DISCRIMINATIVE APPROACH TO DISCOVERY IMPLICIT KNOWLEDGE

Sergiy Chalyi, Olga Kalynychenko, Sergiy Shabanov-Kushnarenko, Vira Golyan

Abstract: The process of finding implicit knowledge based on the well-known DIKW concept ("data - information - knowledge - wisdom") is considered. It has been shown that the transformation of data into information and then into knowledge is carried out by means of the implicit context (implicit knowledge). The DIKW concept is added by elements of implicit knowledge. A general algorithm to identify it is offered. Based on the extended concept of DIKW, a difference approach to identification of implicit knowledge is presented. The approach is based on a comparison of the results of the data mining process for similar processes performed at different times and finding the difference between the graphs obtained. This is considered as formalized implicit dependencies.

Keywords: process mining, implicit dependences, explicit dependences, implicit knowledge, explicit knowledge, concept "data - information - knowledge - wisdom".

ACM Classification Keywords: I.2.6 Learning - Knowledge acquisition

Introduction

Representation and use of implicit knowledge currently provides significant opportunities to improve the effectiveness of artificial intelligence systems. The lack of a well-developed concept requires an analysis of approaches to identify and use such knowledge by man directly, and subsequently, the formation of models of representation, allocation of use of implicit knowledge.

The problem of obtaining implicit knowledge (inseparable from a human being) is viewed not only in the field of artificial intelligence, but also in the philosophical, psychological, economic studies. The importance of this problem is connected with the key features of implicit knowledge:

- Inseparability from a human being (knowledge allows to get results without awareness of gaining the result);
- Tight integration with explicit knowledge.

Explicit, formalized knowledge is usually based on a implicit context [Goodman, 2003]. The latter is understood by a person, but its finding is associated with considerable difficulties. Consequently, the completeness of the knowledge description in artificial intelligence systems is achieved by finding of both of explicit and implicit knowledge components that determines the relevance of this study.

The problem of implicit knowledge application in philosophical, psychological, economic studies

Before holding the detailed discussion of the problem and the concept of implicit knowledge finding, it is useful to consider various aspects of the interaction of explicit and implicit components in terms of philosophical, psychological, and economic positions. This consideration makes the case for the importance and the ability to reveal knowledge on the basis of the data sets analysis.
The problem of personal knowledge was discussed in philosophical works by M. Polanyi [Goodman, 2003], [Tandem Computers Inc, 1996]. He singled out this kind of knowledge as man's inherent advantage over animals. He believed proficiency to be the main characteristic of this knowledge. When finding man's knowledge we may consider separately "what-knowledge" and "how-knowledge." The first of these is selected and formalized, and the second is implied (in fact, it is implicit) [Goodman, 2003]. In other words, "what-knowledge" usually can be easily explained and documented by man. Knowledge of the second type is characterized by the fact that one can see only the result, and the reasons of its achievement are very difficult to explain. For example, a person can easily solve the problem of face recognition, but he usually finds it difficult to explain his internal recognition algorithm.

Thus, from a philosophical point of view we can highlight the importance of the implicit component, i.e. the "how-knowledge" component (Fig. 1).

In the field of psychology of intelligence, human intelligence is considered in relation to his knowledge. Simple and hierarchical models are used to describe knowledge in the given area.

In a simple one-layer model intelligence reflects the current level of psychological development of an individual, and it is expressed through a variety of psychological manifestations [Kholodnaya, 2002]. Knowledge in such a model reflects intelligence functionality and can be represented by a set of factors taking into account the types of mental activities, the object with which mental actions have been performed and the final result of the actions [Ngai, 2009]. This simplified model gives an opportunity to get a single set of explicit and implicit dependencies in human knowledge.

Further details of natural intelligence and knowledge are the three-layer hierarchical models. In the papers [Konar, 2000], [Ian, 2011] the following layers are identified as general intelligence; general intelligence capacities; special capacities of a human being. Note that the second layer in this model reflects verbal, numerical, spatial intelligence capacities, and the third layer reflects professional capacities of a human being, namely: algorithmic, technical thinking, math skills, etc. As one can see from the structure of the model, knowledge, in this case, is distributed into layers as follows: the upper layer mostly presents implicit knowledge, the lower layer mostly presents explicit and subject to formalizing knowledge.

The three-layer hierarchical model of natural intelligence [Bondarenko, 2011] offers a different distribution of knowledge at the second layer setting verbal and nonverbal abilities of a human being. At the third layer we
distinguish human intellectual capacities which use explicit and implicit knowledge in separate areas of activity. Thus, the current intelligence models specify integral use of explicit and implicit knowledge, the advantage of an explicit (implicit) component being determined by the layer of the model hierarchy, as shown in Fig. 2.

Fig. 2. Distribution of implicit knowledge by using a three-layer model of natural intelligence

In today's economy, knowledge has become more important than the traditional factors of production. The complex of explicit and implicit knowledge forms knowledge of an organization (organizational knowledge). The latter provides enterprise activity both due to document forms of representation, and the knowledge, experience and skills of the employees.

For the first time, the term "organizational knowledge" was proposed by I. Nonaka and H. Takeuchi [Tsuchiya, 1993] in studying the origin and development of innovation in Japanese companies. The authors view organizational knowledge as the knowledge which integrates the totality of knowledge, staff experience at the organizational level as a whole. The knowledge of an organization comprises both formalized corporate knowledge about its functioning, and the implicit knowledge of individual employees (Fig. 3).

Fig. 3. The interaction of explicit and implicit knowledge: the economic aspect
Organizational knowledge includes the following components inherent to a human-being:

- Competence of performers which includes education, skills, experience and practical skills to perform employment duties;
- Labor intellectual assets as "total intelligence" of the company employees which are determined by education and qualifications and also depend on prior intellectual activities of employees;
- A common corporate culture that embraces non-formalized set of knowledge on how to perform business processes and interaction between employees;
- Communication resources that cover the knowledge and experience to organize relations with the counterparty of the company.

The explicit component of organizational knowledge in particular includes:

- Formalized description of business processes of the company as well as the general management culture;
- Intellectual property of the company;
- Knowledge resources about managerial, financial, scientific and legal production technologies being used;
- Information resources which record the results of knowledge application in daily activities of the organization.

Thus, the research in the field of economics emphasizes the importance of integrated use of explicit and implicit knowledge, justifying their continuous interaction in the process of economic activity. Moreover, information resources, as components of the explicit knowledge, may reflect the results of using implicit knowledge in the management of the organization.

The conducted analysis of explicit and implicit knowledge in the philosophical, psychological, and economic aspects reveals the following features of their use and interaction (Fig. 4):

- By using explicit knowledge one can highlight a practical result and also to explain, write down and formalize the used regularities;
- By using implicit knowledge one can only get a practical result. To explain the way it is obtained is usually difficult;
- Information about the activity (of an individual, organization) contains "traces" of the application of both explicit and implicit knowledge.

![Diagram of explicit and implicit knowledge](image-url)

Fig. 4. Comparison of the key features of explicit and implicit knowledge application
The given features of implicit knowledge suggest that implicit knowledge can be duplicated traditionally only by means of informal methods (communication, learning combined with the accumulation of human experience), which makes its use difficult in artificial intelligence systems. It is also important to note that the lack of inter-relationships and the presence of only the external manifestations also complicate the formalization of such knowledge.

At the same time it is possible to search for both explicit and implicit knowledge based on information files (data sets), resulting from the application of relevant knowledge.

Separation of explicit knowledge in this case is carried out by data-, process mining techniques. Separation of implicit knowledge by analyzing data sets requires further investigation.

The above mentioned reasoning specifies the relevance of search and formalization of implicit knowledge for future application in artificial intelligence systems.

**Problem setting**

The above key features of implicit knowledge illustrate the difficulties of its search and formalization in the general case due to the influence of the human factor. At the same time it is possible to extract and formalize the hidden knowledge from the patterned array of data in the problems of data-, process- and web-mining.

This possibility is based on the effect of implicit knowledge on the formation of the dependencies resulting from intelligent analysis. Such dependencies can be displayed in the form of graphs and, in fact, represent explicit knowledge of the processes (structured objects) of the subject domain.

However, the resulting knowledge of identical processes or objects, based on the analysis of data sets during different time intervals, differ in many cases. The reason for these differences is largely determined by the use of formalized, hidden knowledge while executing initial processes and forming corresponding data sets.

These considerations testify that, in principle, it is possible to identify implicit dependencies in the analysis of structured objects (processes) resulting from the data set research.

All this shows the importance of model development problems for finding implicit knowledge based on the analysis of relevant data sets.

**DIKW concept as the basic scheme for discovering implicit knowledge**

Before further consideration of the problem of discovering implicit knowledge it is necessary to analyze the overall sequence of person’s work with knowledge. This sequence is based on the alternating use of explicit and implicit knowledge. In the learning process there occurs knowledge transformation from one form into another.

Conversion of explicit knowledge in a implicit form is performed by training. In this process, the implicit and explicit knowledge components are initial ones. Explicit knowledge is presented as well-known strategies, technologies and documented materials. Implicit knowledge is owned by people who teach and who are trained. Those who teach help "absorb" presented material; integrate it into the world view existing in student’s mind. As a result, explicit knowledge is converted into skills, abilities and experience. Implicit context is transferred by an individual who teaches and provides (facilitates) learning.
Conversion of knowledge from the explicit form into a implicit one is performed by means of finding and subsequent documenting of the implicit component. At the same time additional explicit knowledge (e.g. formal documentation rules) is used.

Further expansion of the transformation sequence of explicit knowledge into a implicit form and vice versa leads us to the well-known DIKW concept "data - information - knowledge - wisdom [Tsuchiya, 1993]."

In this paper this model is of interest due to the fact that in the process of transformation between the levels of the model implicit knowledge is used. Formalization of such transformations creates conditions for the use of implicit knowledge in artificial intelligence systems.

Let us briefly examine the levels of the given model, adapting them with regard to the peculiarity of the problem to reveal implicit knowledge.

The first level of the model being considered covers the original data sets, which are then converted into information and knowledge. The main features of the given data set are as follows:

- It is a direct result of observations;
- It covers a set of arbitrary symbols whose meaning at this level is not considered;
- It usually has a specific form of representation, so the conversion of this form for further use may be needed.

When considering the problems posed in this paper concerning the use of implicit knowledge in artificial intelligence systems as the first-level data $\{p_1, p_2, ..., p_n\}, i = 1, n$ of the DIKW model, it is advisable to use databases and structured text files.

The second level – the information level has the following differences from the level of data:

- The links between the data that determine the value of a data set and allow drawing conclusions about the data available are given;
- The possibility of using the data in the current level is not determined.

When using predicate models, the relationship between the data is given in the form of a predicate $I(p_1, p_2, ..., p_n)$ that determines the structure of information. This predicate can be mapped to a system of binary predicates $l_i$, $i = 1, m$, which is represented as a relational network for parallel processing [Mitra, 2003].

The third level - the level of knowledge - has the following features:

- Knowledge integrates information, and it is practically useful.
- Knowledge is presented in the form of individual interconnections, integration of knowledge and the creation of new knowledge at this level is not considered.

Knowledge at this level can be represented as predicates that define the relationship between the elements of previous levels.

The level of wisdom allows us to find fundamentally new understanding of the existing knowledge. From the standpoint of artificial intelligence the level of meta-knowledge can be represented as a second-order predicate (a predicate from predicates) $M(l_1, l_2, ..., l_k)$, given at the set $\{l_i, i = 1, k\}$.

In some papers meta-knowledge and wisdom are separated; in this case a hierarchy of 5 levels is formed. However, the level of wisdom is inherent to human intellect exclusively. At this level human intelligence operates with abstract values differentiating, for example, between good and evil, bad and kind.

Therefore, in accordance with the problem being solved, we single out 4 levels, and the last level is considered as the level of meta-knowledge.
In the DIKW model we identify a level of understanding as a process of creating new knowledge out of the existing one. The key function of understanding is a function of teaching new knowledge. Understanding allows us not only to create new knowledge, but also to apply this knowledge to perform useful (in the sense of achieving the expected results) actions. As we have seen previously, understanding requires the use of explicit and implicit knowledge. Therefore, understanding is based on the levels of knowledge, information, data, and uses implicit knowledge in the process of transition from one level to another (e.g. the implicit context).

Thus, the DIKW concept expands the previously discussed process of work with explicit and implicit knowledge and, consequently, allows us to pass to the modeling of finding implicit knowledge.

All the above reasoning shows the importance of identifying the role of implicit knowledge in the hierarchy of knowledge, and also requires the formalization of impact of hidden knowledge in the concept of "data - information - knowledge - meta-knowledge". Building such a formal model allows us to justify finding implicit knowledge in general and implicit dependencies in particular based on the analysis of structured data sets.

An important feature of knowledge, which is used in natural intelligence, is to represent knowledge as a process, as opposed to being an object of knowledge in artificial intelligence systems. Interrelation of data, information and knowledge in explicit and implicit forms define a process aspect of knowledge in accordance with the DIKW concept (Fig. 5). Knowledge in the process approach is not only the object of use, but it also can play an active role.

![Diagram](image-url)

**Fig. 5. The DIKW concept and use of implicit knowledge**

Thus, the source of knowledge represents its implicit knowledge in a structured way, in the form of information. In the process of structuring, explicit knowledge is used - for example, about the required structure and the form of presenting information. The information obtained contains knowledge in the hidden
form. It can be transformed into knowledge only after its interpretation by using additional knowledge, both implicit and explicit.

Then knowledge is transferred as a data set. Indeed, any documented rules, formulas, texts, diagrams, etc. are simply a collection of characters as long as their interpretation is made. Knowledge receiver performs interpretation of links between data, getting information, then the interpretation of templates obtaining implicit knowledge. In the process of interpretation at this stage a procedural aspect of knowledge is commonly used.

Transformation of knowledge from the implicit into the explicit form is performed by means formalizing previously obtained templates at the last (optional) stage. The given stage completes the transfer of knowledge in natural intelligence. The explicit knowledge obtained can be further used in AI systems.

Note that in accordance with the DIKW concept the patterned array of data (object) corresponds to the information level.

The last key property of implicit knowledge is that it is directly related to performing various actions, such as in technological processes, business processes, design processes, etc. Unlike explicit knowledge, which is separated from an individual and, therefore, can be studied, modified and used at any given time, implicit knowledge manifests itself when fixing operations as part of the above processes, namely in the form of structured data sets.

The important feature of the considered process of knowledge transfer is as follows: for two people to be able to transfer implicit knowledge to each other, they must have collective knowledge (both explicit and implicit). This means that their structuring and interpretation knowledge systems should correspond to each other [Mitra, 2003]. Thus, implicit dependencies can be found by a system of artificial intelligence based on the interpretation of structured data sets under the following conditions:

- Interpretation technique is consistent with the technique, by means of which implicit knowledge has been transformed into a structured data object;
- There is a predefined set of explicit (formalized) dependencies, which were derived from structured data sets.

These features of implicit knowledge lead to the following interim conclusions:

- In natural intelligence explicit and implicit knowledge are complementary. Implicit knowledge provides the formation of explicit knowledge in data and information processing;
- Implicit knowledge is not perceived by man and can be obtained only on the basis of the analysis of actions related to it;
- To discover implicit knowledge it is necessary to use general explicit knowledge concerning the subject domain.

**Discriminant model of implicit knowledge finding**

The considered pattern of extracting implicit knowledge in terms of the DIKW concept is the basis for the model of finding implicit knowledge from data sets based on process mining results.

The basis of this model is the process representation of knowledge, which describes knowledge as a process and is characterized by the following features:

1) Explicit knowledge representation as a graph that shows a sequence of related activities and events in the subject area.
2) Invariant representation, which means the possibility of its various imaging while maintaining a predetermined interrelation between vertices and arcs.

3) The ability to integrate knowledge presented in the form of individual processes, since a group of similar graphs covers explicit knowledge about a class of similar objects.

4) The ability to verify the completeness and consistency of knowledge.

The third feature among the considered properties of process representation of knowledge provides the ability to discover implicit knowledge according to the DIKW concept by performing the following steps (Fig. 6):

1) Formation of a set of standardized processes for a given subject domain \( \{P_j\} \), represented as a set of vertices \( V_j \) and arcs \( E_j \). Graph vertices of the process reflect its activity and, therefore, in addition to its \( N_j \) identities they also can be characterized by a set of \( A_j \) attributes: \( P_j = \{V_j, E_j\}, V_j = \{N_j, A_j\} \). The given set displays explicit knowledge about the \( P_j \) processes (or about structured objects in general) and is usually available directly as a result of initial process development. Consequently, this step corresponds to the first three levels of the DIKW concept.

2) Gaining knowledge of this or similar processes based on the analysis of data arrays (process log) by means of process mining techniques in the form of the \( \{P_{jk}\} \) set. As a result of executing this step for each process \( P_j \) we obtain \( k \) – graphs reflecting the knowledge of each of its \( k \) – implementations, at that \( P_{jk} \) graphs can "slightly" differ from standardized, specified in the first step. These differences may be caused either by incomplete particular implementation as compared to the original model or by dependencies that are not reflected in the graphs of processes and, therefore, are implicit. In general, such implicit dependencies do not allow us to get isomorphic graphs.

3) Finding implicit dependencies. At this step it is necessary to reveal and formalize the differences between the graphs describing identical processes or objects. These differences may represent a formalized part of implicit dependencies. This step corresponds to the expanded DIKW concept shown in Fig. 5. Finding these differences will be formalized later in the discriminant model.

4) Adding a formalized implicit component obtained in the previous step to the process model of knowledge representation.

Original data in the process of finding implicit knowledge in accordance with the above algorithm are the original model of the process \( P_j \) and its log \( L_j \). The process consists of a set of activities whose performance is fixed in the form of time-bound events in the process log. The process log consist of a set of traces of these activities: \( L_j = \{S_{jk}\} \). Each \( S_{jk} \) trace corresponds to one-shot execution of the \( j \) – process and represents a set of events reflecting the sequence of the performed \( s'_{jk} \in S_{jk} \) activities. Thus, the process log integrates all of its documented implementations. The differences between the original and the final models of the process built on the basis of the analysis of each of the process log traces can be divided according to the following classification criteria:
Fig. 6. The general pattern of finding implicit knowledge from data sets

the original model of the \( P_j \) process contains \( s'_j \) activities that are missing in the final \( P_k \) model obtained as a result of the analysis of the \( k \)-trace:

\[
\exists \{ s'_j \} : (s'_j \in P_j) \land (s'_j \notin P_k),
\]

(1)

the final \( P_k \) model contains activities that are missing in the original \( P_j \) model:

\[
\exists \{ s'_j \} : (s'_j \notin P_j) \land (s'_j \in P_k),
\]

(2)

In the first case, we can say that in the process of specific implementation not all the opportunities available in the original model have been used.

The second case shows the incompleteness of the original model, which may be associated with the use of implicit knowledge in the process implementation.

In accordance with the criterion (2) for the implementation of the considered sequence of steps it is proposed to use the discriminant model of finding implicit knowledge based on pair-wise comparison of \( P_j \) and \( P_k \) graphs for \( k \)-execution of the \( j \)-process.

Note that the discrepancy between \( P_j \) and \( P_k \) graphs represents the information level in the DIKW model.

Indeed, the differences between the graphs are represented by the subgraph that contains a set of vertices and arcs connecting them which are available in the \( P_k \) graph and are missing in the \( P_j \) graph. Thus, in the general case, this subgraph contains only a fragment of the original process in the form of structured
information. But we are interested in implicit knowledge resulting in such a difference between the graphs. In accordance with the concept discussed above, implicit knowledge is a process that leads to a change in the structure of information. Therefore, to formalize this knowledge it is proposed to find a sequence of transformations of the $P_j$ graph into the $P_\alpha$ graph, if the condition (2) is fulfilled. Then the model of implicit knowledge, which is identified on the basis of differences between the original and final models, is based on the use of the $Add$ operator, which adds the missing vertices to the $P_j$ graph, as well as arcs connected by the given vertices.

Let us define the $Add$ operator as follows:

$$Add(s'_j, s'_s, s'_P) \iff \forall s'_j, (s'_j \not\in P_j) \land (s'_s \in P)$$

(3)

The model of finding implicit knowledge is based on the cyclic elimination of the discrepancy between the original and final graphs of processes.

$$M^j_\alpha = Add(s'_j, s'_s, s'_P) \iff (s'_j \not\in P_j) \land (s'_s \in P_\alpha)$$

(4)

The application result is represented as the process of changing the original $P_j$ graph that reflects the procedural nature of implicit knowledge.

**Knowledge representation and the use of implicit dependencies**

As it has been shown previously, the use of obtained implicit dependencies is based on the representation of knowledge in the form of graphs of processes. Formalized implicit dependencies complement the graph with new vertices and arcs, reflecting previously hidden cause-and-effect relationships between the individual elements of the process.

As a whole, the process representation of knowledge is the development of script representation. Such representation contains a sequence of frames describing a stereotyped sequence of events taking into account the context and is a way to represent procedural knowledge [Mitra, 2003]. Process representation of knowledge allows us to combine procedural and declarative knowledge and has the following features:

- Knowledge is presented as a set of graphs reflecting possible sequences of tasks (actions), as well as the events that are required to perform these tasks;
- Knowledge integration in the form of processes carried out by means of uniting on the basis of common events;
- Derivation restrictions in the form of additional rules are used.

Thus, in this case, knowledge representation has a two-layer structure combining a flexible scheme of interaction between the fragments of knowledge in the form of processes and a relatively rigid structure of processes (Fig. 7).

At the level of knowledge representation in the form of processes, it is necessary to take into account the starting event of the process, one or more tolerable end events and, probably, the process priority, as well as the rules - restrictions for the process.
At the process level we consider the graph of the process describing the set of its tasks as well as the sequence of their execution. Existence of implicit dependencies is possible at this level of knowledge representation. They are not represented in the process model prior to their identification, which is why there arises a problem of the model adequacy. After finding and formalizing implicit dependencies, they become explicit, form part of the model and do not differ from the previously considered explicit knowledge.

Conclusions

The performed analysis of implicit knowledge has shown that hidden knowledge is characterized only by external manifestations in the process of its application, and, therefore, it is usually identified and copied only with the participation of an individual. This fact complicates its use in artificial intelligence systems.

An approach to finding implicit knowledge based on the DIKW concept is offered. This concept focuses on the use of implicit context when converting data into information and knowledge that has allowed developing a generalized algorithm for finding implicit knowledge using a difference approach to its identification. The approach is based on finding the difference between the original model of the process and the end model, obtained as the result of the process log analysis by means of process mining. If the original model is incomplete compared to the end model, a process that complements the original model is formed, this process can reflect implicit knowledge in a formalized manner.

Bibliography


Authors’ Information

Sergiy Chalyi − D.Sc, professor of Information Control Systems department, Kharkov National University of Radio Electronics Ukraine; Lenin av., 14, 61166 Kharkov, Ukraine; e-mail: s_chalyi@mail.ru

Sergiy Shabanov-Kushnarenko − D.Sc, professor of Software Engineering department, Kharkov National University of Radio Electronics Ukraine; Lenin av., 14, 61166 Kharkov, Ukraine; e-mail: sergabaev@mail.ru

Olga Kalynychenko - PhD, associate professor of Software Engineering department, Kharkov National University of Radio Electronics Ukraine; Lenin av., 14, 61166 Kharkov, Ukraine; e-mail: okalinichenko@mail.ru

Vira Golyan - PhD, Software Engineering department, Kharkov National University of Radio Electronics Ukraine; Lenin av., 14, 61166 Kharkov, Ukraine; e-mail: veragolyan@yandex.ru