda_r, \lambda_r) \) where the constituent parts refer to the question, the semantic relation, the extra label, and the participant. The extra label for relations other than synonym, cohyponym, and comeronym is an empty string \( \varepsilon \).

\[
\forall \lambda \in \Lambda[\lambda_p \notin \{ \text{synonym}, \text{cohyponym}, \text{comeronym} \} \Rightarrow \lambda_r = \varepsilon]
\]
The function $W: \Lambda \rightarrow [0, +\infty)$, assigns a weight to an answer $\lambda \in \Lambda$ and is defined as $W(\lambda) = \tau_{\text{knowl}} * k(\lambda_p) + \tau_{\text{expert}} * e(\lambda_p)$ where $k: P \rightarrow N$ and $e: P \rightarrow N$ are the functions that assign values to the demonstrated knowledge and the declared expertise of the participant $\lambda_p$ and $\tau_{\text{knowl}}, \tau_{\text{expert}} \geq 0$ are the relevant coefficients.

For any question $q \in Q$, any semantic relation $\varphi \in \Phi$, and any extra label $r \in R$ the total weight function $Total: Q \times \varphi \times R \rightarrow [0, +\infty)$ is defined as $Total(q, \varphi, r) = \sum_{p \in P, (q, \varphi, r, p) \in \Lambda} W(q, \varphi, r, p)$ which is the sum of weights of the answers that have provided extra label $r$ for the identified relation $\varphi$ between the labels in question $q$. To compute this summation for all extra labels for the relation $\varphi$ in question $q$, the function $Total(q, \varphi) = \sum_{r \in R} Total(q, \varphi, r)$ will be used.

$\Pi$ is the set of final semantic relations to be used for building the PAO (Equation 6). For every question $q$, the relation $\varphi_{\text{max}}$ with the maximum total weight will be selected if its total weight is more than a weight threshold $\theta_w$.

Furthermore, the extra label $r_{\text{max}}$ with the maximum weight for the relation $\varphi_{\text{max}}$ between labels of the question $q$ will be approved.

$$\Pi = \{(q, \varphi_{\text{max}}, r_{\text{max}}) \in Q \times \Phi \times R | \varphi_{\text{max}} \in \text{arg}_{\varphi} \max Total(q, \varphi) \land Total(q, \varphi_{\text{max}}) \geq \theta_w \land r_{\text{max}} \in \text{arg}_{r} \max Total(q, \varphi_{\text{max}}, r)\}$$

**VI. EVALUATION**

The Quality Guardian Rosebud game has been implemented as a Java web application deployed on a Tomcat server. This gamified system is domain-independent; that is, no domain-specific considerations are made in the design of this approach. The semantic relations distinguished in this approach, e.g., synonymy, antonymy, hypernymy, and holonymy, are general and can exist between activity labels from different domains, as they are adapted from the semantic relations defined in WordNet [16]. Also, the motivational drives and game elements supported in the gamified system design apply to participants from different domains, as they are based on generic gamification frameworks, e.g., Octalysis [63] and Self-Determination Theory [64]. However, for the evaluation, a public real-life event log [81] that is the record of the building permit application processes executed in a Dutch municipality is used as a case study. Considering

![FIGURE 3. The view of the user profile page of the game](image-url)
Spin the wheels to reveal a question. Then answer that by selecting the semantic relation between the two activity labels. Overall, you have 6 attempts to spin, answer, and earn points.

**FIGURE 4.** The view of the main board of the game

**FIGURE 5.** The view of the topic summative feedback of the game
the constraints of the evaluation, e.g., public vs. private event logs and recruiting sufficient participants, evaluating this gamified system with event logs from multiple domains is out of the scope of this study. Details of the evaluation are described in the following sections.

A. DATA

The public BPIC15_2 event log [81] was used for the evaluation. This event log contains data related to a building permit application process in a Dutch municipality covering applications received over a period of approximately four years. The original event log contains 832 cases, 44354 events, and 410 distinct activity labels. However, incomplete cases (identified by the data attribute ‘caseStatus’) and infrequent events (identified by the ProM plugin ‘Filter Log using Simple Heuristics’ with threshold 85%) were filtered out from the log to focus on the main process behaviour. The filtered log\(^6\) contains 616 cases, 30813 events, and 82 distinct activity labels.

In the game setup phase, Equations 1–4 were used to generate \( n = 36 \) game questions from the filtered log. The weighted average similarity \( S_{avg}(a, b) \) was computed based on the control flow, resource, data, time, and string similarities [75] with the weights 2, 2, 1, 1, and 1, respectively. Higher weights are assigned to the control flow and resource dimensions because they are more informative for assessing the similarity of activities [75]. A similarity threshold \( \theta \) of 0.4 was used so that more activities could be investigated by the similarity of activities [75].

A total of 36 label pairs (i.e., questions), made up of 47 distinct activity labels, and 9 control questions were included in the game. The control question bank was generated using the available resources in the municipality domain\(^7\). The label questions were divided into three topics representing three stages of the building permit application process: (1) application lodgement and initial checks, (2) decision making, and (3) notification and application closure activities.

The ‘action_code’ data attribute in the log, which indicates the stage of an activity, is used to categorise questions into the three topics. Each topic consists of three levels: level 1, which is made up of three label questions and one control question, level 2, which consists of four label questions and one control question, and level 3, which consists of five label questions and one control question.

B. PROCEDURE

The game was made publicly available online for 45 days. It was advertised to QUT staff and students, BPM and process mining researchers and practitioners on LinkedIn and through the international academic networks. The participants played the game and were scored using stars and diamonds as described in Section V-B. The maximum number of stars and diamonds that could be earned in each topic was 900 and 300, respectively. The four XP badges Good Starter, Newbie Contributor, Phenom Contributor, and Chief Contributor were granted to participants when they accumulated 100, 600, 1500, and 2700 stars. Similarly, the three knowledge badges Problem Solver, Expert, and Master were awarded once 200, 500, and 900 diamonds were collected.

In the post-game phase, the answers were weighted according to function \( W \) described in Section V-C. The function \( k \) assigns the values 1, 2, 3, or 4 to the demonstrated knowledge of a participant if they have received zero, one, two, or three knowledge badge(s), respectively. Similarly, function \( e \) assigns the values 1, 2, 3, or 4 to the declared expertise of a participant if they have declared (in the sign-up form) to know nothing, a little, a fair bit, and a lot about the building permit application process. Equation 6 is applied to choose the best semantic relations. The value of the threshold \( \theta_w \) is selected in such a way that the type of semantic relation between two activities is approved only if they have a weighted vote of at least half of the maximum possible vote. A comprehensive list of parameters is available online\(^8\).

The PAO (referred to as the game PAO) is created from the selected semantic relations by a Java program incorporating the OWL API [82] and the HermiT [70] reasoner\(^9\).

A ground truth set of semantic relations between candidate label pairs of the BPIC15_2 event log [81] was created through an interview with a domain expert, who is very familiar with the process and the activity labels in the event log\(^10\). The expert was introduced to different types of semantic relations between activity labels: synonymy, hyponymy, holonymy, and antonymy. The candidate label pairs that were provided as the input to the gamified system were presented to the domain expert who then identified the semantic relation of each pair. The collected relations were used to create the ground truth PAO through the same Java program which created the game PAO.

C. ANALYSIS METHOD

The similarity between the game PAO and the ground truth PAO is computed. The Semantic Cotopy (SC) notion defined by Maedche et al. [83] is adapted for this purpose. The authors used SC to compare two ontologies \( O_1, O_2 \) based on their Taxonomy Overlap (TO), where taxonomy is the superclass sub-class hierarchy. More specifically, the semantic cotopy of a class \( C_1 \) in the ontology \( O_1, SC(C_1, O_1) \), is the set of all classes of \( O_1 \) that are either super-class or sub-class of \( C_1 \). The TO of \( O_1 \) and \( O_2 \) based on \( C_1 \) is the ratio between the size of the intersection and the union of \( SC(C_1, O_1) \) and \( SC(C_1, O_2) \). The overall TO of \( O_1 \) and \( O_2 \) is the average TO over all classes in \( O_1 \).

The SC notion [83] is expanded to cover other types of semantic relations. Specifically, four cotopy notions are defined

\(^6\)Available at https://event-log-quality.s3.amazonaws.com/access/log.xes.
\(^7\)Such as https://business.gov.nl/regulation/all-in-one-permit-building.
\(^8\)At https://event-log-quality.s3.amazonaws.com/access/Pars.pdf.
\(^9\)The program codes are available at https://event-log-quality.s3.amazonaws.com/access/Code.zip.
\(^10\)It is acknowledged that the ground truth ontology is not built by a group of experts.
to evaluate the PAO: (1) Hypernymy Cotopy, which looks at super-class sub-class relations defined in Section IV-C2 (same as the SC [83]), (2) Holonymy Cotopy, which focuses on part-of relations defined in Section IV-C3, (3) Synonymy Cotopy, which looks at equivalence relations defined in Section IV-C1, and (4) Antonymy Cotopy, which focuses on disjoint relations defined in Section IV-C4. The four similarity measures Overall Hypernymy Overlap (HyO), Overall Holonymy Overlap (HoO), Overall Synonymy Overlap (SO), and Overall Antonymy Overlap (AO) are computed based on each cotopy notion, respectively.

While approaches to ontology learning are already proposed in the literature [5], [13], [27], they cannot be directly compared with this approach. The reason is that the type of input data they accept is not an event log. Even the few ontology learning approaches [44]–[46] in the area of process mining take inputs other than event logs, e.g. a set of process models. Furthermore, the structure of the ontology that these approaches create is different from the structure of the PAO: they do not formalise semantic relations of activity labels in their ontologies. Another possible setting for the evaluation could be to create an ontology of activity labels in an event log using a purely technical method and compare the quality of the acquired ontology with the ontology created through gamification. However, we did not opt for that setting in this paper. We only focused on how gamification can be used to improve user engagement in the tedious task of ontology creation.

To measure participants’ engagement in the gamified system, they were asked to fill out a survey consisting of 15 questions with 5 point Likert scale answers (strongly agree–strongly disagree). Each question is about one desirable feature of a gamified system based on the Octalysis framework [63], the GameFlow evaluation framework [78], and Self-Determination Theory (SDT) [64], [65] or the Player Experience of Need Satisfaction (PENS) model [84]. GameFlow aims to measure people’s enjoyment with a gamified system. The items are mapped from the flow theory [85] to games. SDT describes psychological needs that motivate humans to act, and PENS adds further game-related concepts to this theory. Table 1 shows a comprehensive list of the items measured in each question and the relevant feature in each of the frameworks. The Cronbach’s alpha coefficient [86] is computed to measure the reliability of the survey responses.

D. RESULTS

A total of 40 participants played the game, out of which 35 participants were considered for the evaluation because they completed at least one topic and the survey. More specifically, 9 participants finished one topic, 2 completed two topics, and 24 answered the questions of all three topics.

Figure 6 illustrates the semantic relations identified by the game participants compared to the ground truth (GT)\(^\text{11}\). As shown in Figure 6, the semantic relation of 30 out of 36 label pairs was identified through the use of the gamified system (i.e., \(|II| = 30\) in Equation 6). All semantic relations were identified in the first topic, which had the highest number of participants (i.e., 35). There were 2 missing relations in the second topic (i.e., for label pairs number 17 and 24) with 26 participants, and 4 missing relations in the third topic (i.e., for label pairs number 27, 28, 32, and 34) with 24 participants. A higher number of participants results in more approved relations because it increases the weight of responses allowing them to pass the threshold \(\theta_w\). No antonymy relation has been identified by participants, which conforms to the ground truth. Overall, the type of relations recognised in the game match the ground truth except for 5 questions (i.e., numbers 2, 12, 16, 22, and 33). A lower weight threshold \(\theta_w\) can lead to more approved answers and a higher threshold can result in more correct answers.

Table 2 summarises the similarity measures \(\text{HyO}, \text{HoO}, \text{SO},\) and \(\text{AO}\) between the game PAO and the ground truth PAO. As reported in Table 2, there is a high overall hypernymy, holonymy and synonymy overlap between the two ontologies. The overall antonymy overlap is 0 because there exists no antonym relation in any of the ontologies. The results of Figure 6 and Table 2 confirm that the gamified system has been successful in creating a high-quality process activity ontology.

Figure 7 demonstrates the user engagement survey results categorised based on the frameworks (a) Octalysis [63], (b) GameFlow [78], and (c) SDT [64]. The overall Cronbach’s alpha coefficient of 0.8918 in this survey shows that it is of acceptable reliability. Participants found the system engaging overall (\(M=3.74, SD=1.01\)), which implies that they had a positive user experience. Figure 7(a) shows that development and accomplishment (\(M=3.63, SD=1.06\)), epic meaning (\(M=3.59, SD=1.11\)), and creativity and feedback (\(M=3.53, SD=1.08\)) were the top motivations for participants to engage with the gamified system. However, participants were neutral about the effectiveness of the unpredictability and curiosity (\(M=2.86, SD=1.24\)) and the social influence and relatedness (\(M=3.1, SD=1.32\)) motivation drives. The former was mainly because unpredictability was not given as much prominence in the game design, it being a secondary motivation drive, and the latter can be explained by the survey being presented to the participants after the first topic with most of them not having seen the leaderboard yet. The 11 participants who completed the survey at the end of the second or the third topic were more influenced by the social motivation (\(M=4.05, SD=1.07\)); this is because participants can only choose the next topic to work on through their profile page, where the leaderboard is located (Figure 3). As depicted in Figures 7(b) and (c), the average score for all features is between 3 and 4. Overall, participants’ responses show evidence that gamification is a promising method to achieve a good user experience in the task of identifying semantic relations between activities.
TABLE 1. The survey items mapped to the relevant features in the three frameworks: Octalysis [63], GameFlow [78], and SDT/PENS [64], [65], [84]

| #  | Item                              | Octalysis [63]                | GameFlow [78] | SDT/PENS [64], [65], [84] |
|----|-----------------------------------|------------------------------|---------------|---------------------------|
| 1  | Clear game goals                  | Epic meaning and calling     | Clear goals   | Purpose                   |
| 2  | Useful feedback                   | Development and accomplishment| Feedback      | Competence                |
| 3  | Competition                       | Social influence and relatedness | Social Interaction | Relatedness |
| 4  | Cooperation                       | Social influence and relatedness | Social Interaction | Relatedness |
| 5  | Controllable interface            | -                            | Control       | Intuitive Controls        |
| 6  | Appropriate level of challenge    | Development and accomplishment | Challenge     | Competence                |
| 7  | Appropriate pace of challenge growth | Development and accomplishment | Challenge     | Competence                |
| 8  | Learning                          | Development and accomplishment | Player skills | Competence                |
| 9  | Concentration                     | -                            | Concentration | -                         |
| 10 | Involvement in the game           | -                            | Immersion     | Presence/Immersion         |
| 11 | Availability of options and choices | Creativity and feedback  | -            | Autonomy                  |
| 12 | Feeling of liberty                | Creativity and feedback      | -            | Autonomy                  |
| 13 | Helpfulness for the community     | Epic meaning and calling     | -            | -                         |
| 14 | Curiosity to reveal the fortune wheels | Unpredictability and curiosity | -            | -                         |
| 15 | Overall engagement                | -                            | -            | -                         |

FIGURE 6. The semantic relations identified in the game compared to the ground truth (GT)

FIGURE 7. Summary of the survey scores for the features of the frameworks (a) Octalysis [63], (b) GameFlow [78], and (c) SDT [64]

TABLE 2. The similarity measurements between the game PAO and the ground truth PAO

| (HyO) | (HoO) | (SO) | (AO) |
|-------|-------|------|------|
| 0.8840 | 0.8693 | 0.8904 | 0.0000 |

VII. CONCLUSION

A novel approach to learning a process activity ontology from event logs is proposed. The method uses crowdsourcing and gamification techniques to enhance user experience in the task of ontology learning. The semantic relations covered in the process activity ontology are synonymy, hypernymy, holonymy, and antonymy between activity labels. The results with 35 participants show that the semantic relations of 30 out of 36 label pairs are identified through the gamified approach. A higher number of participants leads to detecting more semantic relations. Out of the 30 semantic relations diagnosed, 25 match the ground truth. Also, the PAO created through the gamified system is highly similar to the ground truth PAO. The survey results show that participants found the system engaging overall. After creation, a process activity ontology can be used for activity label quality improvement, i.e., to repair synonymous labels and activity labels at different levels of granularity in event logs. Furthermore, the ontology can be extended in the future to include other semantic relations, e.g., control flow relations between activities where it is indicated, for example, that an activity must always be performed before another activity in a process and can
never occur before. Such relations can further contribute to the repair of data quality problems in event logs.

VIII. ACKNOWLEDGMENT

We are grateful for the contributions of Dr Joos Buijs, who acted as the domain expert. This study has been approved by the QUT Human Research Ethics Committee (approval number 2000000047).

REFERENCES

[1] W. M. P. van der Aalst, Process Mining: Data Science in Action. Springer, 2016.

[2] R. J. C. Bose, R. S. Mans, and W. M. P. van der Aalst, “Wanna Improve Process Mining Results - It’s High Time We Consider Data Quality Issues Seriously,” in Computationl Intelligence and Data Mining Symposium. Singapore: IEEE, 2013, pp. 127–134.

[3] S. Suriadi, R. Andrews, A. H. M. ter Hofstede, and M. T. Wynn, “Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs,” Information Systems, vol. 64, pp. 152–160, 2017.

[4] T. Baier, J. Mending, and M. Weske, “Bridge abstraction layers in process mining,” Information Systems, vol. 46, pp. 123–139, 2014.

[5] M. Shamsfard and A. Barforoush, “The state of the art in ontology learning: a framework for comparison,” The Knowledge Engineering Review, vol. 18, no. 4, pp. 293–316, 2003.

[6] D. Dou, H. Wang, and H. Liu, “Semantic data mining: A survey of ontology-based approaches,” in International Conference on Semantic Computing. Anaheim, CA, USA: IEEE, 2015, pp. 244–251.

[7] A. Conde, M. Larrañaga, A. Arruarte, and J. A. Elorriaga, “A combined approach for eliciting relationships for educational ontologies using general-purpose knowledge bases,” IEEE Access, vol. 7, pp. 48 339–48 355, 2019.

[8] Y. Yang, K. Cormcian, and M. Yu, “Ontology learning for systems engineering body of knowledge,” IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 1039–1047, 2021.

[9] A. Hipsipley, D. Cheng, and K. Ahmad, “The head-modifier principle and multilingual term extraction,” Natural Language Engineering, vol. 11, no. 2, pp. 129–157, 2005.

[10] S. A. Caraballo, “Automatic construction of a hypynnem-labeled noun hierarchy from text,” in Annual Meeting of the Association for Computational Linguistics. Maryland, USA: ACL, 1999, pp. 120–126.

[11] R. Lima, B. Espinispe, H. Oliveira et al., “An inductive logic programming-based approach for ontology population from the web,” in Database and Expert Systems Applications, ser. LNCS, vol. 8055. Prague, Czech Republic: Springer, 2013, pp. 319–326.

[12] J. Mending, H. Leopold, and F. Pitke, “25 challenges of semantic process et al., “Predicting the quality of process model matching,” in Business Process Management Conference (BPM), ser. LNCS, vol. 9850. Rio de Janeiro, Brazil: Springer, 2016, pp. 124–140.

[13] M. Hazman, S. R. El-Beltagy, and A. Rafea, “A survey of ontology learning approaches,” International Journal of Computer Applications, vol. 22, no. 9, pp. 36–43, 2011.

[14] H. Davalcu, S. Vadruv, S. Nagarajar, and I. Ramakrishan, “Ontominer: bootstrapping and populating ontologies from domain-specific web sites,” IEEE Intelligent Systems, vol. 18, no. 5, pp. 24–33, 2003.

[15] C. Ma and B. Molnar, “Ontology learning from relational database: Opportunities for semantic information integration,” Vietnam Journal of Computer Science, 2020.

[16] N. Tuy, D. J. Dong, and V. T. Thao, “Ontology construction by collaboration of non-experts and crowd-outsourcing: The case of the effect of diet on health,” in International Conference on Semantic Web Conference (ISWC), ser. LNCS, vol. 2342. Sardina, Italy: Springer, 2002, pp. 221–235.

[17] J.-H. Kim, H. Kwon, D.-H. Kim et al., “Building a service-oriented ontology for wireless sensor networks,” in International Conference on Computer and Information Science (ICIS). Portland, Oregon,USA: IEEE, 2008, pp. 649–654.

[18] A. Zrebacke, A. Adel, K. Jacksi, and A. Selamat, “Designing an ontology of e-learning system for duhok polytechnic university using Protégé OWL tool,” Journal of Advanced Research in Dynamical Control Systems, vol. 11, no. 5, pp. 34–37, 2019.

[19] A. Maedch and S. Staab, “Ontoedit: Collaborative ontology development for the semantic web,” in International Semantic Web Conference (ISWC), ser. LNCS, vol. 3242. Sardina, Italy: Springer, 2002, pp. 221–235.

[20] H. Dong and F. K. Hussain, “SOF: a semi-supervised ontology-learning-based focused crawler,” Concurreny and Computation: Practice and Experience, vol. 25, no. 12, pp. 1755–1770, 2013.

[21] Q. Wang, G. Ding, and S. Yu, “Crowdsourcing mode-based learning activity flow approach to promote subject ontology generation and evolution in learning,” Interactive Learning Environments, vol. 27, no. 7, pp. 965–983, 2019.

[22] E. A. Kardinata and N. A. Rakhmawati, “Integration of crowdsourcing and multilingual term extraction,” in Knowledge and Agent Technology: Interdisciplinary Perspectives on the Semantic Web, ser. LNCS, vol. 124, no. 3, pp. 317–327, 2020.

[23] M. N. Asim, M. Wasiim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, “A survey of ontology learning techniques and applications,” Database, vol. 2018, p. bau101, 2018.

[24] J. Howe, “The rise of crowdsourcing,” Wired magazine, vol. 14, no. 6, pp. 1–4, 2006.

[25] S. Deterding, M. Sicart, L. E. Nacke et al., “Gamification: Using Game-Design Elements in Non-Gaming Contexts,” in International Conference on Human Factors in Computing Systems, Extended Abstracts Volume. Vancouver, BC, Canada: ACM, 2011, pp. 2425–2428.

[26] G. A. Miller, “WordNet: a lexical database for English,” Communications of the ACM, vol. 38, no. 11, pp. 39–41, 1995.

[27] H. van der Aa, A. Gal, H. Leopold et al., “Instance-Based Process Matching Using Event-Log Information,” in International Conference on Advanced Information Systems Engineering (CAiSE), ser. LNCS, vol. 10253. Essen, Germany: Springer, 2017, pp. 283–297.

[28] A. Koschmider, M. Ullrich, A. Heine et al., “Revising the Vocabulary of Business Process Element Labels,” in International Conference on Advanced Information Systems Engineering (CAiSE), ser. LNCS, vol. 9091. Stockholm, Sweden: Springer, 2015, pp. 69–83.

[29] N. Tax, E. Alasgarov, N. Sidorova et al., “Generating Time-Based Label Refinements to Discover More Precise Process Models,” Journal of Ambient Intelligence and Smart Environments, vol. 11, no. 2, pp. 165–182, 2019.
et al.

[59] G. Rallapalli, D. G. Saunders, K. Yoshida, A. Edwards, C. A. Lugo, H. Hodson, “Gamers unleash swarms of nanoparticles on tumours,”

[50] A. Rebmann and H. van der Aa, “Extracting semantic process information”

[49] S. De Cesare, D. Juric, and M. Lycett, “Toward the automation of business natural language processing: a multiple case study,”

[47] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, N. Z. Legaki, N. Xi, J. Hamari, K. Karpouzis, and V. Assimakopoulos, M. A. Luengo-Oroz, A. Arranz, and J. Frean, “Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears,”

[51] B. M. Good, S. Loguercio, O. L. Griffith, M. Nanis, C. Wu, and A. I. Su, “The Cure: Design and Evaluation of a Crowdsourcing Game for Gene Selection for Breast Cancer Survival Prediction,”

[52] R. Tinti, M. Luczak-Roesch, E. Simperl, and W. Hall, “An investigation of player motivations in Eyewire, a gamified science project,”

[53] M. A. Luengo-Oroz, A. Arranz, and J. Frean, “Crowdsourcing to process ontology models to process ontology: A case study,”

[54] A. Coskuncay, O. Gurbuz, and O. Demirors, “Transformation from business process models to process ontology in the context of statistics education,”

[55] T. Edwards, “Challenging games help students learn: An empirical study with a multiplayer online game,”

[56] Y.-K. Chou, Actionable gamification: Beyond points, badges, and leaderboards. Packt Publishing Ltd, 2019.

[57] E. L. Deci and R. M. Ryan, “Self-determination theory,” in Handbook of theories of social psychology. Sage Publications Ltd, 2012, vol. 1, ch. 20, pp. 416–436.

[58] D. H. Pink, Drive: The Surprising Truth About What Motivates Us. Riverhead Books, 2009.

[59] T. Berners-Lee, J. Hendler, and O. Lassila, “The Semantic Web,” Scientific American, vol. 284, no. 5, pp. 34–43, 2001.

[60] T. Berners-Lee and M. Fischetti, Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor. Harper San Francisco, 1999.

[61] G. Antoniou and F. Van Harmelen, “The Semantic Web Vision,” in A Semantic Web Primer, 2nd ed. MIT press, 2008, ch. 1, pp. 1–21.

[62] A. Rector, N. Drummond, M. Horridge et al., “OWL: Pizzas: Practical Experience of Teaching OWL-DEL: Common Errors & Common Patterns,” in International Conference on Engineering Knowledge in the Age of the Semantic Web, ser. LNCS, vol. 3257. Whitley Hall, UK: Springer, 2004, pp. 63–81.

[63] B. Glimm, I. Horrocks, B. Motik et al., “Hermit: an OWL 2 reasoner,” Journal of Automated Reasoning, vol. 53, no. 3, pp. 245–269, 2014.

[64] D. Tsarkov and I. Horrocks, “PaCT++ Description Logic Reasoner: System Description,” in International Joint Conference on Automated Reasoning, ser. LNCS, vol. 4130. Seattle, WA, USA: Springer, 2006, pp. 292–297.

[65] V. Haarslev, K. Kidde, R. Möller et al., “The RacerPro knowledge representation and reasoning system,” Semantic Web Journal, vol. 3, no. 3, pp. 267–277, 2012.

[66] I. Horrocks, P. F. Patel-Schneider, and E. Van Harmelen, “From SHIQ and RDF to OWL: the making of a Web Ontology Language,” Journal of Web Semantics, vol. 1, no. 1, pp. 7–26, 2003.

[67] S. Bunk, F. Pittke, and J. Mendling, “Aligning process model terminology with hypernym relations,” in International Symposium on Data-Driven Process Discovery and Analysis, ser. LNBP, vol. 244. Vienna, Austria: Springer, 2017, pp. 105–123.

[68] S. Sadeghi-Nasiri, A. H. M. ter Hofstede, M. T. Yynn, and S. Suriai, “A contextual approach to detecting synonymous and polluted activity labels in process event logs,” in International Conference on Cooperative Information Systems (CoopIS), ser. LNCS, vol. 11877. Rhodes, Greece: Springer, 2019, pp. 76–94.

[69] L. Von Ahn, B. Maurer, C. McMillen et al., “reCAPTCHA: Human-based Character Recognition via Web Security Measures,” Science, vol. 321, no. 5895, pp. 1465–1468, 2008.

[70] S. S. Iyengar and M. R. Lepper, “When choice is motivating: Can one desire too much of a good thing?” Journal of personality and social psychology, vol. 79, no. 6, pp. 995–1006, 2000.

[71] P. Tweetser and P. Weyt, “Gamflow: a model for evaluating player enjoyment in games,” Computers in Entertainment, vol. 3, no. 3, p. 3, 2005.

[72] S. Nobel, M. Bgee, S. Schneider, and G. D. Rey, “The higher the score, the higher the learning outcome? heterogeneous impacts of leaderboards and choice within educational videogames,” Computers in Human Behavior, vol. 65, pp. 391–401, 2016.

[73] M. S. Hardas and L. Purvis, “Bayesian vote weighting in crowdsourcing systems,” in International Conference on Computational Collective Intelligence, ser. LNCS, vol. 7653. Ho Chi Minh City, Vietnam: Springer, 2012, pp. 192–203.

[74] van Dongen B.F., “BPI Challenge 2015 Municipality 2. 4TU.ResearchData. Dataset.” https://doi.org/10.4121/uuid: 63a8435a-077d-4ece-97cd-2c76d394d99c.

[75] M. Horridge and S. Bechhofer, “The OWL API: A Java API for OWL Ontologies,” Semantic Web, vol. 2, no. 1, p. 11–21, 2011.

[76] A. Maedche and S. Staab, “Measuring similarity between ontologies,” in International Conference on Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, ser. LNCS, vol. 2473. Sigenza, Spain: Springer, 2002, pp. 251–263.

[77] R. M. Ryan, C. S. Rigby, and A. Przybylski, “The Motivational Pull of Video Games: A Self-Determination Theory Approach,” Motivation and Emotion, vol. 30, no. 4, pp. 384–396, 2006.

[78] M. Csikszentmihalyi, Flow: the psychology of optimal experience. Harper & Row, 1990.

[79] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” Psychometrika, vol. 16, no. 3, pp. 297–334, 1951.
MRS SAREH SADEGHIANASL is a PhD candidate at the Business Process Management Group of the School of Information Systems, Queensland University of Technology in Australia. Her research focuses on improving the quality of input data, that is, an event log, to prepare it for process mining, which concerns analysing organisational processes. More recently, she is looking at using gamification techniques for data cleaning to make it more fun, as it is currently the most tedious and time-consuming task of every data analysis project. She has a Computer Science background and has skills in software engineering, algorithm design, and programming.

PROF ARTHUR H.M. TER HOFSTEDE received his PhD degree from the Katholieke Universiteit Nijmegen (since renamed to Radboud Universiteit), Nijmegen, The Netherlands, in 1993. He has worked at Queensland University of Technology, Brisbane, Australia, since 1997, and is currently a professor and the Head of the School of Information Systems. His research interests lie in the area of business process management, in particular business process automation and process mining. He was involved in the well-known workflow patterns initiative and at QUT he has managed the well-known YAWL initiative. He is a co-author on over 200 publications, including over 100 journal publications.

PROF MOE THANDAR WYNN completed her PhD in the area of workflow management in 2007 from Queensland University of Technology, Brisbane, Australia. She leads the Business Process Management Research Group at QUT. She is Vice-Chair and one of the steering committee members of the IEEE Taskforce on Process Mining. She has published 80+ refereed papers, including 30+ journal articles. Her ongoing research focuses on process-oriented data mining (process mining), data quality and robotic process automation for the digital transformation of processes.

DR SELEN TURKAY is a lecturer in Human Computer Interaction at Queensland University of Technology (QUT). Her research interests include the design of personalised, interactive, and collaborative immersive environments. Specifically, she studies the effects of design choices on user agency and wellbeing outcomes, as well as user experiences including engagement and motivation in variety of contexts including education and health. Prior to joining QUT, Dr Turkay worked as a research scientist and a post-doctoral research fellow at Harvard University. She earned her doctoral degree in Instructional Technology and Media at Columbia University Teachers College.

PROF TRINA MYERS is a Computer Scientist and the Associate Dean (Learning and Teaching), Faculty of Science at Queensland University Technology. Trina is the current President of the Australian Council of Deans of ICT (AC-DICT) that aims to promote and advance ICT education, research and scholarship on behalf of Australian universities. Trina’s research interests span both Technology-based research and Learning and Teaching-based research. Trina’s Technology-based research interests focus predominately on semantic technologies, data and knowledge management, collective intelligence and the Internet of Things (IoT). She works predominantly in multi-disciplinary research projects that have included other disciplines such as Health Science, Conservation Management and Business.