Electronic Preprint for Journal of Information Processing Vol.24 No.1

Regular Paper

An Ontology-based Framework for Semantic Reconciliation in Humanitarian Aid in Emergency Information Systems

Pasinee Apisakmontri[1,2,a) Ekawit Nantajeewarawat[2,b)] Mitsuru Ikeda[1,c)]
Marut Buranarach[3,d)]

Received: April 2, 2015, Accepted: October 2, 2015

Abstract: Humanitarian aid in an emergency information system involves information from multidisciplinary environments. A lot of information is stored in relational databases. Semantic interoperability between existing relational databases and ontologies still remains a major practical issue. In order to avoid a combinatorial explosion of terminology alignment among different systems, we designed a pivot ontology framework, and present a pivot construction methodology and a PivotOntology-to-Database schema matching methodology. The first methodology is adopted from an ontology engineering technique, and the second one is based on a linguistic relation approach. To integrate humanitarian aid in emergency information from several databases, the Humanitarian Aid for Refugees in Emergencies (HARE) ontology has been proposed. Coverage of the HARE ontology is evaluated with respect to comparison against knowledge sources, and matching with existing systems. The evaluations demonstrate that the HARE ontology is broadly compatible with existing database schemas.

Keywords: Ontology, Pivot Ontology, Humanitarian Aid in Emergency, Relational Database, schema matching, interoperability, knowledge representation

1. Introduction

During a critical period of a disaster, various humanitarian actors such as governments, relief organizations, volunteers, and affected people often gather information and create information systems independently with little consolidation. Such information is incomplete and sometimes creates a conflicting picture of humanitarian needs [17], resulting in limited collaboration between different systems for information collection, sharing, and dissemination. Interoperability between humanitarian aid information systems is needed in order that they perform more smartly and more effectively. Achieving that goal requires collaborative technologies, which often deeply involve information sharing and domain knowledge. Some of the Disaster Management Services (DMSs) have emerged from disaster situations in the past. For example, the Sahana disaster information system, which is an open-source software application, was initiated after the Sumatra-Andaman earthquake in 2004 [6].

Not only for human communication, collaboration should create new ways for computer communication through which humanitarian actors can disseminate their information. Humanitarian aid information, including information on the occurrences of disaster situations, victims, shelters, resources, facilities, etc., is usually heterogeneous, rapidly changeable, ambiguous, and large. It is widely distributed and owned by different organizations [38], and as such, it is stored diversely in distinct heterogeneous data sources in different locations. Successful and innovative collaboration solutions are limited by a large number of humanitarian actors and incompatible information. An important challenge of information integration in the humanitarian aid domain is to identify the correlation of data from multiple sources [10].

An ontology enables one to reuse and share application domain knowledge using a common vocabulary across heterogeneous applications. By its definition ([22], [34]), an ontology is a hierarchically structured set of terms and some specification of their meanings to define a structure on the domain and constrain the probable interpretations of terms that can be used as a skeletal foundation for a knowledge base. An ontology provides a promising approach for dealing with semantic heterogeneity problems. Geographic information, which is closely related to disaster information, has been represented using ontologies [18], [37]. An ontology for emergency preplanning [16] and a sensor ontology for emergency management [7] have been developed. Although there are existing ontologies for disaster management, these ontologies are application-dependent.

As mentioned above, information integration is a major is-

1 School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa 923–1211, Japan
2 School of Information, Computer, and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University, Khlong Luang, Pathum Thani, Thailand
3 National Electronics and Computer Technology Center, Khlong Luang, Pathum Thani, Thailand
a) pasinee.a@jaist.ac.jp
b) ekawit@siit.tu.ac.th
c) ikeda@jaist.ac.jp
d) marut.buranarach@nectec.or.th

sue in this domain. Most developed applications are based on Relational-Databases (RDBs). Many humanitarian aid organizations developed their own databases. Large-scale database integration becomes a critical process. A database-to-database integration is a general problem that normally requires database-to-database schema matching [27]. Schema matching between a large number of different database schemas often heavily relies on many database administrators and is time consuming. Our objective is to provide a basis for common understanding of terms related to humanitarian aid in the disaster management domain in order to facilitate information interoperability and provide a foundation for knowledge interoperability. To achieve this objective, we designed a pivot ontology framework that consists of two main components: pivot ontology construction and PivotOntology-to-Database schema matching. In general, a pivot ontology is an ontology that has correspondences to other neighbor ontologies [21].

Most existing humanitarian aid information systems have not been represented using ontologies. Rather, these systems are stored in relational databases and are not initially developed for supporting information integration. Our proposed framework employs a pivot ontology as intermediate conceptualization for semantic reconciliation among existing databases. We introduce an ontology engineering methodology for developing a pivot ontology for Humanitarian Aid for Refugees in Emergencies (HARE), which can support interoperation among heterogeneous systems and provide guidelines for PivotOntology-to-Database schema matching.

The paper is organized as follows: Section 2 provides a pivot ontology framework for humanitarian aid information systems. Section 3 describes our HARE ontology construction methodology. Section 4 describes a combining matcher strategy for PivotOntology-to-Database schema matching. Section 5 explains the correspondence analysis between existing humanitarian aid information systems through the HARE ontology. Section 6 concludes the paper.

2. A Pivot Ontology Framework

With the idea of pivot ontology construction, we aim to fulfill the requirements for heterogeneous information in humanitarian aid. As mentioned above, an information system for humanitarian aid in emergencies often cooperates with information in diverse domains. The problem is further complicated when accessing independent databases, across organizations, where full semantic knowledge of the component databases is most likely not available [20]. For example, system *A* provides disaster reports with disaster names, victims, and affected areas. System *B* also provides disaster reports with overlapping details, e.g., disaster names, missing persons, and affected areas. System *C* provides information on facilities, e.g., facility names, types, and their locations. System *D* provides medical information, e.g., infirmary names and physician information. It is necessary for humanitarian actors to collect relevant information from these systems to support their decision-making processes. For this, the schemas of the four systems have to be integrated. Information from autonomous database schemas possibly has similar meanings but appears in structurally different forms in different databases. To cooperate between this information, semantic conflict is often problematic, leading to mismatching integration.

An ontology enables one to reuse and share application domain knowledge using a common vocabulary across heterogeneous applications. It provides a hierarchically structured set of terms and some specification of their meanings to define a structure on the domain and constrain the possible interpretations of terms that can be used [22], [34].

What is required, in fact, is not a combination of all application-dependent knowledge. An important function of a pivot ontology is to support very broad semantic interoperability among domain knowledge. As depicted in **Fig. 1**, the pivot ontology should be generally designed as an application-independent ontology for sharing humanitarian information across independent databases. We focus our attention on five humanitarian aid processes, i.e., refugee registration, identification of persons of concern, emergency planning, distribution of assistance, and donation, which involve several humanitarian actors. The pivot ontology should be constructed based on existing upper ontologies and lexical databases, which together provide general concepts, and domain-specific knowledge from international standards, such as the United Nations High Commissioner for Refugees (UNHCR) handbooks [8], [24], [25], and the Sphere handbook [26]. With these information sources, the pivot ontology can be constructed more rapidly and more reliably.

In order to be able to capture the pivot ontology's core concepts and the upper ontologies' broad concepts, all of these concepts are integrated into a single ontology. Domain knowledge is integrated into the domain level of HARE. The domain ontology will be classified by exploiting the generic concepts in the upper ontologies. Since the upper ontologies represent very broad concepts, the domain ontology cannot be seamlessly generalized by the upper ontologies. WordNet [*1] has been chosen to reconcile the domain ontology and the upper ontologies. The pivot ontology plays an important role in matching with existing systems using a PivotOntology-to-Database schema matching approach.

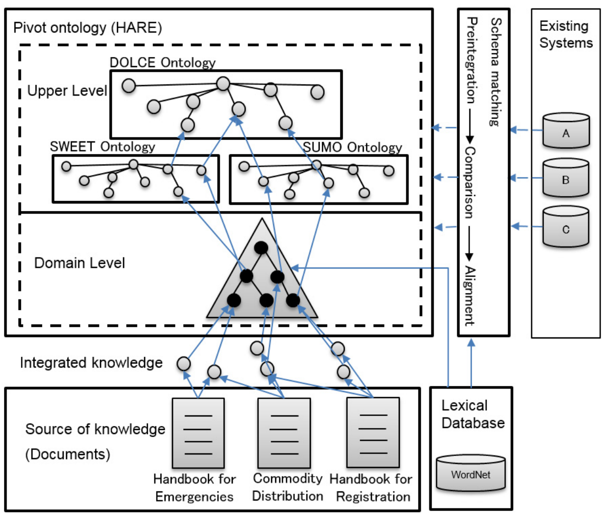

Fig. 1 Pivot Ontology Framework for humanitarian aid domain.

*1 https://wordnet.princeton.edu/

3. A Pivot Ontology Construction Method

Existing approaches to ontology construction include the Uschold and King's method [14], [33], [35], which is intended for enterprise ontology construction, the METHONTOLOGY methodology [9], intended for building life cycles based on evolving prototypes, the SENSUS methodology [31], intended for linking domain terms to a large-scale ontology, and the On-To-Knowledge methodology [30], intended for utilizing ontologies for improving knowledge management quality. The Uschold and King's method and the METHONTOLOGY methodology apply application-independent strategies, and the SENSUS methodology uses an application-semi-dependent strategy. By contrast, the On-To-Knowledge methodology employs an application-dependent strategy. An ontology created using an application-independent strategy is likely to be more reusable, compared to that developed using an application-dependent strategy [11].

To design a common ontology for Humanitarian Aid for Refugees in Emergencies (HARE), an application-independent strategy, techniques for reusing existing ontologies, semantic hierarchical conceptual models, and ontology engineering techniques for solving interoperability problems [13] are applied. We adopt the basic steps from the Uschold and Kings method [33], which consists of the following phases: (i) purpose identification, (ii) ontology capture, and (iii) coding and integrating. In their original forms, these phases do not precisely describe the reuse of existing ontologies and hierarchical conceptual models. To meet our objective, they are extended and tailored for the construction of the HARE ontology as follows [1], [2]:

(1) Identifying purposes and scope
  (a) Getting requirements of refugees in emergencies
  (b) Creating use case diagrams and use case descriptions
(2) Building an ontology
  (a) Ontology capture - considering knowledge models from the use case diagrams
  (b) Ontology coding and integrating
    (i) Integrating with existing upper ontologies
    (ii) Finding hypernyms of each concept to create a hierarchical structure
(3) Evaluation - Verification with FaCT++, UNHCR handbooks, and existing schemas

3.1 Identifying Purposes and Scope

Getting requirements of refugees in emergencies: The HARE ontology integrates domain knowledge from relevant chapters in the Handbook for Emergencies [24] and related documents [8], [25] to undertake the abstraction and processes of refugee emergencies from UNHCR. The operations of UNHCR cover many areas in refugee emergencies, including health, food, sanitation and water, as well as key field activities corroborating the operations such as logistics, community services, and registration. Such operations are managed and controlled by many associate organizations. In this step, information should be extracted carefully from documentation.

Creating use case diagrams and use case descriptions: The domain, scope and purposes of the identified operations are determined, and Unified Modeling Language (UML) use case diagrams are developed for specifying typical user-visible functions of a humanitarian aid information system and for graphically representing and envisioning the relationships between use cases and actors. The flows of interaction between actors and the system in each use case is specified using a textual use case description.

3.2 Ontology Capture

The use case diagrams and use case descriptions obtained from the previous phase provide a source of requirements for establishing ontological conceptualization for developing the HARE ontology. Based on these diagrams and descriptions, concepts and relationships between them are identified and extracted. Resulting core concepts include Commodity, Distribution Cycle, Family, Household, Head of Family, Head of Household, Refugee, Registration Card, RefugeeActivity, RefugeeNeed, Person, Plan, Project, Organization, and Staff. After the core concepts are defined, subclasses and disjoint decompositions are also identified; for example, a food product is identified as a specific type of Commodity.

The implementation of the HARE ontology requires an appropriate ontology editor and development environment. The Prot*égé* development platform, which contains the Prot*égé*-OWL ontology editor for the Semantic Web, is used in this research.

3.3 Coding and Integrating

Integrating with upper ontologies: An upper ontology (a top-level ontology or a foundation ontology) describes general common concepts for many knowledge domains and provides a mechanism for interoperation across domain-specific systems [12], [32]. The concepts obtained from the previous phase are associated with more general concepts in three relevant upper ontologies, i.e., the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [*2], the Semantic Web for Earth and Environmental Terminology (SWEET) [*3], and the Suggested Upper Merged Ontology (SUMO) [*4]. The DOLCE ontology provides clear cognitive artifacts aiming to make explicit the rationale behind ontological modeling decisions. The SWEET ontology provides an extensive vocabulary in the domain of geography and environments. The SUMO ontology merges a number of existing upper ontologies into 11 sections, e.g., the structural ontology containing relations for defining a proper ontology, and the unit-of-measure ontology providing definitions of standard unit systems [23].

An integration problem often arises when HARE concepts are generalized into general concepts in several upper ontologies. To address the integration among upper ontologies, priority levels of upper ontologies should be arranged. The DOLCE ontology is given the highest priority level because of their abstract concepts for cognitive ontological categorization. Both SUMO and SWEET partly comprise domain ontologies; however, SUMO contains more abstract concepts, whereas SWEET contains more specific concepts related to humanitarian aid. SUMO therefore

*2 http://www.loa.istc.cnr.it/old/DOLCE.html
*3 http://sweet.jpl.nasa.gov/
*4 http://www.adampease.org/OP/

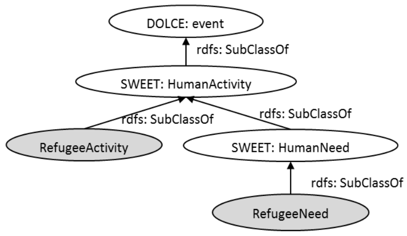

Fig. 2 Integrating with upper ontologies.

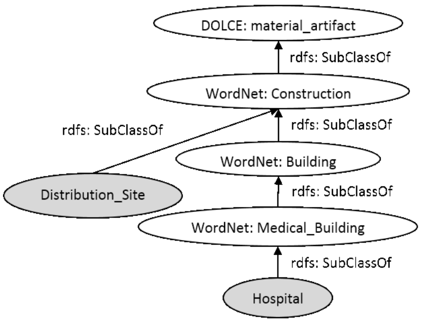

Fig. 3 Integrating with WordNet.

takes priority over SWEET. **Figure 2** exemplifies the integration among DOLCE, SWEET, and HARE, where a grey oval represents a HARE concept.

Finding hypernyms of each concept to create a hierarchical structure: A hypernym is a word or phase whose meaning includes the meanings of other words. A broad meaning of a hypernym constitutes a category into which words with more specific meanings fall. For ease of understanding and interoperation, core HARE concepts are organized into a hierarchy by using word hyponyms from WordNet. WordNet can also be used for bridging the gap between HARE and upper ontologies. For instance, consider the core concepts 'Distribution_site' and 'Hospital.' Each of them is a material artifact, which is a top-level concept from DOLCE. More concrete representations of abstract concepts are required to connect top-level concepts defined in an upper ontology with the core concepts defined in the previous phase. Some WordNet concepts such as 'Medical_Building,' 'Building,' and 'Construction' are more specific than 'DOLCE:material_artifact.' Likewise, the 'Construction' concept is more general than 'Distribution_site,' and the 'Medical_Building' concept is more general than 'Hospital.' **Figure 3** shows a generalization hierarchy of these concepts.

3.4 Pivot Ontology Evaluation

The obtained HARE ontology provides a common conceptualization of humanitarian aid for integrating related systems. **Figure 4** shows examples of top-level concepts in the HARE ontology. In total, the HARE ontology contains 446 elements (268 classes, 105 object properties, and 73 data properties), 90 of which are taken from the upper ontologies and a lexical database, i.e., 38 elements from DOLCE, 4 elements from SWEET, 7 elements from SUMO, and 41 elements from WordNet.

For evaluation purposes, the coverage of the HARE ontology is evaluated based on a compatibility comparison against two existing systems, i.e., Sahana [6], [28], and Ushahidi [36], by using

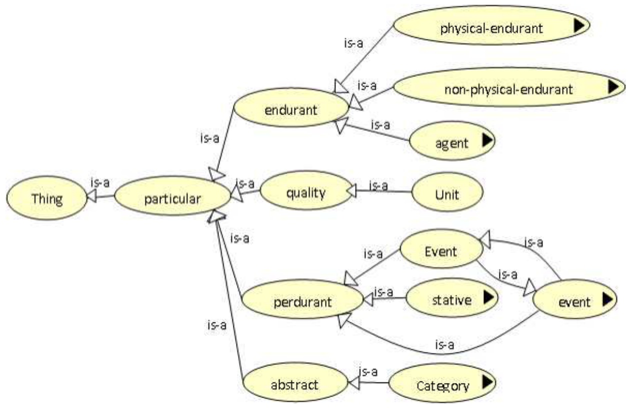

Fig. 4 Top-level concepts in the HARE ontology.

the schema matching technique described in Section 4. The evaluation details will be given in Section 5.

4. Schema Matching

Schema matching is the process of identifying correspondences between elements of two schemas. Schema-based matching and related techniques for database matching have been studied in Refs. [5], [19], [27], [29]. Two main levels, i.e., element and structure levels, were considered in these studies. In the element level, matching elements are computed by analyzing entities, e.g., using linguistic matching, an auxiliary information technique, and constraint-based matching. In the structure level, matching elements are computed by analyzing the structure of entities, e.g., using graph matching, usage-based matching, and document link similarity.

To enhance information sharing for an emergency response, humanitarian information integration among diverse databases is necessary. Reconciliation of the structure and terminology of heterogeneous database schemas is required to solve a database schema integration problem. For database integration, a global schema is useful to eliminate duplication, avoid problems of multiple updates, and minimize inconsistencies across systems [4]. A global schema requires the establishment of explicit semantics and knowledge reuse. In knowledge engineering, the idea of ontology has been introduced to support wider usability of a knowledge base. In this research, the HARE ontology is employed as a global schema and a PivotOntology-to-Database schema matching methodology is designed for the database integration in the humanitarian aid domain. This section explains the PivotOntology-to-Database schema matching model and matching algorithms.

4.1 A PivotOntology-to-Database Matching Model

The HARE ontology contains general terminologies from WordNet and the three aforementioned upper ontologies. A relation between database schema elements and HARE concepts can be found using lexical matching. Our lexical matching method uses WordNet for finding synonyms and hyponyms in order to determine lexical entailment [15] between database schema elements and HARE concepts.

Our PivotOntology-to-Database matching model is depicted in **Fig. 5**. Three ontology-database matching techniques, i.e.,

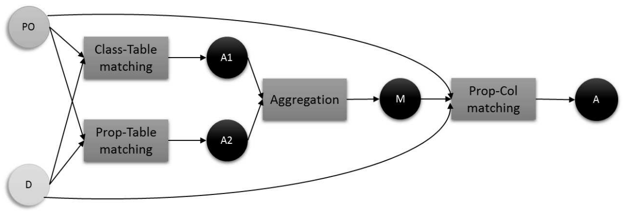

Fig. 5 A PivotOntology-Database matching model.

Class-Table, Property-Table, and Property-Column matching techniques, developed in our previous work [2] are adopted for ontology-to-database matching. The matching model consists of two phases. The first phase consists of two different processes, i.e., Class-Table and Property-Table matching, running independently for finding correspondences between a given pivot ontology, say *PO*, and a given database, say *D*. An alignment is a set of correspondences obtained from each matching process. The resulting alignments, say Alignments $A1$ and $A2$, are aggregated into a combined alignment, say M. The second matching phase takes M, *PO*, and D as input data for determining the final alignment, say A, using Property-Column matching.

4.2 Matching Algorithms

Based on the correspondences between their elements (Class-Table, Property-Column, and Property-Table correspondences [2]), a given relational database D is matched against the pivot ontology with the assistance of a domain expert using the following algorithms:

- Algorithm 1: The main structure for calling other algorithms and returning the output.
- Algorithm 2: Used for checking correspondences of junction tables against HARE concepts.
- Algorithm 3: Used for checking correspondences between HARE classes and database tables.
- Algorithm 4: Used for checking correspondences between HARE properties and table columns.

Element-level matching techniques, i.e., string-based matching and matching using linguistic resources, are applied in Algorithms 2, 3, and 4. Structure-level matching techniques based on internal structure, domains, ranges, foreign keys, and property types are applied in Algorithms 2 and 4. Resulting alignments from Algorithms 2 and 3 are aggregated. Columns and their corresponding properties in the aggregation result are then checked by using Algorithm 4.

Data: C is a concept in a given pivot ontology and T is a table in a given database.
Result: Alignment A between Concept C and Table T
for *(i=0;C_i exists;i++)* **do**
  for *(j=0;T_j exists;j++)* **do**
    if *(T_j is a junction table)* **then**
      if *(PropTableMatching(Prop(C_i),T_j))* **then**
        ADD($A2$, corr(*Prop*(C), T_j));
      end
    else
      if *(ClassTableMatching(C_i,T_j))* **then**
        ADD($A1$, corr(C_i, T_j);
      end
    end
  end
  M = Merge($A1$, $A2$);
end
for *(i=0;M_i exist;i++)* **do**
  A = PropColMatching(M_i(C),M_i(T));
end
Return A;

Algorithm 1: PivotOntology-Database matching.

Data: T is a junction table; $T0$ and $T1$ are referenced tables of ForeignKey(T); C is Domain(Prop(C)); $C1$ is Range(Prop(C)).
if *(ClassTableMatching(C, $T0$))*
AND (ClassTableMatching($C1$, $T1$)) **then**
  Return true;
else
  Return false;
end

Algorithm 2: Property-Table matching.

if *Synonym(T,C)* **then**
  Return true;
else if *Hyponym(Synset(T),Synset(C))* **then**
  Return true;
else
  Return false;
end

Algorithm 3: Class-Table matching.

for *(i=0; $Prop_i$(C) exist; i++)* **do**
  for *(j=0; Col_j(T) exist; j++)* **do**
    if *Datatype($Prop_i$(C)), Col_j(T))* **then**
      if *Synonym($Prop_i$(C)), Col_j(T))* **then**
        ADD(A, corr($Prop_i$(C), Col_j(T));
      else if *Hypernym($Prop_i$(C)), Col_j(T))* **then**
        ADD(A, corr($Prop_i$(C), Col_j(T));
      end
    end
  end
end
Return A;

Algorithm 4: Property-Column matching.

5. Evaluation

The main goal of the evaluation is threefold: (i) to investigate the PivotOntology-to-Database matching compared to direct matching without using the HARE ontology, (ii) to examine the compatibility of the HARE ontology against existing systems, and (iii) to explore a case study on database integration via the HARE ontology. In addition, semi-automatic matching is also investigated so as to reduce matching time for dealing with large-scale database schemas. Strategies for semi-automatic matching and combinations thereof are evaluated.

The two open-source disaster management systems, i.e., Sahana [6], [28] and Ushahidi [36], are used in our case study. Sahana was developed by members of the Sri Lankan IT community, including experts in emergency and disaster management, and dedicated to helping people by providing information management solutions. Ushahidi was created by a development team from different countries, and dedicated to gathering crisis information. The two systems have different schemas and provide different features in the humanitarian aid domain. Sahana contains modules such as organization registry, Project Tracking, Messag-

Table 1 Manual matching experiments.

Group	Scenario	Description
1	$\mathcal{S} \leftrightarrow \mathcal{U}$	Matching between Sahana and Ushahidi without HARE
2	Test Case 1: $\mathcal{S} \leftrightarrow \mathcal{H}$	Matching between Sahana and HARE
	Test Case 2: $\mathcal{U} \leftrightarrow \mathcal{H}$	Matching between Ushahidi and HARE
	Test Case 3: $\mathcal{S} \leftrightarrow \mathcal{H} \leftrightarrow \mathcal{U}$	Analyzing common correspondences obtained from Test Cases 1 and 2

ing, Scenarios Events, Human Resources, Inventory, Assets, Assessment, and Map. Ushahidi provides relatively fewer features. By collecting information via text messages, emails, twitter and web-forms, it tracks reports on maps, filters data by time, and observes occurring time and locations of events. The schema of Sahana is far larger than that of Ushahidi, i.e., Sahana contains 3,296 elements, while Ushahidi contains 388 elements.

5.1 Manual Matching

The goal of the evaluation is to show how the HARE ontology works by comparing two values from two groups of experiments, i.e., the number of correspondences in the integration between two database schemas through the HARE ontology, and the number of those in the database integration without the HARE ontology. Ideally, the results of both groups should be close or equal. The results are shown in Table 4 (Section 5.1.3). The two groups of experiments are shown in **Table 1**, i.e.:

- Matching between Sahana and Ushahidi without using the HARE ontology, denoted by $\mathcal{S} \leftrightarrow \mathcal{U}$: The number of correspondences is regarded as an expected matching result.
- Matching between Sahana and Ushahidi through the HARE ontology: This group consists of 3 test cases:
  - Test Case 1: Matching between Sahana and HARE, denoted by $\mathcal{S} \leftrightarrow \mathcal{H}$.
  - Test Case 2: Matching between Ushahidi and HARE, denoted by $\mathcal{U} \leftrightarrow \mathcal{H}$.
  - Test Case 3: Analyzing the matching results of Test Cases 1 and 2 for finding common matching.

5.1.1 Test Case 1: Matching between Sahana and HARE

The event concept is a core concept in disaster management. The HARE ontology is designed to facilitate interoperability among existing humanitarian aid databases by providing broad humanitarian aid vocabularies and their relationships. An event information is described by the HARE ontology. Each event uses an asset. An event has related activities in a particular time period. The activity concept is a subclass of the event concept. An activity contains the information to identify its location. Each of an event, an activity, and a location has a name as a data type property. A location is identified by its latitude, longitude, and address. In the portion of the Sahana database illustrated in **Fig. 6**, an event has many activities at a location in a particular period. Assets used by an event are kept in a location. An activity belongs to a project. Events, activities, and locations have names as their attributes. A location is identified by a latitude, a longitude, a street, and a postcode. The alignment resulting from matching Sahana with HARE consists of 107 correspondences, some of which are shown in **Table 2**. According to this table,

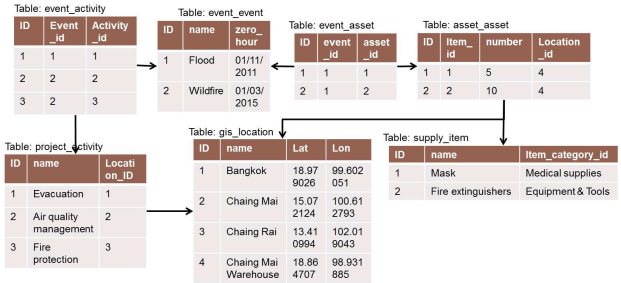

Fig. 6 Example of the Sahana schema.

Table 2 Examples of correspondences between Sahana and HARE.

Correspondence pairs	
Sahana	**HARE**
event_event.name	Event.has-event-name
event_activity	Event.has-activity.Activity
project_activity.name	Activity.has-activity-name
project_activity.location_id	Activity.has-location.Location
gis_location.name	Location.has-location-name

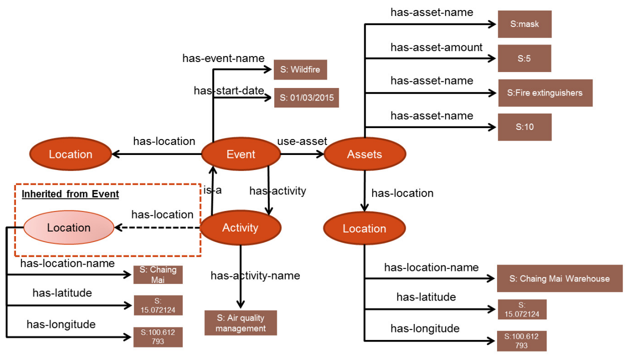

Fig. 7 Sahana and HARE matching.

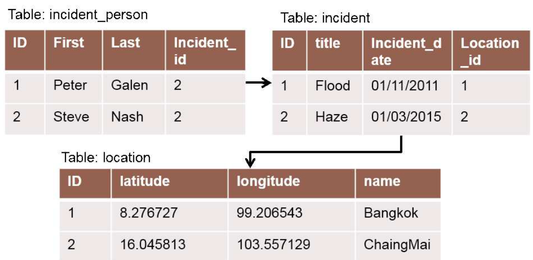

Fig. 8 Example of Ushahidi schema.

information in Sahana can be shared with other systems through three HARE concepts, i.e., an event, an activity, and a location. **Figure 7** depicts information and schema integration between Sahana and HARE. A red oval represents a concept in the HARE ontology. A label on a line represents a property of a concept. A brown rectangle denotes information embedded in HARE after matching. Properties of the event concept are inherited through the 'is-a' relation to the activity concept. As a result, an activity also has the 'has-location' property.

5.1.2 Test Case 2: Matching between Ushahidi and HARE

In the Ushahidi database illustrated in **Fig. 8**, an incident person will be recorded to identify the incident in which he/she participates. An incident contains the information to identify its location. A location is identified by its latitude and longitude.

We also consider the possibility of generalizing elements for interoperability. Linguistic relations are employed for database

Table 3 Examples of correspondences between Ushahidi and HARE.

Correspondence pairs		Relation
Ushahidi	HARE	
incident	Event	Hyponym
incident_person	Person	Hyponym

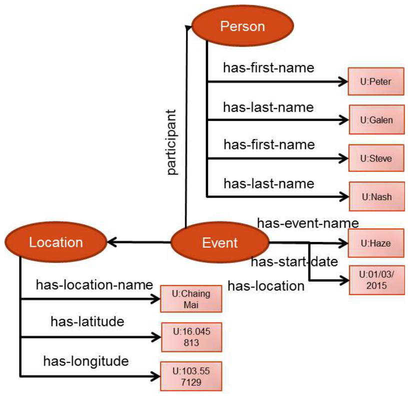

Fig. 9 Ushahidi and HARE matching.

Table 4 The experimental results for manual matching.

Group	Scenario	#Correspondences
1	$\mathcal{S} \leftrightarrow \mathcal{U}$	13
2	Test Case 1: $\mathcal{S} \leftrightarrow \mathcal{H}$	107
	Test Case 2: $\mathcal{U} \leftrightarrow \mathcal{H}$	24
	Test Case 3: $\mathcal{S} \leftrightarrow \mathcal{H} \leftrightarrow \mathcal{U}$	13

matching. A linguistic resource, i.e., WordNet, is used for finding linguistic relations, e.g., synonyms, hypernyms, hyponyms, and equality. The alignment result consists of 24 correspondences. **Table 3** shows examples of some correspondences and their relations. For instance, the concept incident is more specific than the concept event. An implication is that information about an incident can be represented by an event, but some information concerning an event may be not represented by an incident. **Figure 9** depicts information and schema integration between Ushahidi and HARE.

5.1.3 Test Case 3: Analyzing Matching Results from Test Cases 1 and 2 for Finding Common Correspondences

Table 4 shows the results of manual matching. The number of correspondences in Test Case 3, i.e., $\mathcal{S} \leftrightarrow \mathcal{H} \leftrightarrow \mathcal{U}$, is equal to that of direct matching $\mathcal{S} \leftrightarrow \mathcal{U}$ without using HARE. However, the use of HARE greatly extends the possibility of integration and fusion of information in Sahana and Ushahidi (Fig. 12). The 13 common correspondences obtained from Test Case 3 provide a bridge connecting the correspondences in Test Case 1 and those in Test Case 2, i.e., correspondences in Test Cases 1 and 2 can be joined using these common HARE-based correspondences. Information from Sahana and Ushahidi flows through such join operations. Without using HARE, the possibility of joining Sahana elements to Ushahidi elements is limited.

5.1.4 Compatibility of HARE

Figure 10 depicts the percentage of compatibility of Sahana against Ushahidi and HARE. For information diversity, the HARE ontology has overcome Sahana. The labels above the axes of Fig. 10 show twelve categorized terms and the total number of Sahana elements in each categorized term. The categorized terms are asset, shelter, document, event, location, hospital, human resource, organisation, person, project, request, and supply.

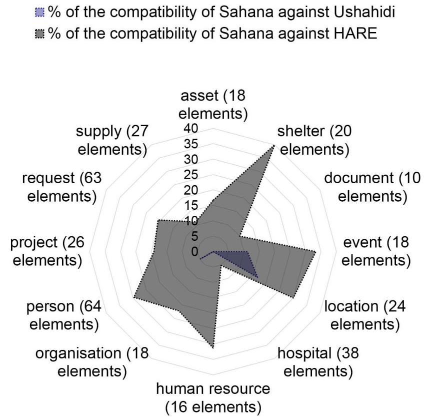

Fig. 10 Compatibility percentage of Sahana against Ushahidi and HARE.

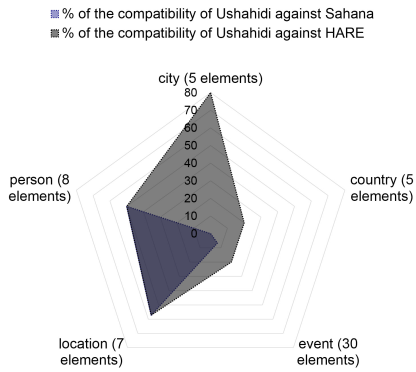

Fig. 11 Compatibility percentage of Ushahidi against Sahana and HARE.

The black area represents the compatibility percentage of Sahana against HARE, while the purple area points out the compatibility percentage of Sahana against Ushahidi. The purple area is apparently smaller than the black area, i.e., the compatibility of Sahana against HARE is higher than the compatibility of Sahana against Ushahidi. Usability of Sahana increases as the compatibility of Sahana against HARE expands. The higher compatibility of a pivot ontology and anonymous databases can be achieved by ontology modification, which is a process in ontology engineering. An ontology is flexible and changeable under control of the main structure of the pivot ontology.

Figure 11 depicts the compatibility percentage of Ushahidi against Sahana and HARE. For information diversity, the HARE ontology has overcome Ushahidi. There are five categorized terms, e.g., city, country, event, location, and person. The black area points out the compatibility percentage of Ushahidi against HARE. The purple area depicts the compatibility percentage of Ushahidi against Sahana. The black area includes the purple area, i.e., the compatibility of Ushahidi against HARE is higher than that of Ushahidi against Sahana. Usability of Ushahidi also increases as the compatibility of Ushahidi against HARE expands.

The two schemas share some common information, i.e., persons, locations, and events. **Figure 12** represents the information integrated through the HARE ontology. There are three zones in the figure, i.e., Sahana, Ushahidi, and their common zone. The 'event' concept provides a bridge connecting Sahana information, e.g., 'Wildfire,' and Ushahidi information, e.g., 'Haze.' This connection is derived from the HARE concept 'location,' which has corresponding concepts with the same instance in Sa-

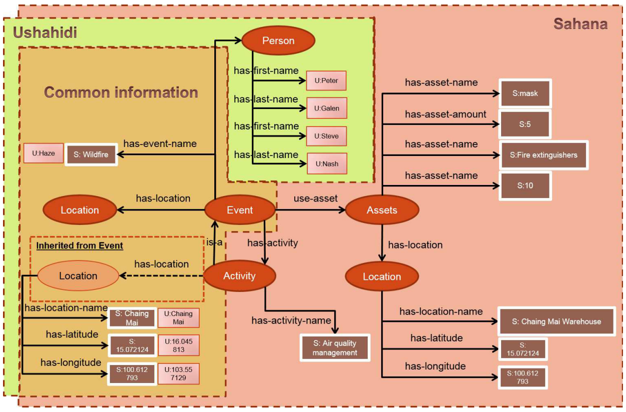

Fig. 12 Information sharing between Sahana and Ushahidi through HARE.

Matching approaches	Strategies					
	Name	Name and Synonym	Path	Leave	Parent	Datatype
Case A: Simple	Trigram	-	Trigram	Trigram	Trigram	-
Case B: Datatype	Trigram	-	Trigram	Trigram	Trigram	Datatype similarity
Case C: Synonym	Trigram	Trigram	Trigram	Trigram	Trigram	-
Case D: Synonym + Datatype	Trigram	Trigram	Trigram	Trigram	Trigram	Datatype similarity

Fig. 13 Semi-automatic matcher composition.

hana and Ushahidi, i.e., S:Chiang_Mai and U:Chiang_Mai. Moreover, the HARE ontology indicates that not only the 'location' concept but also the 'person,' 'activity,' and 'asset' concepts have relations to the 'event' concept. For example, the 'event' concept has a 'has-activity' relation to the 'activity' concept, which contains instances of Sahana, and also has a 'participant' relation to the 'person' concept, which contains instances of Ushahidi. Through these relations, instances of the two databases can be connected by using corresponding concepts, e.g., the activity instance 'S:Air_quality_management,' which is related to the event 'S:Wildfire' or 'U:Haze,' can be connected with the person instances 'U:Peter' and 'U:Steve,' which participate in that event.

5.2 Semi-automatic Matching

The evaluation described above is based on manual matching. Next, we exploit semi-automatic matching in the two groups previously used for manual matching (cf. Table 1 in Section 5.1) and investigate appropriate matching strategies. We measure the precision, recall, F-measure, and the number of correspondences to determine the effectiveness of the HARE ontology and combinations of matching strategies for database integration through HARE. We construct four experimental cases, i.e., Cases A-D, shown in **Fig. 13**, for combination of matching strategies, using COMA++ [3] as a matching tool. Trigrams break a word up into a set of three-letter sequences and computed trigram similarity with another set. A matching result is a similarity value between two elements. By experiments, we use 0.4 as a low-threshold similarity value for finding correspondences. The results of both experiment groups are expected to be close or equal. The PivotOntology-to-Database matching is computed in both element and structure levels, e.g., lexical matching and data type matching. Thus, an assumption of evaluation is that Case D, i.e., combination of synonym and data type strategies, can be an appropriate matching strategy for both groups of experiments.

Table 5 Semi-automatic matching results.

Group	Scenario	% of manual matching			
		Case A	Case B	Case C	Case D
1	$\mathcal{S} \leftrightarrow \mathcal{U}$ (ER_m=13)	23.08%	23.08%	7.69%	38.46%
2	Test Case 1: $\mathcal{S} \leftrightarrow \mathcal{H}$ (ER_m=107)	43%	27.1%	58.88%	66.36%
	Test Case 2: $\mathcal{U} \leftrightarrow \mathcal{H}$ (ER_m=24)	20.83%	16.67%	16.67%	29.17%
	Test Case 3: $\mathcal{S} \leftrightarrow \mathcal{H} \leftrightarrow \mathcal{U}$ (ER_m=13)	15.38%	7.69%	15.38%	38.46%

Table 5 shows the results of semi-automatic matching compared with the results of manual matching (Table 4), where ER_m is the number of correspondences obtained from manual matching. The most effective matching strategy composition is Case D (Combination of synonym and data type strategies). The aim of this empirical study is to investigate appropriate semi-automatic matching strategies. Improvement of the performance of semi-automatic matching techniques is beyond the scope of this research. However, according to Table 5, the results of semi-automatic matching are in line with those of manual matching. In particular, using the strategy combination of Case D, the result of Test Case 3 ($\mathcal{S} \leftrightarrow \mathcal{H} \leftrightarrow \mathcal{U}$), i.e., 38.46%, is the same as that of Group 1 ($\mathcal{S} \leftrightarrow \mathcal{U}$).

6. Conclusions

With a vast increase of humanitarian aid information, various systems are individually developed by different organizations. Collaboration support between individual systems is essential and challenging. In this paper, we have designed a pivot ontology framework. The HARE ontology is proposed based on our pivot ontology construction methodology, in which several ontology engineering techniques are applied. The core concepts and their relationships are extracted from standard handbooks. Upper ontologies are used in the construction for generalization of a pivot ontology. PivotOntology-to-Database schema matching has been introduced as a guideline for cross-system integration based on a pivot ontology. Linguistic relations are applied for finding alignments between an ontology and database schemas.

Our experiments show that two existing humanitarian aid information systems, i.e., Sahana and Ushahidi, can be extensively integrated by using the HARE ontology, with the mainstream of their schemas being well covered. Our case study not only demonstrates PivotOntology-to-Database schema matching, but also typically presents the integrated information extended by using the HARE ontology. In addition, several semi-automatic matching strategies are investigated.

Future work includes an extension of the scope of the HARE ontology to cover broader humanitarian aid knowledge. We plan to implement ontology-based emergency response services for decision-making support concerning providers (donors), recipients (affected populations), and implementers (e.g., governments, foundations, Red Cross, NGOs, and UN agencies). Such services will address issues related to how to facilitate the transmission and use of information seamlessly across humanitarian aid services developed by different providers, recipients, and implementers during all major phases of disaster management.

References

[1] Apisakmontri, P., Nantajeewarawat, E., Buranarach, M. and Ikeda, M.: Towards the Use of Upper Ontologies for Refugee Emergencies in Disaster Management, *the Second Asian Conference on Information Systems* (2013).

[2] Apisakmontri, P., Nantajeewarawat, E., Buranarach, M. and Ikeda, M.: Ontology Construction and Schema Merging Using an Application-Independent Ontology for Humanitarian Aid in Disaster Management, *Semantic Technology*, Lecture Notes in Computer Science, Vol.8943, pp.281–296, Springer International Publishing (2015).

[3] Aumueller, D., Do, H.-H., Massmann, S. and Rahm, E.: Schema and Ontology Matching with COMA++, *Proc. 2005 ACM SIGMOD International Conference on Management of Data, SIGMOD '05*, pp.906–908, New York, NY, USA, ACM (2005).

[4] Batini, C., Lenzerini, M. and Navathe, S.B.: A Comparative Analysis of Methodologies for Database Schema Integration, *ACM Comput. Surv.*, Vol.18, No.4, pp.323–364 (1986).

[5] Bernstein, P.A., Madhavan, J. and Rahm, E.: Generic schema matching, ten years later, *Proc. VLDB Endowment*, Vol.4, No.11, pp.695–701 (2011).

[6] Careem, M. et al.: Sahana: Overview of a Disaster Management System, *International Conference on Information and Automation, 2006, ICIA 2006*, pp.361–366 (2006).

[7] Compton, M. et al.: The SSN ontology of the W3C semantic sensor network incubator group, *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol.17, pp.25–32 (2012).

[8] Division, O.S.: *Commodity Distribution*, United Nations High Commissioner for Refugees (1997).

[9] Fernández-López, M., Gómez-Pérez, A. and Juristo, N.: Methontology: From ontological art towards ontological engineering (1997).

[10] Foster, I. and Grossman, R.L.: Data Integration in a Bandwidth-rich World, *Comm. ACM*, Vol.46, No.11, pp.50–57 (2003).

[11] Gábor, N.: Ontology Development, *Semantic Web Services*, Studer, R., Grimm, S. and Abecker, A. (Eds.), pp.107–134, Springer Berlin Heidelberg (2007).

[12] Gómez-Pérez, A. and Fernández-López, M.: *Ontological engineering: With examples from the areas of knowledge management, e-commerce and the semantic web*, Springer (2010).

[13] Gómez-Pérez, A. et al.: *Ontological Engineering: With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web* (*Advanced Information and Knowledge Processing*), Springer-Verlag New York, Inc., Secaucus, NJ, USA (2004).

[14] Grüninger, M. and Fox, M.S.: Methodology for the Design and Evaluation of Ontologies (1995).

[15] Herrera, J. et al.: Textual Entailment Recognition Based on Dependency Analysis and WordNet, *Machine Learning Challenges, Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, Vol.3944, pp.231–239 (2006).

[16] Hui, W. et al.: An owl-based knowledge representation framework for searching semantically emergency preplan, *ISECS International Colloquium on Computing, Communication, Control, and Management, 2009, CCCM 2009*, Vol.3, pp.128–131 (2009).

[17] IASC: Multi-Cluster/Sector Initial Rapid Assessment (MIRA), available from ⟨https://docs.unocha.org/sites/dms/Documents/mira_final_version2012.pdf⟩ (accessed 2014-02-24).

[18] Klien, E., Lutz, M. and Kuhn, W.: Ontology-based discovery of geographic information services-An application in disaster management, *Computers, Environment and Urban Systems*, Vol.30, No.1, pp.102–123 (2006).

[19] Madhavan, J. et al.: Generic Schema Matching with Cupid, *Proc. 27th International Conference on Very Large Data Bases, VLDB '01*, pp.49–58, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (2001).

[20] Naiman, C.F. and Ouksel, A.M.: A Classification of Semantic Conflicts in Heterogeneous Database Systems, *J. Organ. Comput.*, Vol.5, No.2, pp.167–193 (1995).

[21] Navas-Delgado, I. et al.: Semantic Fields: Finding Ontology Relationships, *Database and Expert Systems Applications*, Lecture Notes in Computer Science, Vol.5690, Springer Berlin Heidelberg, pp.427–434 (2009).

[22] Patil, R., Knight, K. and Russ, T.: Towards distributed use of large-scale ontologies, *AAAI Spring Symposium on Ontological Engineering* (1997).

[23] Pease, A., Niles, I. and Li, J.: The suggested upper merged ontology: A large ontology for the semantic web and its applications, *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*, Vol.28 (2002).

[24] Preparedness, T.E. and Section, R.: *Handbook for Emergencies*, United Nations High Commissioner for Refugees, UNHCR Headquarters, Case Postale 2500, Switzerland, 3 edition (2007).

[25] Profile, P., the Population and Section, G.D.: *UNHCR Handbook for Registration*, United Nations High Commissioner for Refugees (2003).

[26] Project, S.: *The Sphere Handbook, Humanitarian Charter and Minimum Standards in Humanitarian Response*, Practical Action Publishing (2011).

[27] Rahm, E. and Bernstein, P.A.: A survey of approaches to automatic schema matching, *The VLDB Journal*, Vol.10, No.4, pp.334–350 (2001).

[28] Sahana: Sahana Software Foundation, available from ⟨http://sahanafoundation.org/⟩.

[29] Shvaiko, P. and Euzenat, J.: A Survey of Schema-Based Matching Approaches, *Journal on Data Semantics IV*, Spaccapietra, S. (Ed.), Lecture Notes in Computer Science, Vol.3730, pp.146–171, Springer Berlin Heidelberg (2005).

[30] Staab, S., Studer, R., Schnurr, H.-P. and Sure, Y.: Knowledge processes and ontologies, *IEEE Intelligent systems*, Vol.16, No.1, pp.26–34 (2001).

[31] Swartout, B., Patil, R., Knight, K. and Russ, T.: Toward distributed use of large-scale ontologies, *Proc. 10th Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*, pp.138–148 (1996).

[32] Tripathi, A. and Babaie, H.A.: Developing a modular hydrogeology ontology by extending the {SWEET} upper-level ontologies, *Computers & Geosciences*, Vol.34, No.9, pp.1022–1033 (2008).

[33] Uschold, M. and Gruninger, M.: Ontologies: Principles, methods and applications, *The Knowledge Engineering Review*, Vol.11, No.02, pp.93–136 (1996).

[34] Uschold, M. and Jasper, R.: A Framework for Understanding and Classifying Ontology Applications, *IJCAI'99 Workshop on Ontology and Problem Solving Methods: Lessons Learned and Future Trends* (1999).

[35] Uschold, M. and King, M.: Towards a Methodology for Building Ontologies, *In Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95* (1995).

[36] Ushahidi: Ushahidi, available from ⟨http://www.ushahidi.com/⟩.

[37] Xu, W. and Zlatanova, S.: Ontologies for Disaster Management Response, *Geomatics Solutions for Disaster Management*, Li, J., Zlatanova, S. and Fabbri, A. (Eds.), Lecture Notes in Geoinformation and Cartography, pp.185–200, Springer Berlin Heidelberg (2007).

[38] Zhang, D., Zhou, L. and Nunamaker Jr, J.F.: A Knowledge Management Framework for the Support of Decision Making in Humanitarian Assistance/Disaster Relief, *Knowledge and Information Systems*, Vol.4, No.3, pp.370–385 (2002).

Pasinee Apisakmontri is a Ph.D. student within Doctoral Dual Degree Program (Sirindhorn International Institute of Technology (SIIT) and Japan Advanced Institute of Science and Technology (JAIST)). She received her B.Sc. degree in Computer Science (First Class Honors) from Prince of Songkla University in 2007. She received her M.S. degree in Computer Science from Asian Institute of Technology in 2012. Her research interests are Knowledge Management, Semantic Technology, Ontology, and Disaster management.

Ekawit Nantajeewarawat received his D.Eng in Computer Science from the Asian Institute of Technology in 1997. His research interests include knowledge representation, information extraction, automated reasoning, and formal ontology languages.

Mitsuru Ikeda received his B.E. and M.E. from Utsunomiya University in 1984 and 1986, respectively. He received his D.E. from Osaka University in 1989. He is currently a professor in Japan Advanced Institute of Science and Technology. His research interests include service science, ontological engineering research in the field of e-learning, knowledge management, semantic web. He has received best paper award of Japanese Society for Artificial Intelligence and outstanding paper award of International Conference on Computers in Education in 1996 and 1999, respectively.

Marut Buranarach received his B.Eng. degree in Engineering from King Mongkuts Institute of Technology Lardkrabang in 1995. He received his M.S. and Ph.D. degrees in Information Science from the University of Pittsburgh in 1998 and 2004, respectively. He has been with the Language and Semantic Technology Lab at NECTEC in Thailand since 2005.