Advanced Engineering Informatics 30 (2016) 467–480

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Automated experiential engineering knowledge acquisition through Q&A contextualization and transformation

Bo Song [a,b], Zuhua Jiang [a,*], Lijun Liu [a]

[a] Department of Industrial Engineering & Management, Shanghai Jiao Tong University, Shanghai, China
[b] China Institute of FTZ Supply Chain, Shanghai Maritime University, Shanghai, China

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 6 June 2016
Accepted 10 June 2016

Keywords:
Experiential knowledge
Knowledge acquisition
Question and answer
Context-aware
Tacit knowledge

ABSTRACT

Experiential knowledge (EK) in the brain of proficient engineers is an important asset for manufacturing enterprises. As a kind of tacit knowledge, EK is hard to describe clearly and often requires a lot of human efforts to be acquired in a computer-operable form. In this paper we propose a context-aware mechanism to acquire EK in an automatic and timely manner. The proposal comprises a formal description of EK using ontology and default logic, a machine learning-based method that discovers Q&A from the context of collaborative engineering tasks, and a semantic mapping step transforming the discovered Q&A into ontological concepts and relations. An application case shows that the EK of a group of engineers collaborating over a finite element analysis task can be automatically captured from their desktop information flow. The effectiveness of the proposed method with respect to other knowledge acquisition approaches is demonstrated through quantitative and qualitative comparison.

1. Introduction

In recent years, with the increasing mobility of employees, the loss of skilled workers has become a serious problem for many companies. The phenomenon has led to the need of acquiring skilled workers' experiential knowledge (EK) in computer-operable forms, via which the knowledge can be stored and reused easily after the person leaves the company. However, EK cannot be easily acquired because it resides in human brains and is difficult to be described formally. In the previous research, EK was often acquired through interviewing proficients and then manually tidying the narratives of the interviewees [1–3]. This approach does not only require the intensive intervention of knowledge engineers but also makes the acquired knowledge uncontextualized – the interview is not situated in real time task processing and thus has the risk of rendering the acquired knowledge isolated from necessary application details.

In this paper, we use context awareness to solve the problem of real-time EK acquisition. The real-time acquisition of EK also implies the automated nature of the acquisition method, since manual processing of the large volume of contextual information is not realistic. With computer-based analysis of contextual information of a knowledge worker, the developed EK acquisition tool can tell whether any EK is showing up in the current working environment as well as what problems the emerging EK is meant to solve. So far context awareness has been adopted by several knowledge management systems to realize active knowledge reuse [4–6], but little research has explored context awareness for knowledge acquisition. This study makes use of a special kind of contextual information – Q&A, to realize EK acquisition. Being the most common way of knowledge communication, Q&A signifies both the communicated knowledge and the usage condition of the knowledge, thus becomes a suitable carrier of personal experience. To find out the Q&A containing EK, we employ machine learning techniques to classify sentences into useful Q&A elements and useless information. Then the captured EK is changed into a computer-operable form by mapping the identified Q&A to the semantics defined in a domain ontology called the EK ontology. The semi-automatic construction of the EK ontology will also be discussed in this paper.

The paper is organized as follows. Section 2 introduces some recent studies about experiential knowledge and its acquisition. Section 3 defines the concept of EK in this paper and provides EK's formal representation. Section 4 proposes our context-aware EK acquisition method. Using a real-world application case, Section 5 shows how the proposed method is used. Section 6 concludes the paper and points out some future research directions.

* Corresponding author.
E-mail address: zhjiang@sjtu.edu.cn (Z. Jiang).

http://dx.doi.org/10.1016/j.aei.2016.06.002
1474-0346/

2. Related research

2.1. Experience and tacit knowledge

In the knowledge management literature, experience is often mentioned with tacit knowledge [7–10]. Tacit knowledge is a kind of knowledge residing in human brains and being difficult to tell, imitate and disseminate [11]. Experience is often regarded as the source of tacit knowledge or a part of tacit knowledge. For example, Noh et al. think that tacit knowledge is related with people's intuition, insight, faith, and skills, and people's problem-solving experience is stored in their memories as tacit knowledge [8]. Brockmann and Anthony define tacit knowledge as the practical know-how originating from a person's experience of achieving a goal at work [9]. Armaghan and Renaud believe experience is the knowledge people gain after solving a problem [12]. Azadeh et al. define experience as the practical knowledge for dealing with complicated situations where nonlinear, time-varying and fuzzy characteristics are hard to be described by rigorous mathematics [13]. D'Eredita and Barreto investigate the formation of experience and reach the conclusion that experience is a series of associated scenarios encompassing the goal an individual tries to reach, the stimuli the individual receives when achieving the goal, the explanation to the stimuli the individual makes, and how the individual react to the stimuli [14]. Foguem et al. propose that the representation of experience should include the dimensions of *context*, *event*, *analysis* and *solution* [15]. Based on the summarization of previous research, Chen proposes three basic elements of EK, namely *problem*, *cause*, and *solution*. Chen also outlines eight features of EK, deeming EK as being tacit, hierarchical, descriptive, causal, procedural, associative, action-oriented and skill-oriented [16]. Gavrilova and Andreeva think that the EK of employees can be transformed into explicit knowledge through using metaphors, analogies and models in discussions [17]. Some research on case-based reasoning (CBR) also mentions experience, treating cases as the encapsulation of experience [18,19]. Ruiz et al. propose an experience management framework called Experience Feedback (EF) to draw lessons from the positive and negative events in an enterprise's database [20]. Ruiz et al. also distinguish EF from CBR by pointing out that the former is a tacit knowledge externalization mechanism while the latter does not concern tacit knowledge externalization.

From the above literature one can see that experiential knowledge has aroused the interest of many researchers and has been studied from its definition, representation to acquisition methods. However, inheriting the tacit nature from tacit knowledge, experiential knowledge is hard to describe clearly, and currently there is no authoritative definition of EK. While common understanding of experiential knowledge exists (it is *personal*, *contextualized and problem-solving oriented* knowledge), currently there is a gap between the common understanding and the EK some studies assert to get. To better explain this gap, we provide a survey of relevant research in the following subsection.

2.2. Knowledge acquisition

Knowledge acquisition is to acquire knowledge from experts or other knowledge sources and express the acquired knowledge in a computer-operable form. Table 1 lists the recent studies focusing on acquisition of "non-explicit" knowledge. The listed works are analyzed from four perspectives – knowledge source, knowledge type, knowledge representation and knowledge acquisition method.

2.2.1. Knowledge source

Knowledge source refers to the data or data-generating mechanism/subject used as the input to a knowledge acquisition method. Typical knowledge sources include database [13,21–23], interview [2,3,24], conversation [25,26], expert [27], MIS (management information system) [20,28], operation [29] and domain text [30]. When using database as the knowledge source, it usually means multiple cases represented as attribute–value pairs in a database are generalized to produce conclusions. The attribute–value representation of cases supports a variety of advanced data mining techniques, but it can only be used when the target problem can be characterized by a few known parameters. Interview and conversation both produce narratives for knowledge acquisition. Composed of natural language, narratives are much more complicated and expressive than attribute–value pairs. The difference between interview and conversation is that the former aims at obtaining knowledge from a single interviewee, while the later involves multiple knowledge contributors expressing opinions alternatively and guidelessly. Experts are an important source of knowledge. Although there is only one paper listed in the "expert" column of Table 1, the interview and conversation are often conducted on/by experts for knowledge acquisition. In [27], experts directly participated in knowledge acquisition as they are asked to assign the cause-effect relation to the decision variables. Featuring explicit business model, data structure and workflow, a management information system can serve as an effective knowledge source due to the clean, structured and categorized data accumulated in the system [20,28]. The operations of experts solving a class of problems can be used to mine operational knowledge. The operational knowledge source consists of a series of sequentially arranged actions to achieve a certain target [29]. Domain text such as failure report can be used as knowledge source [30]. It can be viewed as a mixed case representation containing both attribute values and text.

2.2.2. Knowledge type

Knowledge type refers to how researchers call the knowledge they have acquired. In Table 1, the surveyed literature focus on acquiring knowledge that is experiential, tacit, personal, operational and uncertain. These types of knowledge are different from the "explicit knowledge" which shows itself as textbooks, formulae or computer programs. The meaning of experiential knowledge and tacit knowledge has been discussed in Section 2.1. The personal knowledge literally means all the knowledge that a person owns, but when mentioned with knowledge acquisition, it mainly means the experience and insights possessed by a certain individual and cannot be found in publicly available sources. Operational knowledge is about how to get things done, usually without telling the reason to do so. Uncertain knowledge is the knowledge containing imprecise and incomplete aspects due to the limitation or error of sensors.

2.2.3. Knowledge representation

Knowledge representation determines what kinds of relations are modeled in the knowledge and how these relations connect the elements of the knowledge. Knowledge representation is important as it affects knowledge's expressiveness, readability and machine operability. The if-then rules are the most widely adopted knowledge representation method because it makes a balance between the machine operability and the human readability [13,20–22,28,30]. Concept maps denote the knowledge concepts connected by relations with nodes and edges [2,3,20,27]. Compared with the if-then rules focusing on representing the causal relation, a concept map can represent arbitrary bilateral relations. Text segments, when properly selected and annotated, can represent knowledge in a traditional, easy to understand way. Liu et al. [25] select text segments from expert discussions as experiential tacit knowledge. The wiki techniques used to overcome the knowledge acquisition bottleneck present knowledge as

Table 1
Characteristics of knowledge acquisition research.

	Database	Interview	Conversation	Expert	MIS	Operation	Domain text
Knowledge source	Castro-Schez (2013) [21] Feng (2011) [22] Liu (2011) [23] Azadeh (2010) [13]	Cairo (2012) [24] Kwong (2009) [2] Wang (2009) [3]	Liu (2014) [25] Wagner (2006) [26]	Cheah (2011) [27]	Ruiz (2014) [20] Juarez (2009) [28]	Jin (2006) [29]	Wang (2010) [30]

	Experiential	Tacit	Personal	Operational	Uncertain
Knowledge type	Ruiz (2014) Liu (2014) Castro-Schez (2013) Liu (2011) Azadeh (2010) Juarez (2009) Wang (2009)	Liu (2014) Cairo (2012) Wang (2010) Kwong (2009)	Cheah (2011) Kwong (2009) Wagner (2006)	Jin (2006)	Feng (2011)

	If-then rule	Concept map	Annotated text	Ad-hoc
Knowledge representation	Ruiz (2014) Castro-Schez (2013) Feng (2011) Azadeh (2010) Wang (2010) Juarez (2009)	Ruiz (2014) Cheah (2011) Kwong (2009) Wang (2009)	Liu (2014) Wagner (2006)	Cairo (2012) notations Liu (2011) probabilistic graph Jin (2006) sequence

	Data mining	Manual	Fuzzy math	Wiki technique	Rough set
Knowledge acquisition method	Liu (2014) Ruiz (2014) Castro-Schez (2013) Liu (2011) Cheah (2011) Wang (2010) Wang (2009) Jin (2006)	Ruiz (2014) Cairo (2012) Azadeh (2010) Juarez (2009) Kwong (2009)	Castro-Schez (2013) Cheah (2011) Azadeh (2010) Wang (2009)	Wagner (2006)	Feng (2011)

hyperlink-enriched texts [26]. Researchers have also designed ad-hoc knowledge representation for their own problems. The graphical representation of knowledge can be extended with probability to signify the possibility of a state of a consequence node [23]. The branched sequences of actions are suitable for representing operational knowledge [29].

2.2.4. Knowledge acquisition method

Data mining, fuzzy math and manual extraction [2,13,20,24,28] are the most often used knowledge acquisition approaches. The data mining techniques used for acquiring knowledge include association rule mining [20], inductive learning [21], Bayesian network [23], regression [3], key graph algorithm [25], interesting pattern discovery [29] and neural network [30]. Liu et al. [25] propose a two-stage method for acquiring experiential tacit knowledge. With the recorded utterance of experienced engineers, they first obtain the tacit knowledge in the form of natural language and then apply the key graph algorithm to obtain the core content of experiential tacit knowledge. The use of data mining techniques helps to automate the process of knowledge acquisition, thus is indispensable for the application of knowledge acquisition in today's information-intensive work environment. Fuzzy math is usually used with other methods to accomplish knowledge acquisition. For example, Azadeh et al. use fuzzy math to transform the failure data of pumps into inexact concepts, with which they form a series of imprecise linguistic rules as the experiential knowledge [13]. Hence fuzzy math provides the bridge between numerical data and a knowledge concept. Manual work is often required in knowledge acquisition since the information in the knowledge source is high dimensional and can contain redundancies and contradictions. Cairo and Guardati try acquiring tacit knowledge through interview and manual modeling. They invite the interviewees to describe the knowledge usage events, and then use a special symbol system to express the concepts, processes, problems, structures and solutions found in the narratives [24]. To overcome the bottleneck of knowledge acquisition, Wagner proposes conversational knowledge management, which suggests the use of Wiki techniques to let the individuals share their personal knowledge spontaneously [26]. Although different knowledge acquisition methods have been proposed, the existing research shows a gap between the definition and acquisition of EK. As can be seen from Table 1, some researchers claim that they have acquired tacit or experiential knowledge [13,21], but the knowledge source they use is the objective data that has never been processed by a human brain, therefore the knowledge they have acquired does not match with the "personal" trait of tacit/experiential knowledge.

2.3. Context-aware knowledge management

The introduction of context awareness to knowledge management is driven by the consideration that knowledge is generated and applied in certain context. Context is any information that can be used to characterize the situation of an entity [31–33]. In knowledge management practices, context usually includes information dimensions like project, domain, activity, role, task, milestone and document. Being aware of the context allows a knowledge system to understand what the user is currently doing and what knowledge best matches with his/her needs. Generally, there are two ways to realize context awareness in knowledge management. The first is to use MIS as the source of contextual information [5,6,34,35]. The KnowMore system [5] is a typical system of this type. It observes the workflow of a business process to examine if any predefined knowledge need patterns are matched. If so, the system will retrieve knowledge suitable for solving the current problem and push it to the user. Shen et al. [35] conduct research aimed at providing active support for knowledge workers. They train a multinomial Naïve Bayes predictor to recognize the context that triggers knowledge needs. The training data is the task samples obtained from the Lotus Activities System. The second source of contextual information is the desktop of a knowledge worker [36–38]. Desktop activities can reflect the work content and target of a user. Lokaiczyk et al. [26] propose a just-in-time e-learning system using the desktop information including operating system events, user files, network stream and clipboard content. The PASTREM system [37] employs LDA (latent Dirichlet allocation) and topic model to identify the theme of a user's desktop activities and then recommends relevant knowledge to the user. To improve the precision of context awareness, Rath et al. [38] propose a User Interaction Context Ontology to map the raw desktop information into a unified data model. The problem with the existing context-aware knowledge management is that while the current research has exploited context awareness for knowledge reuse and recommendation, few studies have explored context awareness for knowledge acquisition.

3. Model of experiential knowledge

3.1. Definition of EK

In this paper, EK is defined by synthesizing the traits of experiential knowledge mentioned in existing literature. The traits listed in Table 2 are extracted in a literal fashion, for instance, in [8] Noh et al. describe experiential knowledge as "despite success or failure, his/her problem-solving experience is stored in their memory as tacit knowledge", so the "tacit" and "problem-solving oriented" traits can be extracted for this mentioning. In this example the word "his/her" implies that the owner of the experiential knowledge is an individual, so the "personal" trait can also be extracted.

Summarizing the traits listed in Table 2, we can define EK as below:

> Experiential knowledge (EK) is the tacit knowledge gained by an individual in the problem-solving context. It contains the problem encountered by an individual and solution reached by the individual. EK can be partially externalized in a case-like fashion.

The above definition emphasizes the problem-solving function and personal characteristic of EK. While the problem-solving function determines that EK should contain problem and solution as content, the personal characteristic determines EK's source can only be persons. Since personal knowledge is often incomplete and error-prone, EK cannot be represented as consistent and complete knowledge. So the best way to represent EK is to encapsulate the problem and solution in a certain case and allow contradictions to exist between different cases.

3.2. Representation of EK

In the proposed method EK is represented as a four-tuple $\langle C, P, R^{cl}, R^{df}\rangle$, where C is a set of concepts, P is a set of relations, R^{cl} is a set of classical inference rules and R^{df} is a set of default rules. Note that C, P and R^{cl} constitute an ontology called the experiential knowledge ontology (EK-Onto). Being an explicit description of conceptualization, ontology can function as a unified information structure capable of mediating the semantics of different information sources [39]. Through mapping the personal utterances to the EK-Onto, EK can be formally acquired and endowed with computer operability. Specifically, the EK-Onto contains 3 kinds of concepts and 12 kinds of relations. The 3 kinds of concepts are the nominal, status and action concepts. The 12 kinds of relations are listed in Table 3.

The nominal, status and action concepts are nouns, adjectives and verbs respectively. Such a classification of concepts helps us to maintain the linguistic legality of the relations formed in the knowledge acquisition and inference process. R^{cl} is used for inferring the facts that are not manifested in the environment. For example, the rule $Has\_status(?x, ?y) \wedge \text{Part\_}of(?x, ?z) \Rightarrow Has\_status(?z, ?y)$ in R^{cl} indicates that the status owned by a part is also owned by the whole. In Section 5 we will show how R^{cl} helps to decide which EK item to use when solving a problem.

The default rules in R^{df} are the core content of EK. Elicited from people's utterance and activities in problem solving, this part of EK reflects the idea of an individual about what problems may occur in a situation and what strategies are effective for solving the problems. Since EK is not precise knowledge where each concept and relation plays a definite role, it is natural for EK to include unnecessary information and leave out useful information. In light of this, it is inappropriate to require all the conditions in an EK item be exactly matched in order to reuse it. The default logic [40] is a non-monotonic logic adopting the following principle in reasoning: *as long as there is no evidence suggesting that the current conclusion does not hold, then the current conclusion holds*. Through combining the default logic

Table 3
Relations in EK-Onto.

Relation name	Explanation
Act_on	Current action is applied to an object represented by a nominal
Has_action	Current object receives an action
Has_status	Current object has a status
Has_subject	Current status is owned by an object
Is_a	Current concept is the subclass of another concept
Antonym	Two concepts have the opposite meaning
Cause	Current action/status causes another status
Caused_by	Current status is caused by an action or another status
Before	Current action is executed before another action
After	Current action is executed after another action
Part_of	Current object is a part of another object
Has_part	Current object has another object as a part

Table 2
Traits of experiential knowledge.

Trait of EK	Problem solving oriented	Contextual	Personal	Tacit	Is know-how	Is case	Externalizable
Related research	[8] [12] [28] [15] [16] [25]	[28] [13] [14] [15] [16] [25]	[8] [9] [10] [17] [25]	[8] [10] [16] [25]	[9] [28] [16] [25]	[18] [19]	[17] [20]

with the EK-Onto, the imprecise character of EK can be addressed.

A default rule in R^{df} has the following form:

$$\frac{\mathrm{INCLUDE}(\gamma,\beta) \wedge \mathrm{ANSWER}(\alpha,\beta) : \mathrm{CONSISTENT}(\alpha,\gamma)}{\mathrm{SOLVE}(\alpha,\gamma)} \tag{1}$$

where α, β and γ are propositions connected by the conjunction symbol "⋀". A proposition is composed of the instantiated EK-Onto relations. The meanings of the notations in formula (1) are:

α: the content of an EK item, which corresponds to the solution of a problem;
β: the usage condition of an EK item, which corresponds to the description of a problem;
γ: the context of knowledge activity, which corresponds to the information appearing on the desktop of a user;
INCLUDE(γ, β): is true if and only if γ contains all the propositions in β;
ANSWER(α, β): is true if and only if α is the solution of β;
CONSISTENT(γ, α): is true if and only if each proposition in α is consistent with the propositions in γ;
SOLVE(α, γ): is true if and only if α is helpful for solving γ.

Formula (1) describes a default rule since it does not require INCLUDE(γ, β), ANSWER(α, β) and CONSISTENT(γ, α) be all true to confirm SOLVE(α, γ). Given that INCLUDE(γ, β) and ANSWER (α, β) are satisfied, following the reasoning principle of default logic, unless there is known contradiction between a proposition from α and a proposition from γ, SOLVE(α, γ) will be inferred as true even if some propositions in α cannot be judged whether they are in accordance with γ. As such, representing EK with default rules makes the reasoning with EK more flexible.

4. Context-aware EK acquisition

4.1. Q&A as EK source

EK is generated in human brains and is intangible until it is said out or used to solve a problem. To acquire EK, one needs a tangible source of EK to use as input. In this study, we use the Experiential Q&A (EQA) as the source of EK. EQA refers to the Q&A taking place in the problem-solving process and about how to achieve a goal. For example, a Q&A with the topic "how to model layered material in bending analysis" is an EQA, while questions like "when was Lincoln born" do not produce EQA. As EQA contains the personal ideas about problem solving, it matches the definition of EK in this paper.

EQA itself can have different sources, e.g., face-to-face communication, online forum discussion, email correspondence, instant messages, etc. EQA from different sources has different traits. For example, the EQA in online forums is usually contributed by people working for different organizations, so the question askers have to describe their problems with a full explanation of the context in order for others to understand. For the email-based EQA contributed by a group of collaborating people, the context may not be deliberately explained as it is already shared by the group members. In this paper, focusing on acquiring EK from a group of collaborating people, the contextualization of the EQA – finding the information related with but not directly adjacent to a question, must be resolved.

4.2. The overall process of EK acquisition

The overall process of EK acquisition is shown in Fig. 1. The process requires three types of input:

- Structured domain text, which is used to build the EK-Onto.
- EQA accumulated in the online forums, which is used to obtain the Q&A term association.

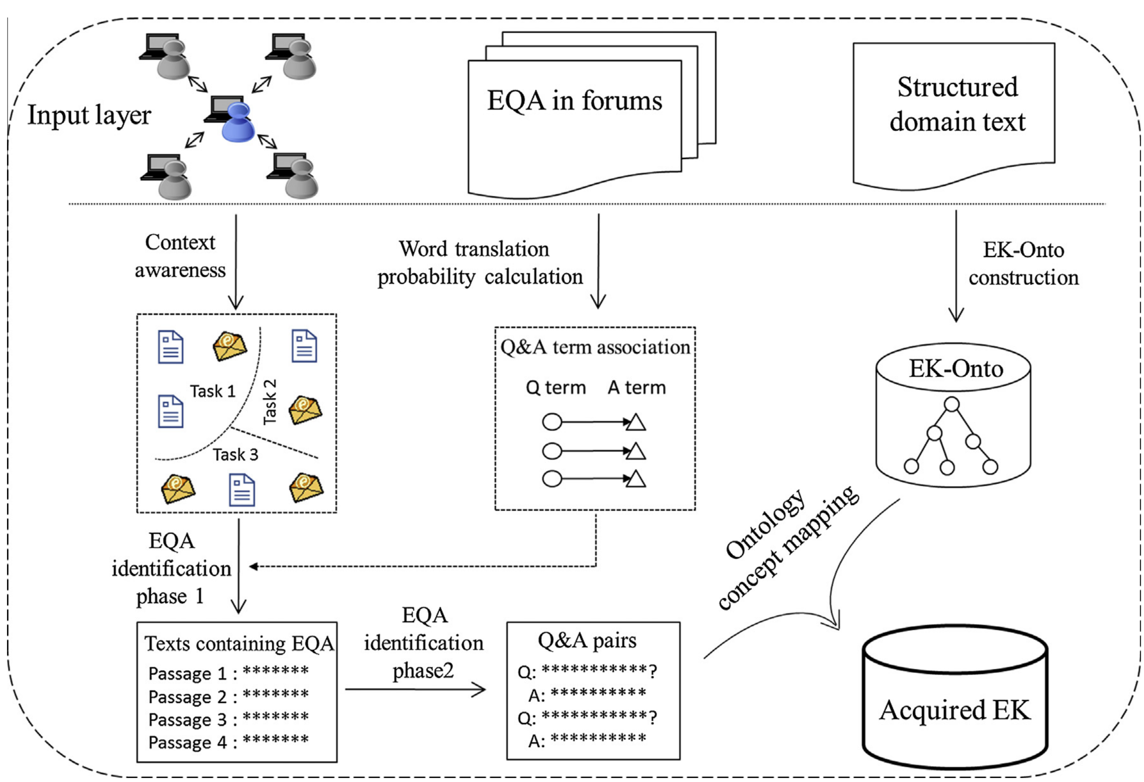

Fig. 1. The overall process of EK acquisition.

- A user's desktop knowledge activities, which serve as the source of EQA.

The structured domain text refers to the pieces of text enumerating the basic concepts and relations in the knowledge domain, e.g., a domain thesaurus or the catalogue of a textbook. With this type of text as input, the ontology codifying the domain knowledge can be automatically built. The EQA accumulated in online forums contains many problem descriptions and the according solutions. Through analyzing them it can be known that for a given concept appearing in the question what concepts are more likely to appear in the answer. Such an association between the Q&A terms enables us to find the text segments that are more likely to form Q&A pairs. However, in a multitasking work environment the identification of accurate Q&A pairs cannot be so easy since the engineers' activities may be targeted at different tasks. If the text segments from different tasks are selected to form Q&A pairs, mismatches are likely to happen. To solve this problem, a clustering-based context awareness mechanism is proposed to group the desktop information according to which task it belongs to. After this, the first round of EQA identification is performed within each information cluster that represents an independent task. The second round of EQA identification is performed by a CRF (Conditional Random Field) model trained over some manually labeled forum discussion threads which contain multiple rounds of questioning and answering. At last, the texts in the Q&A pairs are parsed sentence by sentence and the phrases in them are mapped to the concepts and relations defined in the EK-Onto. The acquisition of EK is finished when the default rule corresponding to each EQA is constructed.

4.3. EK-Onto construction using structured domain text

In order to develop the ontology with efficiency, we use natural language processing techniques to extract the concepts and relations composing the EK-Onto from the structured domain text. A piece of the structured domain text is shown in Table 4, which is excerpted from the catalogue of ANSYS software tutorial.

From the content and the hierarchical structure of the catalogue, we can obtain the following information for building the EK-Onto:

- concepts representing the names of entities, e.g., *Analysis*, *Model*, *Solution Control*, etc.
- concepts representing the status of entities, e.g., *Static*
- concepts representing the action of people, e.g., *Perform*, *Build*, *Set*, etc.
- relations between concepts, e.g., Has_status(analysis, static), Act_on(perform, analysis), Has_part(analysis, model), Before (build, set), Is_a(displacement, load), etc.

To obtain the above information, we apply part-of-speech tagging to the text and then match the tagged text with a series of semantic patterns. For example, using the pattern of an adjective followed by nouns, the instances of the Has_status relation can be found. The instances of the Has_part relation can be found by detecting a noun appearing in the parent directory and a noun appearing in the child directory. For the instances of Antonym and Cause relation, they are captured by the following two ways. The first way is to use a semantic dictionary such as WordNet. For example, according to WordNet the status "static" has the status "dynamic" as its antonym, and the action "divide" can cause the status "separate". But in some specific knowledge domain, the terminologies may have Cause and Antonym relation that is not defined in a general dictionary, so the second way of acquiring the relation instances is to ask domain experts to define them. In this paper, semantic patterns, WordNet and manual definition are collaboratively used to establish the EK-Onto.

Table 4
Example of structured domain text.

Performing a Static Analysis
Build the Model
Set Solution Controls
......
Apply the Loads
Load Types
Displacements (UX, UY, UZ, ROTX, ROTY, ROTZ)
Velocities (VELX, VELY, VELZ, OMGX, OMGY, OMGZ)

4.4. Translation probability-based Q&A term association

Given a lot of Q&A pairs, the terms in them can show some statistical associations. For example, when a certain term appears in the question, some terms may be more likely to appear in the answer. To quantify this type of association, Xue et al. compute the translation probability between a question term and an answer term by treating all the questions as a language to be translated and all the answers as the language that the questions are translated into [41]. Given a set of Q&A pairs $\{(\boldsymbol{q}_1, \boldsymbol{a}_1), (\boldsymbol{q}_2, \boldsymbol{a}_2), \ldots, (\boldsymbol{q}_N, \boldsymbol{a}_N)\}$, $P_{TR}(w|w')$, the probability of translating term w' into term w, is computed as follows:

$$
\begin{aligned}
P(w|w') &= \lambda_{w'}^{-1}\sum_{i=1}^{N} c(w|w';\boldsymbol{a}_i,\boldsymbol{q}_i) \\
c(w|w';\boldsymbol{a}_i,\boldsymbol{q}_i) &= \frac{P(w|w')}{\sum_{w''\in\boldsymbol{q}_i}P(w|w'')}\#(w,\boldsymbol{a}_i)\#(w',\boldsymbol{q}_i) \qquad (2) \\
\lambda_{w'} &= \sum_{w}\sum_{i=1}^{N} c(w|w';\boldsymbol{a}_i,\boldsymbol{q}_i)
\end{aligned}
$$

where $\#(w, \boldsymbol{a}_i)$ is the number of times the term w appears in answer $\boldsymbol{a}_i$, and $\#(w', \boldsymbol{q}_i)$ is the number of times the term w' appears in question $\boldsymbol{q}_i$. The computation is iterative: starting by arbitrarily setting the initial value of $P(w|w')$, the equations are repeatedly called to update $P(w|w')$ until it converges. The converged value of $P(w|w')$ is the $P_{TR}(w|w')$. As $P_{TR}(w|w')$ represents the probability of the term w being in the answer of a question containing the term w', we can use it to measure the correlation between two text segments – if the terms in one of the text segments are likely to be translated into the terms in the other text segment, then the two text segments may form a Q&A pair.

4.5. Context-awareness for engineering knowledge activity

In the field of computer-aided engineering (CAE), a typical knowledge worker (engineer) would use a computer to complete a series of design, calculation, and simulation tasks meanwhile collaborating with other workers via email and meeting. The information generated in these engineering activities becomes the source of EQA and EK. To organize this heterogeneous information and clarify the relationships between the different information objects used in this paper, we present the knowledge activity context model shown in Fig. 2.

As mentioned before, the central functionality of context awareness is to group the context information items according to which task they belong to. To decide which information items should be grouped together, we calculate the following types of *distance* between two information items:

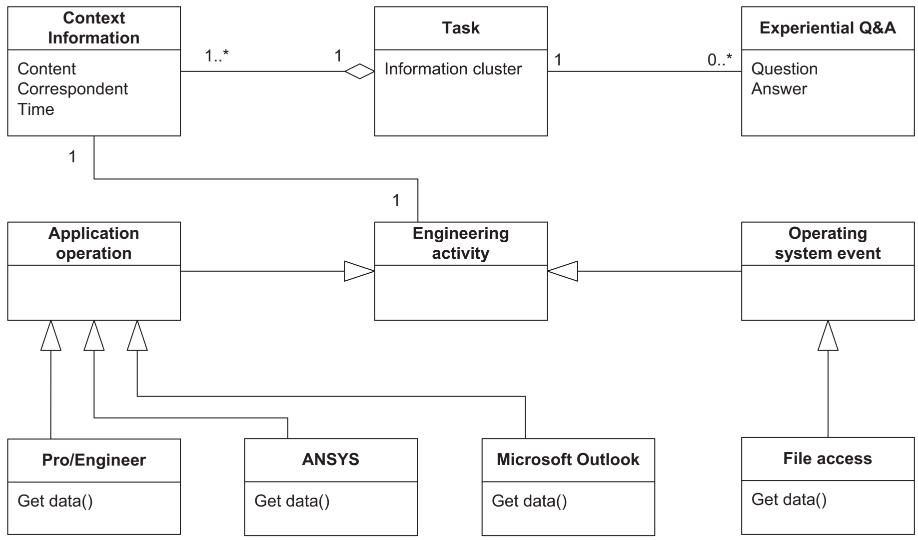

Fig. 2. Context model of CAE knowledge activity.

- Content distance

$$Dis_{con}(s_i, s_j) = 1 - Sim_{cos}(s_i, s_j) \quad (3)$$

where s_i and s_j denote two information items, Sim_{cos} is the cosine similarity between the vectorized representation [42] of s_i and s_j.

- Correspondence distance

$$Dis_{cor}(s_i, s_j) = 1 - \frac{1}{2} \cdot \left(\frac{N(cor(s_i), cor(s_j))}{N(cor(s_i))} + \frac{N(cor(s_i), cor(s_j))}{N(cor(s_j))} \right) \quad (4)$$

where $cor(s_i)$ represents the correspondent of s_i, which refers to the person who sends s_i to the user or the person to whom the user sends s_i. $N(cor(s_i), cor(s_j))$ is the number of times $cor(s_i)$ and $cor(s_j)$ communicate with the user regarding the same task. $N(cor(s_i))$ is the number of times $cor(s_i)$ communicates with the user.

- Time distance

$$Dis_{tim}(s_i, s_j) = \log_{10}(1 + \|time(s_i) - time(s_j)\|) \quad (5)$$

where $time(s_i)$ is the time that s_i is last accessed. The difference between times is measured in hours.

- Total distance

$$Dis(s_i, s_j) = w_{con} \cdot Dis_{con}(s_i, s_j) + w_{cor} \cdot Dis_{cor}(s_i, s_j) + w_{tim} \cdot Dis_{tim}(s_i, s_j) \quad (6)$$

With the total distance between each two information items, we can use the systematic clustering method to generate the information collections representing different tasks. To achieve more accurate context information clustering, we adopt a regression-based method to adjust the three weights w_{con}, w_{cor} and w_{tim}. To do this first 40 context information items were selected and manually clustered into 6 tasks, and then a series of training data were generated by traversing through each pair of the information items meanwhile setting the three types of distance as the independent variables and the classification label ("1" for belonging to the same task and "0" for not) as the dependent variable. After applying logistic regression to the training data, the odds ratios of the three types of distance were obtained. Since (1 − odds ratio) represents how much the possibility of two information items belonging to one task shrinks if the corresponding type of distance is increased by 1, we use this value as the weight of the corresponding distance type in the total distance formula. Regarding how many clusters should be derived for the given information items, we adopt the rule of thumb $\sqrt{n/2}$ [43] as it is not known how many tasks a user is processing, where n is the total number of the information items.

4.6. EQA identification

Extracting EQA from engineering knowledge activity makes sense because the collaboration between engineers often leads to communications regarding the happening and solving of problems. To backup this idea, we show in Table 5 the communication types that frequently occur in the engineering knowledge activity and relate each type of the communication with its potential roles in EQA.

From Table 5 we can see that even without clearly asked questions, the context information of engineering knowledge activity is closely related with EQA. To extract EQA from this context information, we go through the following two phases:

4.6.1. Phase 1: Extracting text segments composing potential EQA

Although the communication patterns enumerated in Table 5 demonstrate a close relationship between the knowledge activity and EQA, they are not directly used for extracting EQA since they cannot be easily recognized by computer programs. To extract the communication content likely to form EQA, we exploit the Q&A term association discovered using the method introduced in Section 4.4. Let $T = (I_1, I_2, \ldots, I_n)$ be a sequence of information items addressing the same task and arranged in ascending order according to their access time, let T_{QA} be the sequence of text segments in T being likely to form EQA. T_{QA} is obtained by performing the following operations:

```
For i = 1 to n, i++
  Choose I_i from T;
  Get text segment g_i = f_toText(I_i);
  If (i = 1 ||g_i contains a question|| g_i is replied by a following
  message)
    Add g_i to T_QA;
  Else
    For each g in T_QA
      If (g_i is a reply to g)
        Add g_i to T_QA;
        Break;
      Else
      If (P_QA(g, g_i) > AvgP_QA)
        Add g_i to T_QA;
        Break;
```

Table 5
Communication types and their potential roles in EQA.

	Comm. Type	Originator	Recipient	Description	Role in EQA
1	Receiving a task	Superiors	User	Initial task information, contains the overall description of a task	Question
2	Result submission	User	Superiors	Integrated solution to the task, containing EK of the user and his/her subordinates	Answer
3	Receiving feedbacks	Superiors	User	Evaluation of problem solution, negative opinion can cause new problems	Question
4	Task allocating	User	Subordinates	Describing sub-problems and reflecting the user's understanding of the problem	Answer
5	Receiving results	Subordinates	User	Solutions to sub-problems, reflecting subordinates' EK	Answer
6	Asking a question	User	Expert	Describing a sub-problem the user encounters	Question
7	Receiving an answer	Expert	User	Solutions to sub-problems, EK of colleagues/experts	Answer
8	Being inquired	Novice	User	Describing a sub-problem	Question
9	Replying to an inquiry	User	Novice	Solutions to sub-problems, reflecting the user's EK	Answer
10	Task processing	User	User	Problem solving operations	Answer

where $f_{toText}(I_i)$ is a function to extract the first 100 consecutive words in an information item and $P_{QA}(g, g_i)$ is the possibility that text segment g_i is the answer of text segment g. $AvgP_{QA}$ is the averaged translation probability between the questions and answers in the chosen domain. $P_{QA}(g, g_i)$ and $AvgP_{QA}$ are calculated as follows:

$$P_{QA}(g, g_i) = \frac{1}{|g|}\sum_{w' \in g} max_{w \in g_i} P_{TR}(w|w') \quad (7)$$

$$AvgP_{QA} = \frac{1}{|D|}\sum_{(q,a) \in D} P_{QA}(q, a) \quad (8)$$

where $|g|$ is the number of words in g, (q, a) is a Q&A pair collected from the domain forums and D is the set of all Q&A pairs.

4.6.2. Phase 2: CRF-based Q&A pair identification

After phase 1, the half-extracted EQA takes the form of sequenced text segments contributed by different people. To accurately recognize the Q&A pairs in these texts, we use the method proposed by Ding et al. in a study aimed at extracting Q&A from online forums [44]. In a forum discussion thread, there may be several questions and answers, but they are usually scattered throughout the discussion and blended with useless information. To extract the accurate Q&A pairs, Ding et al. trained a CRF model to discriminate the role of each sentence in the discussion thread regarding a selected question. In this paper, we train a similar linear chain CRF model to complete the 2nd phase of EQA identification. This includes selecting a number of discussion threads that contain multiple rounds of Q&A from domain online forums and annotating each sentence in them by the label of *question*, *context*, *answer* and *plain*. Then the CRF model is trained with a series of vocabulary, semantic and structure features of the sentences. After applying the trained CRF model to the text segment sequences obtained in phase 1, the accurate EQA can be acquired.

4.7. Ontology concept mapping

The final step of EK acquisition is to transform the text in the recognized EQA into the concepts and relations defined in the EK-Onto, and meanwhile establish the default rules. When accomplishing text transformation, the key operation is to map a phrase in the EQA to its most similar concept in the EK-Onto. Suppose a phrase $\boldsymbol{p}$ contains words $p_1, p_2, \ldots, p_m$ and an ontology concept $\boldsymbol{q}$ contains words $q_1, q_2, \ldots, q_n$, then the semantic similarity between $\boldsymbol{p}$ and $\boldsymbol{q}$ is calculated as:

$$Sim_{phr}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{m} \max_{j} Sim_{JCn}(p_i, q_j) \quad (9)$$

where Sim_{JCn} is the WordNet-based JCn similarity of words [45].

Besides the individual phrases, the semantic relations between the phrases are also mapped to the EK-Onto. This is done by parsing the sentences with Stanford Parser [46]. Taking the sentence "it is trying to simulate the effects of a layered panel" for example, after parsing, the sentence can be resolved into semantic relations including *dobj* (simulate-5, effects-7), *amod* (panel-11, layered-10), *prep_of* (effects-7, panel-11), etc. Among these relations, *dobj* corresponds to the relation Act_on in the EK-Onto, *amod* corresponds to Has_status in the EK-Onto, and the prepositional relations are mapped to Part_of in the EK-Onto.

5. Application of proposed method

5.1. Data preparation

In this paper, the proposed EK acquisition method is applied to the FEA (finite element analysis) tasks in a chain manufacturing company. For this we first build the EK-Onto for the target domain. Since the FEA tasks are using the software ANSYS to simulate the mechanical performance of mooring chains, the structured domain text for building the EK-Onto is chosen as the catalogue of the ANSYS Mechanical APDL Structural Analysis Guide and the standards of the mooring chain components. Fig. 3 shows a part of the established EK-Onto.

The data used for deriving the Q&A term association and training the CRF model are collected from the domain online forums. Totally 16,307 discussion threads were downloaded from two forums, namely the xansys.org and physicsforums.com. By treating the first post of each discussion thread as a question and the first reply in each thread as the corresponding answer, we use the GIZA++ software package implementing the IBM machine translation model to generate the translation probability-based Q&A term association. From all the discussion threads, 380 of them were selected and manually annotated as the training samples for the CRF model. The detail about training the CRF model can be found in [47], which is a previous work of ours studying how to predict people's knowledge needs with the accumulated forum Q&A.

5.2. An EK acquisition case

The proposed EK acquisition method is tested in a chain-manufacturing company where engineers carry out FEA to examine the strength of chain components. FEA is a knowledge intensive task because different product shapes, structures, materials, contacts, loads, boundary conditions and model-solving methods interact deeply to make the situation complicated. FEA requires both the disciplinary knowledge and the practical experience of engineers to be manipulated properly. In the chain-manufacturing company, a three-person group (shown in the dotted box of Fig. 4) is set up to handle the FEA tasks of chain components. In its daily work, the FEA group may receive FEA tasks from clients or the design department of the company. There are often several tasks being processed at the same time by the group in a cooperative manner. When meeting problems in handling FEA tasks, the group members may consult each other or ask for help

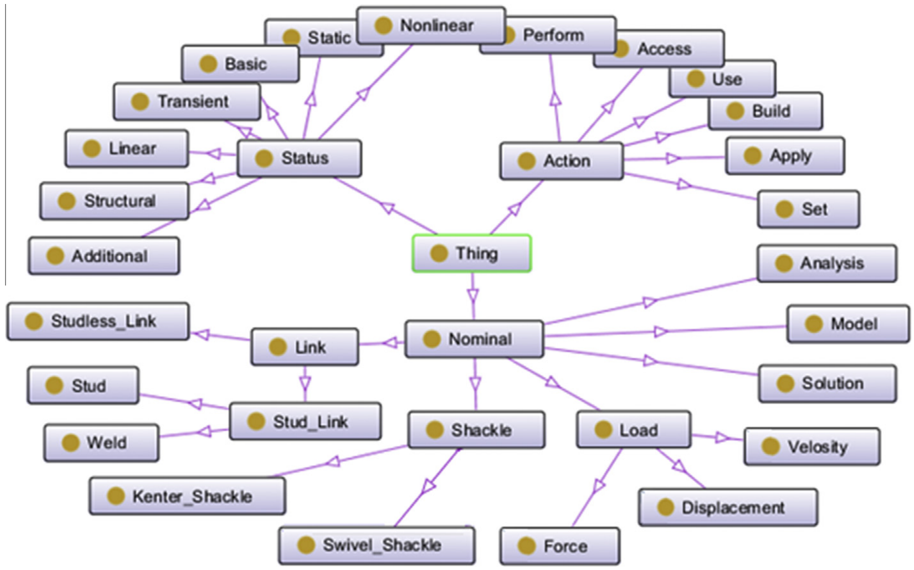

Fig. 3. Part of the EK-Onto developed for mooring chain FEA.

from experts outside the group to get the problems solved. Fig. 4 shows the trajectory of a FEA task's processing. The information item captions shown on the right hand side of Fig. 4 are numbered in a chronological order according to when the item is generated. An arrowed line on the left hand side of Fig. 4 signifies the sender and receiver of an information item of the type "email". The information items appearing next to a person mean they are generated by the person's desktop activities.

In the case shown in Fig. 4, the leader of the FEA group, Zhu.L.F, received an email entitled "NPS job AT14011 Swivel Shackle" (information item 1, referred to as IT1 in the below) from the client Hogsaet.I.S. The email was attached with a file named "Specifications of swivel shackle AT14011.pdf" which was browsed by Zhu. L.F on his desktop (IT2). After reading the email and the attachment, Zhu.L.F divided the incoming task into two parts and sent them (IT3) to the group member Zhang.B.F and Liu.G. Then Zhu.L. F proceeded with his previous work (IT4). After receiving the assigned tasks, Liu.G informed Zhang.B.F that he would wait for the fatigue load simulation result to begin the calculation of chain lifetime (IT5). Zhang.B.F read the specifications of the product (IT6) and then started to build the product's finite element model using ANSYS (IT7). During this phase, Zhang.B.F encountered a problem of modeling a screw connection in the swivel shackle, regarding which he wrote an email (IT8) to ask Zhu.L.F for help. In the reply, Zhu.L.F suggested a simplified simulation of the connection (IT9). Zhang.B.F carried on with the analysis (IT11) and meanwhile sent the fatigue result required for lifetime calculation to Liu.G (IT12). Liu.G wrote a report containing the lifetime calculation result of the product (IT13) and sent it back to Zhang.B.F (IT14) for result integration. The integrated report (IT15) was sent to Zhu.L.F (IT16) for checking, who modified it (IT17) and then sent it to the client (IT18). A few days later, the client returned his comments on the report (IT19), pointing out that the screw connection had been oversimplified. Since no one in the FEA group knew how to accurately simulate a screw connection, Zhu.L.F wrote an email (IT20) to Yao.Y to ask for help.

The information items shown in Fig. 4 are selected regarding the FEA task of the swivel shackle AT14011. In the real situation, these information items are blended with the information belonging to other tasks on an engineer's desktop. Taking the desktop of Zhu.L.F for example, the left hand side of Fig. 5 shows all the information items owned by Zhu.L.F shortly after he replies the ques-

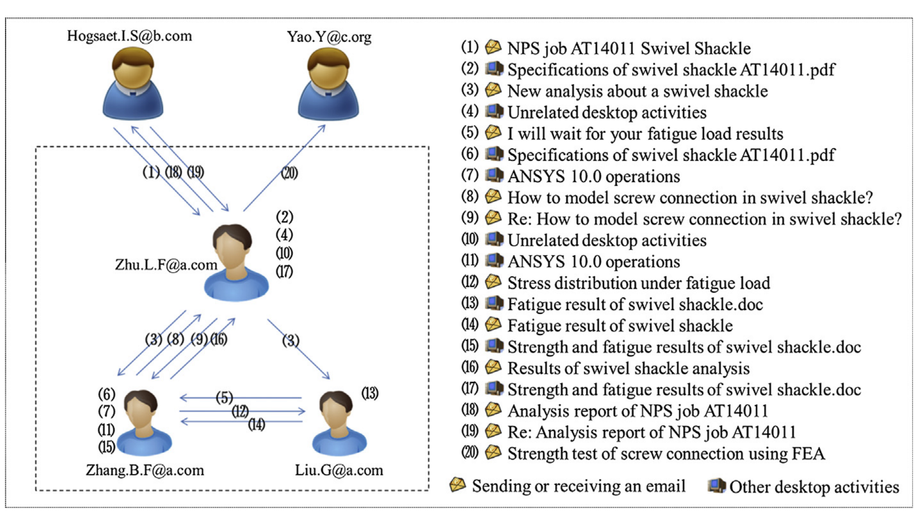

Fig. 4. Information items belonging to a FEA task.

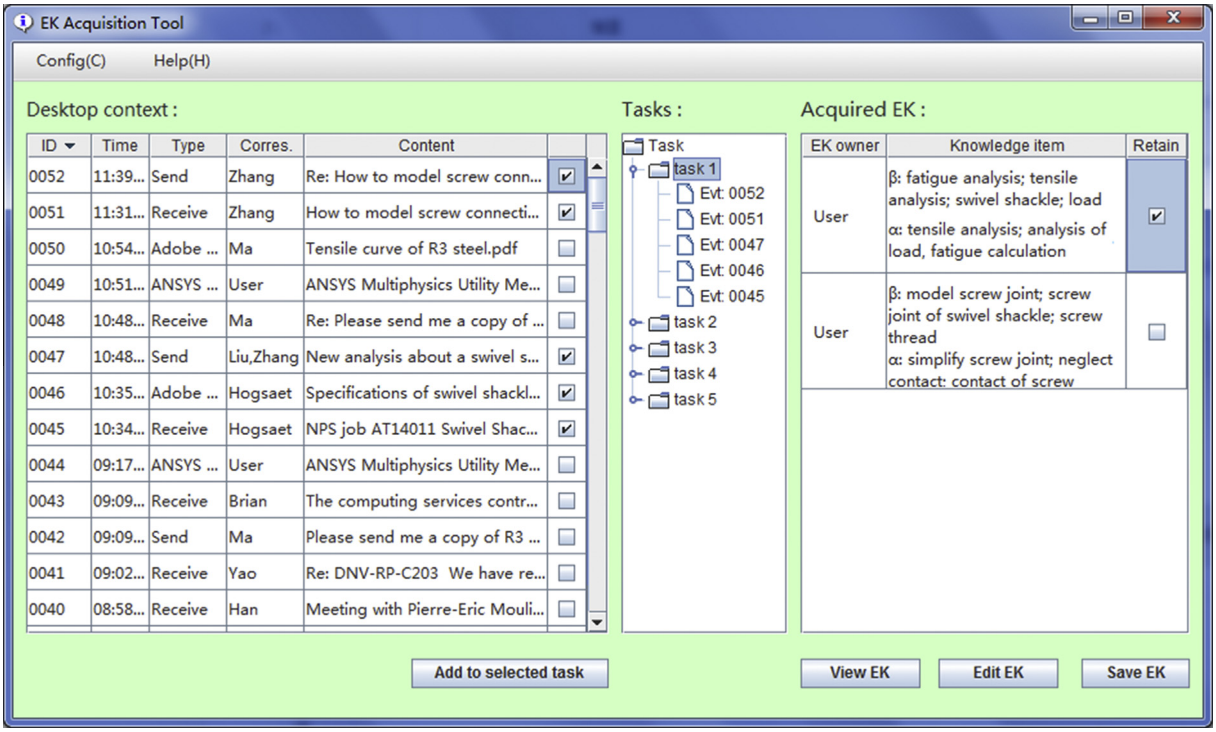

Fig. 5. Interface of EK acquisition.

tion raised by Zhang.B.F. It can be seen that in addition to the task related with the swivel shackle AT14011, there are other tasks processed by Zhu.L.F at the same time, e.g., the task numbered DNV-RP-C203. The user interface shown in Fig. 5 is a component of the EK acquisition prototype system developed in this paper, which visualizes the clustering-based context awareness process described in Section 4.5. The clustering result is shown in the middle of the interface. From Fig. 5 we can see that the information items owned by Zhu.L.F are classified into 5 tasks. The first task currently selected by the user contains the information items 0052, 0051, 0047, 0046 and 0045, which correspond to the FEA of shackle AT14011. The right hand side of the interface shows the acquired EK items regarding the selected user task.

Within each user task, the EQA identification process described in Section 4.6 runs. In the FEA case, the EQAs identified on Zhu.L.F's desktop are shown in Table 6. A sentence labeled with "Q: #" means that the sentence is the #th question in the current task. The label "C: #", "A: #" and "P: #" indicate that the sentence is the context, answer or useless information regarding the #th question. Due to the limitation of space, Table 6 does not show all the sentences in the task context.

The acquired EK items are shown in Table 7. Note that they are tailored to reflect only the information presenting in Table 6. When transforming an EQA into an EK item, we first map the words in the question and the words appearing in both of the context and answer of the question to the concepts in EK-Onto. The concept instances and their relations established in this step are encoded as the usage condition (which is β in formula (1)) of the corresponding EK. Then we map the words in the answer of the question to the EK-Onto and create the content of the EK. Take the EQA consisting of sentence No. 1, No. 2 and No. 3 for example, we can obtain the EK item shown in the first row of Table 7. Note that the context sentence No. 5 and No. 8 do not affect the acquisition of this EK item since they have not reached the answerer when the question is being answered. From the acquired EK, we can see that regarding the same question, different people may contribute different knowledge in which some is proper but other is not. In this case we can retain the last EK item with respect to the same question since the subsequent solutions are usually better than their unsuccessful predecessors. We can also use some simple rules to filter out the EK that is not meaningful. For example, by discarding the EK items containing no action concepts or status concepts besides "existing", we can exclude the EK that cannot guide people to achieve a goal.

5.3. Reusing the acquired EK

The value of the acquired EK can only be demonstrated when it is effectively reused in problem solving. For the EK acquired and represented with the methods proposed in this paper, we provide an example to show how it can be reused to solve engineering problems. Suppose there is the following EK in the knowledge base:

$$\beta = \text{Has\_status}(\text{shell},\ \text{existing}) \bigwedge \text{Has\_status}(\text{node},\ \text{existing}) \bigwedge \text{Has\_part}(\text{model},\ \text{shell}) \bigwedge \text{Act\_on}(\text{mesh},\ \text{joint})$$

$$\alpha = \text{Has\_status}(\text{region}, \text{rigid}) \bigwedge \text{Act\_on}(\text{use}, \text{CERIG})$$

and suppose the current user task involves the following facts:

$$\gamma = \text{Has\_status}(\text{element},\ \text{existing}) \bigwedge \text{Has\_part}(\text{model},\ \text{weld}) \bigwedge \text{Act\_on}(\text{mesh},\ \text{weld})$$

and there is the following common knowledge in the EK-Onto:

1. Is_a(weld, joint)
2. Is_a(shell, element)
3. Is_a(model, NOMINAL)
4. Act_on(mesh, model) $\wedge$ Cause(mesh, existing) $\wedge$ Has_subject (existing, node)

Then the process of inferring the extended fact set γ' from γ is as follows:

Table 6
EQA identification results.

No.	Sentence	Author	Role label
1	Please carry out the required tensile and fatigue analysis for the swivel shackle described in the attachment	Hogsaet. I.S	Q: C: 1 A: P: 2, 3
2	The tensile tests include the proof load test and the break load test	Hogsaet. I.S	Q: C: 1 A: P: 2, 3
3	Zhang is responsible for the tensile load analysis and Liu is responsible for the fatigue life analysis	Zhu.L.F	Q: C: A: P: 2, 3 1
4	How to model screw connection in swivel shackle	Zhang. B.F	Q: C: 2 A: P: 1, 3
5	The screw connection in the shackle is too complicated to model if the screw thread is modeled as 3D solid	Zhang. B.F	Q: C: 1, 2 A: P: 3
6	You may simplify the screw connection by neglecting the thread contact and directly coupling the two cylindrical surfaces underneath the thread	Zhu.L.F	Q: C: 3 A: P: 1 2
7	The calculation results show that the swivel shackle is able to withstand both the proof load and the break load	Zhang. B.F	Q: C: 2 A: P: 3 1
8	Since the screw joint is nonstandard and supports the major load on the shackle, it cannot be overlooked in the finite element analysis	Hogsaet. I.S	Q: C: 1, 3 A: P: 2
9	I have a chain shackle containing a nonstandard screw connection which bears tensile load along the axial direction	Zhu.L.F	Q: C: 3 A: P: 1, 2
10	I want to know if I have to model the screw connection in 3D and deal with the area contact between the threads	Zhu.L.F	Q: C: 3 A: P: 1, 2

Table 7
Acquired EK items.

Question	Owner	EK item
1	Zhu.L.F	β_1 = Has_status(analysis, fatigue) ⋀ Has_status(analysis, tensile) ⋀ Has_status(swivel shackle, existing) ⋀ Has_status(load, existing) α_1 = Has_status(analysis, tensile) ⋀ Part_of(analysis, load) ⋀ Has_status(fatigue calculation, existing)
1	Zhang.B.F	β_2 = Has_status(analysis, fatigue) ⋀ Has_status(analysis, tensile) ⋀ Has_status(swivel shackle, existing) ⋀ Has_status(load, existing) α_2 = Part_of(result, calculation) ⋀ Has_status(swivel shackle, existing) ⋀ Act_on(withstand, load) ⋀ Has_status(result, existing)
2	Zhu.L.F	β_3 = Act_on(model, screw joint) ⋀ Part_of(screw joint, swivel shackle) ⋀ Has_status(screw thread, existing) ⋀ Has_status(screw joint, existing) α_3 = Act_on(simplify, screw joint) ⋀ Act_on(neglect, contact) ⋀ Part_of(contact, screw thread) ⋀ Act_on(couple, surface) ⋀ Has_status(surface, cylindrical) ⋀ Part_of(surface, screw thread)
2	Hogsaet.I.S	β_4 = Act_on(model, screw joint) ⋀ Part_of(screw joint, swivel shackle) ⋀ Has_status(load, existing) ⋀ Has_status(screw joint, existing) α_4 = Has_status(screw joint, nonstandard) ⋀ Act_on(support, load) ⋀ Has_status(load, major) ⋀ Part_of(load, shackle) ⋀ Has_status(finite element analysis, existing)

Table 8
Reasoning rules in R^{cl}.

(1)	X(?x) ⋀ Is_a(X, Y) ⇒ Y(?x)
(2)	Part_of(?x, ?y) ⋀ Part_of(?y, ?z) ⇒ Part_of(?x, ?z)
(3)	Has_action(?x, ?y) ⋀ Part_of(?x, ?z) ⇒ Has_action(?z, ?y)
(4)	Has_ status(?x, ?y) ⋀ Part_of(?x, ?z) ⇒ Has_ status(?z, ?y)
(5)	Before(?x, ?y) ⋀ Before(?y, ?z) ⇒ Before(?x, ?z)
(6)	Antonym(?x, ?y) ⋀ X(?x, ?x1) ⋀ Y(?y, ?y1) ⋀ Has_subject(?x1, ?s1) ⋀ Has_subject(?y1, ?s2) ⋀ S(?s1, ?s2) ⇒ Antonym(?x1, ?y1)
(7)	Act_on(?x, ?y) ⋀ Cause(?x, ?z) ⋀ Has_subject(?z, ?s) ⋀ Act_on(?x1, ?y1) ⋀ X(?x, ?x1) ⋀ Y(?y, ?y1) ⇒ Z(?z, new ?z1) ⋀ S(?s, new ?s1)) ⋀ Has_status(?s1, ?z1)
(8)	Has_status(?x, ?y) ⋀ Cause(?y, ?z) ⋀ Has_subject(?z, ?s) ⋀ Has_status(?x1, ?y1) ⋀ X(?x, ?x1) ⋀ Y(?y, ?y1) ⇒ Z(?z, new ?z1) ⋀ S(?s, new ?s1) ⋀ Has_status(?s1, ?z1)

First, by applying the first rule shown in Table 8 to all the concept instances in γ whose father class is not "NOMINAL", "ACTION" or "STATUS", we get

$\gamma^{(1)}$ = Has_part(model, joint) ⋀ Act_on(mesh, joint)

This step is called generalization inference since it uses a concept's father class to replace itself.

Then, by applying the rule (3) and rule (4) to $\gamma \wedge \gamma^{(1)}$ we obtain

$\gamma^{(3)(4)}$ = Act_on(mesh, model)

Then, by applying the rule (7) and rule (8) to $\gamma \wedge \gamma^{(1)} \wedge \gamma^{(3)(4)}$ we get

$\gamma^{(7)(8)}$ = Has_subject(existing, node) = Has_status(node, existing)

Until now we have

$\gamma' = \gamma \wedge \gamma^{(1)} \wedge \gamma^{(3)(4)} \wedge \gamma^{(7)(8)}$ = Has_status(element, existing) ⋀ Has_part(model, weld) ⋀ Act_on(mesh, weld) ⋀ Has_part (model, joint) ⋀ Act_on(mesh, joint) ⋀ Act_on(mesh, model) ⋀ Has_status(node, existing)

To determine whether INCLUDE(γ, β) holds, we first extract the status and action facts from β and express them in the conjunctive form: Has_status(shell, existing) ⋀ Has_status(node, existing) ⋀ Act_on(mesh, joint). Then we check each of the above facts to see if it is included in γ', whereby we can find the latter two appearing in γ'. For the fact Has_status(shell, existing), although it does not appear in γ' directly, its variation Has_status(element, existing) obtained by the generalization inference is included in γ', so INCLUDE(γ, β) is met. Finally, we check whether the facts in α and γ conflict with each other. The fact Has_status(region, rigid) in α emphasizes that the action Act_on(use, CERIG) applies when there are rigid regions in the model. Although the current γ does not mention whether the model contains a rigid area, but according to the principle of default reasoning, unless γ explicitly mentions a conflicting fact such as Has_status(region, deformable), the conclusion SOLVE(α, γ) will be confirmed and the action mentioned in α would be deemed as helpful for the current user task.

5.4. Comparison with other EK acquisition methods

To verify the effectiveness of the proposed method, we compare our method with existing tacit/experiential knowledge acquisition methods. The potential rivals are the methods listed in Table 1 which use narratives as the knowledge source. The knowledge acquisition methods using the database as the knowledge source are not chosen to compare with the proposed method as there is no direct way to transform the textual input used in this paper to the numerical input required by those methods. The methods using experts or MIS as the knowledge source are also not compared because their inputs are irreproducible and unavailable to the public. Among the studies using narratives as the knowledge source, some are not suitable for the comparison since they mainly use manual methods to acquire knowledge. With these considerations, we choose the method proposed by Liu et al. [25] to compare with.

In Liu's method, experiential knowledge is acquired from experts' discussion record. Although it involves manual annotation of discussion scenarios, the "core content" of experiential knowledge is acquired in an automatic manner using key graph algorithm. By treating the sequentially arranged information items on a user's desktop as the "discussion", we can apply Liu's method in our situation and compare the result with our method's. To quantify the comparison, we adopt the measurement of *precision and recall*. Precision and recall are the standard performance measurement of information retrieval methods. Here we deem knowledge acquisition as a special task of information retrieval – the text segments representing useful EK are the information to be retrieved from all the information on a user's desktop. Precision measures the proportion of useful EK items in all the EK items acquired by a knowledge acquisition method. Recall measures the proportion of useful EK items acquired by a knowledge acquisition method in all the useful EK items existing in the desktop information. Fig. 6 shows the precision and recall of the two evaluated methods in four experiments.

The four experiments use different persons' desktop information as the knowledge source. The desktop information is selected such that in experiment I most of the experiential knowledge exists in the form of explicit Q&A pairs, and there are no obvious Q&A pairs (featuring question marks and interrogatives) in the input information of experiment IV. The number of explicit Q&A pairs in experiment II and III is between I and IV and is increasing from II to III. This is to check how much the proposed method relies on explicit Q&A patterns to acquire knowledge. The decision on whether an acquired EK item is useful is made by the engineer who provides the input desktop information. The identification of all the useful EK in the input is determined by the engineer either. As can be seen from Fig. 6, the proposed method performs better than the key graph-based method regarding both precision and recall, even when there is no obvious Q&A patterns in the input information.

In addition to the quantitative comparison, we compare different knowledge acquisition methods regarding four qualitative factors. The four qualitative factors are: (1) support narrative knowledge source; (2) automated by computer; (3) be in time; (4) support reasoning. The first factor is important because by definition experiential knowledge is a kind of personal knowledge, and the commonest way of expressing experience for an individual is through talking and typing. The automation of knowledge acquisition is also an important factor because manual knowledge acquisition costs too much time and energy, which lowers the efficiency of knowledge acquisition and brings the problem of knowledge acquisition bottleneck. Being in time in knowledge acquisition is pursued by us as a goal in this paper because experiential knowledge is generated and utilized in a dynamic environment, which cannot be fully understood in a retrospective manner. At last, the support of reasoning should be considered when acquiring and formatting knowledge since knowledge reuse should be based on knowledge's semantics rather than words barely. As shown in Table 9, while the proposed method has all the four factors as its advantages, other existing methods do not possess the four advantages at the same time.

Due to the ambiguous definition of tacit/experiential knowledge in the literature, different knowledge acquisition methods may vary significantly from the goal to the implementation. The quantitative comparison between our method and the key

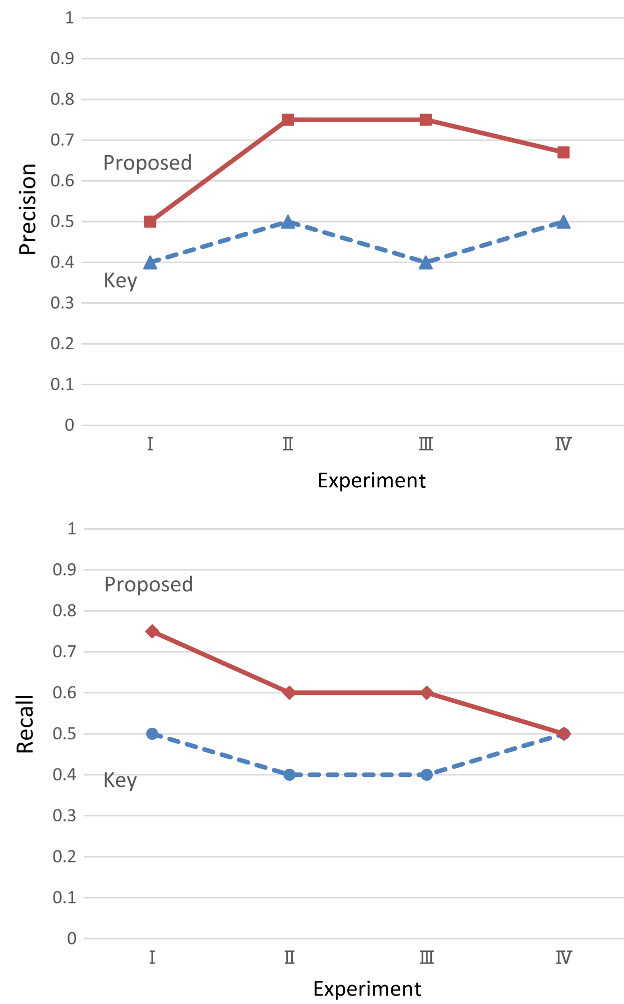

Fig. 6. Precision and recall of two EK acquisition methods.

graph-based method does not mean our method is better than Liu's method in acquiring the "empirical tacit knowledge" (ETK) described in their research. First, the definition of ETK is different from the experiential knowledge (EK) studied in this paper. While ETK emphasizes the role of associations between problem-solving cases as knowledge, EK focuses on the clarification of how an individual gets a particular problem solved. Second, both knowledge acquisition methods contain some manual aspects that cannot be reproduced in the other's situation: for Liu's method, it is the manual annotation of problem-solving scenarios, for the proposed method, it is the annotation of Q&A samples used for training the CRF model. In fact, the selection of the Q&A samples affects somewhat the performance of the proposed method, since if the training examples contains mainly explicit questions, then the trained CRF model will incline to recognize only the Q&A with a question mark

Table 9
Qualitative comparison between knowledge acquisition methods.

	Narrative knowledge source	Automated by computer	Support reasoning	In time
Ruiz (2014)	×	√	√	√
Liu (2014)	√	√	×	×
Castro-Schez (2013)	×	√	√	×
Cairo (2012)	√	×	√	×
Our method	√	√	√	√

or interrogative. In this paper, the training examples contains half explicit questions and half implicit questions (such as ''I want to . . .'' and ''I cannot''), so as to balance the ability of the model to recognize the explicit and implicit Q&A in a user's desktop information. Anyway, since the question mark and interrogatives are still used for identifying a question, the proposed method may experience some performance drop when the incoming desktop information contains few explicit questions.

5.5. Discussion

The knowledge acquired with the proposed method is categorized as experiential knowledge which bears tacitness, personalization and problem-solving orientation as the core traits. With these traits EK may seem similar with tacit knowledge, personal knowledge and know-how, but these are different knowledge types for the following reasons. First, tacit knowledge is the super class of EK and contains components such as intuition, insight and faith beside EK. Personal knowledge denotes all the knowledge owned by a person, containing both tacit knowledge and explicit knowledge. According to the definition provided by Wikipedia [48], know-how is a term for practical knowledge on how to accomplish something, and it is often tacit knowledge in the context of industrial property. In this sense know-how is very similar or even identical to the EK defined in this paper. However, know-how's role as tacit knowledge may not be admitted in other fields. Outside usage in terms of industrial property, know-how is viewed as procedural knowledge [48], and in artificial intelligence procedural knowledge is often represented as a partial or complete finite-state machine or computer program [49]. This explicit representation of procedural knowledge contradicts the understanding of know-how as tacit knowledge, so know-how is not an appropriate term for denoting the knowledge acquired in this research.

An example can be used to show the difference between explicit know-how and tacit EK, meanwhile demonstrating the feasibility of EK acquisition in areas more than FEA. Logistic optimization is an extensively research area in industrial engineering, where a typical optimization problem may be formulated as a mixed integer programming model (MIP) [50]. The formulation of a MIP considering the various objectives, constraints and conditions of a practical problem requires the EK of a researcher, as different researchers usually model the problem differently. The solving of a MIP belongs to explicit know-how, since it has been described in textbooks and there is matured software to do so. For a class of NP-hard MIPs, the time required for obtaining the global optimum of a model is too long to be of practical use. So here EK comes into play – a kind of heuristic methods using the operational rules summarized by domain practitioners and the solution searching strategies designed by domain researchers can help to obtain good enough solutions efficiently, and hence has become a prevailing practice for solving NP-hard logistic problems [50]. In this case, to obtain the rules and strategies as EK, the communication between the practitioners, researchers and the two groups can be used as the task context, to which the proposed method can be applied.

Sometimes an answer provided by a person may be copied from textbooks or other people's opinions. In this case the corresponding EK may not seem to be the ''personal'' knowledge of the answerer. For this, we argue that the situation does not pose a challenge to the ''personal'' trait of EK because knowing when to copy or ''borrow'' is also based on the experience of the answerer. If we need to build a precise incentive mechanism to reward people who contribute innovative knowledge, then knowledge audit may be carried out to investigate the exact knowledge needs, knowledge resources and knowledge flow inside an organization. Plagiarism detection can also be used to check the originality of EK. Anyway, these are beyond the scope of this paper.

Another issue that has not been fully discussed in this paper is the management of the acquired EK. EK management includes the storing, retrieving, reusing and renewing of the EK items. Currently the acquired EK items are stored as formatted texts. When reusing these EK items, they are loaded into the computer memory one at a time, meanwhile restored into the instances and relations defined in the EK ontology, and then matched with the incoming task situation according to the method described in Section 5.3. After the matching, an EK item (its instances and relations) is erased from computer memory to avoid contradictions with the next EK item. If an EK item matches with the current task situation, then it is recommended to the user for reference. When there are plenty of EK items to be checked for reuse, a fast EK retrieving method without reasoning with the full semantics of EK and task situation will be needed. This issue together with the renewing of EK will be left for future research.

6. Conclusion

Aiming at the more effective acquisition of the important yet so far ill-defined experiential knowledge, this paper studies the characterization, representation and context-aware acquisition of experiential knowledge in the computer-aided engineering domain. First, a model of experiential knowledge based on ontology and default logic is proposed, which integrates personal experience with formal domain concepts and endows it with computer operability. Then a context awareness mechanism for desktop engineering activities is developed to recognized different user tasks. Within each task, text segments matching the Q&A patterns are discovered by training a CRF model, and the extracted Q&A is mapped to the ontology to finalize the knowledge acquisition process. Featuring a high automation degree, the proposed method can run continuously on an engineer's desktop and realize in time knowledge acquisition. Such characteristics make the proposed method more efficient and less attention-diverting than the previous methods.

Nevertheless, there are some unfulfilled tasks of the proposed method. First, the experiential knowledge acquired from different individuals often shows similar content but varies largely in quality. To keep the right knowledge for reuse, it is necessary to design a strategy to assess the quality of knowledge items. Second, although the paper has provided a guideline for knowledge reuse, more knowledge management issues including knowledge storage, knowledge retrieval and knowledge renewing, need to be studied in the future.

Acknowledgements

The author is grateful to National Nature Science Foundation of China (71271133), the Research Fund for the Doctoral Program of Higher Education of China (No. 20100073110035), Innovation Program of Shanghai Municipal Education Commission (13ZZ012), and Shanghai Science & Technology Committee Research Project (15590501700) for financial support that made this research possible.

References

[1] C. Linde, Narrative and social tacit knowledge, J. Knowl. Manage. 5 (2) (2001) 160–171.

[2] E. Kwong, W.B. Lee, Knowledge elicitation in reliability management in the airline industry, J. Knowl. Manage. 13 (2) (2009) 35–48.

[3] W.M. Wang, C.F. Cheung, W.B. Lee, et al., A computational narrative construction method with applications in organizational learning of social service organizations, Expert Syst. Appl. 36 (4) (2009) 8093–8102.

[4] B. Song, Z.H. Jiang, Proactive search enabled context-sensitive knowledge supply situated in computer-aided engineering, Adv. Eng. Inform. 27 (2013) 66–75.
[5] A. Abecker, A. Bernardi, K. Hinkelmann, et al., Context-aware, proactive delivery of task-specific information: the knowmore project, Inform. Syst. Front. 2 (3) (2000) 253–276.
[6] H.J. Ahn, H.J. Lee, K. Cho, Utilizing knowledge context in virtual collaborative work, Decis. Support Syst. 39 (4) (2005) 563–582.
[7] I. Nonaka, A dynamic theory of organizational knowledge creation, Organ. Sci. 5 (1) (1994) 14–37.
[8] J.B. Noh, K.C. Lee, J.K. Kim, et al., A case-based reasoning approach to cognitive map-driven tacit knowledge management, Expert Syst. Appl. 19 (4) (2000) 249–259.
[9] E.N. Brockmann, W.P. Anthony, Tacit knowledge and strategic decision making, Group Organ. Manage. 27 (4) (2002) 436–455.
[10] K.U. Koskinen, H. Vanharanta, The role of tacit knowledge in innovation processes of small technology companies, Int. J. Prod. Econ. 80 (1) (2002) 57–64.
[11] M. Polanyi, The Tacit Dimension, Routledge and Keoan Paul, London, 1967.
[12] N. Armaghan, J. Renaud, An application of multi-criteria decision aids models for case-based reasoning, Inf. Sci. 210 (2012) 55–66.
[13] A. Azadeh, V. Ebrahimipour, P. Bavar, A fuzzy inference system for pump failure diagnosis to improve maintenance process: the case of a petrochemical industry, Expert Syst. Appl. 37 (1) (2010) 627–639.
[14] M.A. D'Eredita, C. Barreto, How does tacit knowledge proliferate? An episode-based perspective, Organ. Stud. 27 (12) (2006) 1821–1841.
[15] B.K. Foguem, T. Coudert, C. Béler, et al., Knowledge formalization in experience feedback processes: an ontology-based approach, Comput. Ind. 59 (7) (2008) 694–710.
[16] Y.J. Chen, Development of a method for ontology-based experiential knowledge representation and reasoning, Decis. Support Syst. 50 (1) (2010) 1–20.
[17] T. Gavrilova, T. Andreeva, Knowledge elicitation techniques in a knowledge management context, J. Knowl. Manage. 16 (4) (2012) 523–537.
[18] M.L. Maher, A.G.S. Garza, Case-based reasoning in design, IEEE Expert 12 (2) (1997) 34–41.
[19] D.B. Leake, D.C. Wilson, A case-based framework for interactive capture and reuse of design knowledge, Appl. Intell. 14 (1) (2001) 77–94.
[20] P.P. Ruiz, B.K. Foguem, B. Grabot, Generating knowledge in maintenance from experience feedback, Knowl.-Based Syst. 68 (2014) 4–20.
[21] J.J. Castro-Schez, J.M. Murillo, R. Miguel, et al., Knowledge acquisition based on learning of maximal structure fuzzy rules, Knowl.-Based Syst. 44 (2013) 112–120.
[22] L. Feng, T.R. Li, D. Ruan, et al., A vague-rough set approach for uncertain knowledge acquisition, Knowl.-Based Syst. 24 (6) (2011) 837–843.
[23] W.Y. Liu, K. Yue, M.H. Gao, Constructing probabilistic graphical model from predicate formulas for fusing logical and probabilistic knowledge, Inf. Sci. 181 (18) (2011) 3828–3845.
[24] O. Cairo, S. Guardati, The KAMET II methodology: knowledge acquisition, knowledge modeling and knowledge generation, Expert Syst. Appl. 39 (9) (2012) 8108–8114.
[25] L. Liu, Z. Jiang, B. Song, A novel two-stage method for acquiring engineering-oriented experiential tacit knowledge, Int. J. Prod. Res. 52 (20) (2014) 5997–6018.
[26] C. Wagner, Breaking the knowledge acquisition bottleneck through conversational knowledge management, Inform. Resour. Manage. J. 19 (1) (2006) 70–83.
[27] W.P. Cheah, Y.S. Kim, K.Y. Kim, et al., Systematic causal knowledge acquisition using FCM constructor for product design decision support, Expert Syst. Appl. 38 (12) (2011) 15316–15331.
[28] J.M. Juarez, T. Riestra, M. Campos, et al., Medical knowledge management for specific hospital departments, Expert Syst. Appl. 36 (10) (2009) 12214–12224.
[29] Y. Jin, Y. Ishino, DAKA: design activity knowledge acquisition through data-mining, Int. J. Prod. Res. 44 (14) (2006) 2813–2837.
[30] J. Wang, Y.P. Wu, X.N. Liu, et al., Knowledge acquisition method from domain text based on theme logic model and artificial neural network, Expert Syst. Appl. 37 (1) (2010) 267–275.
[31] P. Brezillon, Context dynamic and explanation in contextual graphs, in: Modeling and Using Context, Springer, 2003, pp. 94–106.
[32] P.J. Brown, J.D. Bovey, X. Chen, Context-aware applications: from the laboratory to the marketplace, IEEE Pers. Commun. 4 (5) (1997) 58–64.
[33] A.K. Dey, Understanding and using context, Pers. Ubiquit. Comput. 5 (1) (2001) 4–7.
[34] L. Zhen, G.Q. Huang, Z.H. Jiang, An inner-enterprise knowledge recommender system, Expert Syst. Appl. 37 (2) (2010) 1703–1712.
[35] J. Shen, W. Geyer, M. Muller, et al., Automatically finding and recommending resources to support knowledge workers' activities, in: Proceedings of the 13th International Conference on Intelligent User Interfaces, 2008, pp. 207–216.
[36] R. Lokaiczyk, A. Faatz, A. Beckhaus, et al., Enhancing just-in-time e-learning through machine learning on desktop context sensors, in: Modeling and Using Context, 2007, pp. 330–341.
[37] B. Schmidt, E. Godehardt, H. Paulheim, PASTREM: proactive ontology based recommendations for information workers, in: Semantic Models for Adaptive Interactive Systems, Springer, 2013, pp. 109–132.
[38] A.S. Rath, D. Devaurs, S.N. Lindstaedt, UICO: an ontology-based user interaction context model for automatic task detection on the computer desktop, in: Proceedings of the 1st Workshop on Context, Information and Ontologies, Greece, 2009, pp. 1–10.
[39] S.C.J. Lim, Y. Liu, W.B. Lee, A methodology for building a semantically annotated multi-faceted ontology for product family modelling, Adv. Eng. Inform. 25 (2) (2011) 147–161.
[40] R. Reiter, A logic for default reasoning, Artif. Intell. 13 (1) (1980) 81–132.
[41] X. Xue, J. Jeon, W.B. Croft, Retrieval models for question and answer archives, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2008, pp. 475–482.
[42] G. Salton, A. Wong, C.S. Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1975) 613–620.
[43] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, Academic Press, London, New York, Toronto, Sydney, San Francisco, 1979.
[44] S. Ding, G. Cong, C.Y. Lin, et al., Using conditional random fields to extract contexts and answers of questions from online forums, in: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, 2008, pp. 710–718.
[45] J.J. Jiang, D.W. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy, in: Proceedings of International Conference Research on Computational Linguistics, 1997.
[46] D.Q. Chen, C.D. Manning, A Fast and Accurate Dependency Parser using Neural Networks, in: Proceedings of EMNLP, 2014.
[47] B. Song, Z.H. Jiang, X.Y. Li, Modeling knowledge need awareness using the problematic situations elicited from questions and answers, Knowl.-Based Syst. 75 (2015) 173–183.
[48] https://en.wikipedia.org/wiki/Know-how.
[49] https://en.wikipedia.org/wiki/Procedural_knowledge.
[50] J. He, Y. Huang, D. Chang, Simulation-based heuristic method for container supply chain network optimization, Adv. Eng. Inform. 29 (2015) 339–354.