Knowledge Based Concept Analysis Method using Concept Maps and UML: Security Notion Case

Miquel Colobran, and Josep M. Basart

Abstract—One of the most ancient humankind concerns is knowledge formalization i.e. what a concept is. Concept Analysis, a branch of analytical philosophy, relies on the purpose of decompose the elements, relations and meanings of a concept. This paper aims at presenting a method to make a concept analysis obtaining a knowledge representation suitable to be processed by a computer system using either object-oriented or ontology technologies. Security notion is, usually, known as a set of different concepts related to “some kind of protection”. Our method concludes that a Security notion is, usually, known as a set of different concepts associated with “some kind of protection”. Our method concludes that a Security notion is a set of different concepts associated with “some kind of protection”.

Keywords—Concept analysis, Knowledge representation, Security, UML.

I. INTRODUCTION

FORMALIZING knowledge is an ancient problem. In the fourth century BC Aristotle included logic in his philosophical system and then the concept was understood as the intellectual representation of an object. The Aristotelian logic remained almost unchanged until the sixteenth century with the work of Leibniz [1] who began to include symbolic notation in logic. In the early nineteenth century, through the work of authors such as Boole [2], logic is related to mathematics through a mathematical system for modeling logical operations and accordingly a concept is a set of logic notions together with a set of rules. The acquisition of concepts has been a topic of study in psychology [3] and even recently, some computer science works focus on the concept notion [4].

Concept Analysis, a branch of analytical philosophy, aims at decomposing the elements, relations and meanings that compose a concept. There are several methods such as the Wilson’s method [5], the Rodgers evolutionary method [6] or the Walker and Avant model [7]. Obtaining the characteristics of a concept is similar to requirement gathering or knowledge elicitation used in Computer Science. Our concept analysis is made with knowledge acquisition with constrained located into the knowledge domain and the knowledge sources. The former is reduced to a concept and the latter appears because of the difficulty to reach experts in the proposed domain.

If Concept Analysis techniques [5],[6],[7] are designed to have a clear and accurate definition of the concept under study. Usually, a concept is taken from a set of sources and, by means of several steps, how it operates and which relations it has with other concepts is revealed. The goal is to obtain a better understanding of the concept. Those techniques are particularly valuable when a concept has more than one meaning. The methods can vary according to the number of steps or the sources used. Some of them are language based and others literature based. The outcome is a language based description. Those methods are stepwise, and any enlargement of the concept or source later made implies redoing the whole analysis. Besides the methods are not suitable to be used in any computational system because they are not formal. There is neither model nor relationship between the elements and there is no detail on the constituents of the elements.

The proposed approach is a 7 step incremental and literature based method aiming at obtaining an outcome suitable to be used in object oriented engineering or ontology technologies. The objective is achieved by means of knowledge elicitation and visual modeling techniques. Knowledge elicitation is used to extract the relevant parts of text related to the concept under study. Concept maps help us to graphically represent the requirements and the Unified Model Language (UML) allows us to show graphically the elements and relations underlying the concept. The outcome reveals the attributes (the value) and behavior (as with what other concepts is related) of these concepts. The resulting graphic (a class diagram) shows these elements. Using UML as a knowledge representation language, further implementation is facilitated. Furthermore, the fact of being incremental allows the enlargement of the model adding new sources, with no need of redoing the former analysis.

This paper is organized as follows: Firstly, an overview of the techniques used to develop the method are introduced briefly after the Introduction. Secondly, the Knowledge Based Concept Analysis (KBCA) method is presented. Thirdly, a case study with the security notion.

II. TECHNIQUES OVERVIEW

Several techniques, briefly described, are used in order to obtain the proposed method (Fig: 1).
A. Concept Analysis

Concepts are multifaceted, abstract representations of reality [8]. Because of that, the concept analysis deals with its vagueness, ambiguity and context in order to clarify its meaning. The formal theories of concepts tries to systematize the way a concept is described such as Unified Concept Theory [4] or Formal Concept Analysis [9].

Concepts, under the view of knowledge, are a cognitive unit of meaning sometimes defined as a “unit of knowledge” (concept describes an abstract idea). The mental concepts describe a class or category. The grouping process is done by relating the aspects and qualities common to many objects. The set of all concepts gives us a representation of the world. Thus, a search for methods to handling concepts, the elements of concepts and the relations among concepts, is inevitable. Most of the used methods, such as Conceptual Maps [10], Formal Concept Analysis [11], Object Oriented techniques [12],[13],[14] or Knowledge engineering techniques [15] begin at their first stages with some sort of analysis. This analysis, in the knowledge field, is called knowledge acquisition.

B. Knowledge Acquisition

Knowledge acquisition is the process of achieving knowledge from a human expert or a group of experts [16]. The goal in knowledge engineering is the creation of knowledge-based systems (KBS).

Under the view of computer science, knowledge acquisition is a step in knowledge engineering. Broadly speaking, the knowledge life cycle include acquisition, design and implementation. Thus, software engineering and software knowledge have common points [15].

Despite there is a range of knowledge acquisition techniques [17], they deal with particular problems. Getting knowledge is made with informal methods such as interviews, questionnaires or unstructured sources, usually but not necessarily, in text form. Communication appears as a big trouble because experts and knowledge engineers have poor understanding of each other’s knowledge area. Usually, experts are not able to express exactly the knowledge and therefore it is difficult to get an overview of the problem to be solved. Besides, this process has a big quantity of informal knowledge which needs to be classified, organized and formalized somehow. In short, the expert has no knowledge on knowledge engineering and the knowledge engineer has poor knowledge on expert knowledge areas.

Knowledge engineer faces other problems, such as Knowledge validation and Knowledge representation. The former is the way to verify if the knowledge is right understood and the latter the way the knowledge is expressed in order to be used to implement the system. To lighten the problem, as in software engineering, Knowledge elicitation requires tools in order to manage requirements. Usually the tools could be a simple spreadsheet, a database or requirement management system.

Mainly, obtaining requirements or knowledge is based on natural language and this presents unique difficulties [18], [19]. Many of the activities involved are cognitive and require creativity as well as knowledge about information technologies and the application domain. Several tools have appeared in order to ease the problem and try eliminating ambiguities. These try to somehow make an interpretation of natural language in order to apply the heuristics [20]. The purpose of some of these tools is to intend to minimize as much as possible the analyst’s personal influence. These are still leaving the final decision of construction schemes in the analyst hands.

Thus, good knowledge gathering relies on the ability of analysts to interpret the model expressed by the user and then be able to express it in a formalized form. In software engineering, Abbot [21] first proposed a technique to gather requirements from texts. One of the big advantages to work with natural language is that it forces the developer to work on the vocabulary and space of the problem. Knowledge gathering stage, thus, is not rigorous because the natural language is ambiguous.

The final result of the knowledge elicitation step is a requirements knowledge representation. It needs to be simple to understand and formal enough to be used as the input of the knowledge implementation stage. That could be achieved by means of knowledge representation and modeling languages.

C. Mind Maps and Concept Maps

A task of concept classification somehow could be achieved using automated tools. Visual classification tools make it easy to classify objects and organize concepts. Currently, different categories of visual tools can be found [22],[23] such as Mind maps, Conceptual diagrams, Visual metaphors, Tree Maps, Flow Maps or Compare and Contrast Maps.

The most well known techniques are concept maps and mind maps. Concept maps are a way to visualize the mental “map” of concepts and their relationships, as well as the structure and hierarchy of these relationships. One important aspect of concept maps is their ability to show large amounts of information in a compact format. In this context, a concept is defined as “a perceived regularity in events or objects, or records of events or objects, designated by a label” [10].

Mind Mapping is a popular related technique devised by Tony Buzan. He describes mind maps as a net starting with a

Fig. 1 KBCA Technologies involved
central word or concept and “around the central word you draw the 5 to 10 main ideas that relate to that word. You then take each of those child words and again draw the 5 to 10 main ideas that relate to each of those words” [24].

D. UML as a Knowledge Representation Language

Visual modeling started in Object Oriented software development methodologies and different methodologies for modeling have existed. But with no doubt, the Unified Modeling Language (UML) closed the discussion.

UML is the modeling language for software systems most well known and used today and is a de facto industry standard approved by the OMG (Object Management Group). UML is a set of specifications for object-oriented notation, which are composed of different diagrams that represent different stages of a software project development. UML combines techniques from data modeling, object modeling and component modeling. It can be used with all processes, along the software development life cycle. UML has synthesized the notations of the Booch method [25], the Object-modeling technique (OMT) [26] and Object-oriented software engineering (OOSE) [27] by fusing them into a single, common and widely usable modeling language.

UML is a graphical language for visualizing, specifying, constructing and documenting a system. The language focuses on the representation of a system and tells us how to create and read the models. However, nothing is said about how to create them. The latter is the goal of development methodologies. Some pros of UML could be found in [28]. The UML model consists of three classes of construction block, elements, relationships and diagrams. Elements are abstractions of real or fictitious things such as objects or actions. Relationships are the way how elements relate to each other. Diagrams reflect collections of elements along with their relationships.

The class diagram exhibits a set of classes, interfaces and relationships. This is the most common diagram in describing the design of object-oriented systems. In order to properly represent a system, UML offers a wide variety of diagrams to visualize the system from several perspectives and UML 2.0 includes 13 types of diagrams. As the aim of UML is to model any type of systems, not just software, it is also used as a knowledge representation language and the construction of ontologies [29], [30], [31].

III. KBCA

Our proposal uses together knowledge elicitation, concept maps and UML in order to produce a graphical representation of a concept. Knowledge elicitation, with constrains, is used for requirements gathering; concept maps are used to produce a graphical representation of the requirements and UML is used to draw the final outcome. The method is named as Knowledge Based Concept Analysis (KBCA) of a concept.

A. Concept Analysis and Knowledge Acquisition

Restrictions

In concept analysis the work is focused on a previously agreed concept. KBS work is focused on the domain defined at the beginning of the life cycle. Concept analysis ends when the concept is fully described and Knowledge engineering ends when the computational system is constructed. Knowledge engineering life cycle includes an analysis phase, but also has the design and implementation stages. In order to move closer concept analysis and knowledge engineering, the following points need to be considered.

- Knowledge engineering could fit purposes other than creating a computational system.
- Knowledge engineering life cycle involves several steps. Using the ones related to analysis and design, a knowledge model is obtained.
- Knowledge engineering domain is extremely flexible and could be as small as a concept.

In knowledge engineering, if the implementation stage is not done just a knowledge model of the domain is obtained. If the domain is a concept, the analysis and design stages will be focused just on that concept, its attributes and its relations. The result will be the knowledge model of a concept and become a type of concept analysis.

Another restriction is needed. When dealing with a concept, reaching the experts could be difficult or even not possible. Let’s suppose a work focused on the Newton’s concept of law of universal gravitation or the Descartes concept of mathematics. The concept description should be described on the basis of their writings or the interpretation of these concepts from other people. Thus, the best sources we can achieve are documents.

B. Method in Detail

KBCA consists of seven main steps as shown in Table I and Fig. 2. First three steps belong to the knowledge requirement gathering phase, fourth and fifth steps are the categorization (ordering phase). Sixth step makes the map of ideas/concepts/notions collected, and the last one converts the concept map into a class diagram.
Fig. 2 KBCA Method

The relevant elements and relations are detected in the second step. This is made by emphasizing the important text pieces. That could be made in ways such as changing the text color in a word processor as well as underlining it. From now these chunks of text are called key text elements (KT). That could be done in parallel to populate the list or database (step 3). For clarity purposes this has been separated into two steps. This step is the most critical part of KBCA. The rest of steps rely on this one, because the final outcome heavily depends on this one being properly taken. Thus, It implies some kind of subjective component.

The third step implies collecting the information gathered into a list. That list could be made as plain text, spreadsheet, database or even a Requirement Management Software. A simple document could create about 150 key texts. Thus, a kind of mechanical tool is highly recommended. The minimal needed fields are:

- Key text number
- A way of connecting somehow the key text to the document. That could be done in many ways such as writing the key number to the document or even copying the key text into the list.
- The category

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Choose knowledge source</td>
</tr>
<tr>
<td>Step 2</td>
<td>Extract key text elements</td>
</tr>
<tr>
<td>Step 3</td>
<td>Collect and number</td>
</tr>
<tr>
<td>Step 4</td>
<td>Categorize</td>
</tr>
<tr>
<td>Step 5</td>
<td>Assign</td>
</tr>
<tr>
<td>Step 6</td>
<td>Create concept map</td>
</tr>
<tr>
<td>Step 7</td>
<td>Create class diagram</td>
</tr>
</tbody>
</table>

In the fourth step, after listing, the category list is created. The possibility the category list is known from the beginning exists. Thus, the list could be created any step before. Once again, for clarity purposes, has been placed in that position.

Every requirement matches a category. This is the fifth step. The assignment is needed in order to detect extra or redundant key text. The following steps use the resulting list.

Step 6 is the most difficult part. Any key text is represented into a conceptual map. A number of situations may appear such as two similar key texts that have no relation at all or the coincidence of two key text which reveal that there is no need of rewriting. The rules to operate with key texts could be summarized as shown in table II.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>KT_i ≠ KT_j  ∀ j</td>
<td>A new element</td>
<td>Add to the graphic</td>
</tr>
<tr>
<td>KT_i ⊇ KT_j</td>
<td>Includes</td>
<td>No change</td>
</tr>
<tr>
<td>KT_i ⊆ KT_j</td>
<td>Included</td>
<td>No change</td>
</tr>
<tr>
<td>KT_i = KT_j</td>
<td>Same</td>
<td>No change</td>
</tr>
<tr>
<td>KT. enlarge KT_j</td>
<td>The key text could add some aspect related to the previous key text</td>
<td>Add</td>
</tr>
<tr>
<td>KT_j</td>
<td>Is a relation, not an element</td>
<td>Add</td>
</tr>
</tbody>
</table>

As a result, a concept map diagram is obtained. Thus, there is a bunch of ideas and relations spread on the map.

In the seventh and last step, a conversion of the key text extraction into a more formal graphical language is made, UML. The final result is clearly understood because a standard methodology has been used.

The final diagram needs to express elements or ideas, relations and cardinality.

It is highly recommended to add some extra information at the bottom of the graphics. The reason relies on the fact that some key text elements determine values of the attributes. Thus, the list of the known values is added.
IV. CASE STUDY OF KBCA

Now, the method is applied to security concept. Barry Buzan stated that “security is a underdeveloped concept” [32] in 1983. From then, there is a plethora of work focused on what the security concept is (structure, elements, relations).

A. Applying KBCA Steps

1) Choose
The article chosen is “The concept of security” by Peter Digerer [33]. The article, unpublished but used with author permission, explores the security concept, its meanings and how it is seen by states and individuals. The work emphasizes that safety is an abstract concept and with no specification the concept is empty of meaning. Just when the elements are filled, security appears as an operational concept.

2) Extract
Text is reviewed and the text key elements are underlined as show in Fig. 3.

At this step just the text that looks relevant to the concept is chosen. An initial and provisional list of categories is feasible at this point.

3) Collect
A simple database is populated with the key text elements. Fig. 4 shows the database structure and Fig. 5 some key text elements.

During this step, probably, some redundant key text elements could be discovered.

4) Categorize
Once all the requirements are collected, the list of categories needs to be made. Typically there are just a few categories. If previously a provisional list has been made, it is used to make the final one. In this case, the categories listed in table III are discovered.

<table>
<thead>
<tr>
<th>TABLE III</th>
</tr>
</thead>
<tbody>
<tr>
<td>SECURITY CATEGORIES</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Security</td>
</tr>
<tr>
<td>Security – concept</td>
</tr>
<tr>
<td>Security – attributes</td>
</tr>
<tr>
<td>Security – risk</td>
</tr>
<tr>
<td>Security – sort</td>
</tr>
</tbody>
</table>

5) Assign
Assigning a category to a requirement helps later on the graphical stage. We have obtained 81 key text elements. The ones which are useful to our purpose are discovered and categorized (table IV).

| TABLE IV |
| KEY TEXT CATEGORIZED |

<table>
<thead>
<tr>
<th>Categories</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Security</td>
<td>4</td>
</tr>
<tr>
<td>Security – concept</td>
<td>40</td>
</tr>
<tr>
<td>Security – attributes</td>
<td>4</td>
</tr>
<tr>
<td>Security – risk</td>
<td>2</td>
</tr>
<tr>
<td>Security – sort</td>
<td>6</td>
</tr>
</tbody>
</table>

6) Concept Map
This is the most “traditional” step. Once the requirements are collected and organized, we have all the ingredients to create the concept map. This step involves reviewing all the requirements, one by one, in order to raise all the relationships between them and related concepts.

The relations are spread all over the text, and so are into the requirement list. The concept map may contain redundancies, i.e. the same concept appears in different requirements with, apparently, no relation or two concepts are linked together in different places.
This step is the first one that reduces the amount of information gathered. Using the mentioned rules in table II or even making new ones should be useful to create the concept map.

The outcome is a set of ideas spread onto the canvas. A lot of redundancy is eliminated as shown in Fig. 6.

7) Class Diagram

This step helps reducing the amount of ideas in the previous stage. The outcome is a class diagram that represents the elements and relations involved (Fig. 7). The class diagram and the elements, using UML terminology, are classes and relations between classes and subclasses.

In order to create the class model, the following actions help.
- Fit each element (concept) in a class box.
- Add the attributes and behavior into the class.
- Create the relationship between elements. Add cardinality.

B. Discussion

Several points emerge from this work.

1) Meaning

The knowledge model obtained is a description of the elements, its components and relations among them. Like UML, the model has no meaning by itself. Thus, the case study of security expresses a range of possible definitions of security that are unveiled when the model is instantiated. At this point, when the components have a value, a security definition (class instantiated) appears. That security definition, using the object oriented paradigm, is unique.

2) Incremental Growth

The nature of the method permits an incremental growth of the knowledge model. An iterative process on other sources leads to a bigger and more detailed knowledge model without losing the knowledge acquired from the other sources. Even, new sources produce smaller or no changes because of the model become more complete at every cycle.

3) Uniqueness

As shown, the knowledge security model is meaningless. What if there are two security instances A and B?

If A = B then all the elements, relations and components are the same, and we can conclude that the security definition is the same.

If A and B are two security objects with B having, for example, a different set of policies or threats, we can conclude that in this scenario, A ≠ B.

Thus, there is no unique security definition. There are just security concept constructions and as many securities as
different security objects we are able to create. This is the reason why the “definition” of the resulting security is different. Therefore, persons, groups or states perceive different notions of security because the defining elements vary remarkably.

Besides, if we create a different construction of security (from other source for example), all the resulting objects will be different security objects (despite being neither semantically nor in practice incorrect).

4) Security Definitions and Computer Security

From Barry Buzan work [32], a wide range of security definitions are identified. For example, the human security from UNPD [34] or the expanded notion of security stated by Emma Rothschild [35] who argued that security notion is extended in “four main forms”. Open questions emerge such as is if all of those securities could be considered a kind of a bigger security model, actually a knowledge security model and how computer security and the existing securities could be peacefully integrated in such model. Because of the fact that computers are social tools, Computer Security needs an interdisciplinary work in order to become another kind of security.

V. CONCLUSIONS REMARKS

A methodology for exploring the underlying elements in a concept and the relationships among them is proposed. The outcome is an abstract concept, which requires specific elements to produce “the definition”. This definition is extremely flexible and can be adapted to almost any framework of any field.

The knowledge based concept analysis (KBCA) proposed method is based on knowledge engineering, concept maps and UML. It’s intended to extract knowledge from any informal source in order to obtain concept class diagram. That outcome could be used in object oriented engineering or knowledge based systems such as ontologies.

Concept analysis can also be made with knowledge elicitation applying some restrictions in the domain and the steps involved. The outcome of design stage in knowledge engineering, when the domain is restricted to one concept, leads to a type of concept analysis. The proposed method is a 7 steps concept analysis and literature based in order to overcome expert elicitation problems.

In the proposed scenario, the knowledge engineering analysis and design stages are focused just on one concept its attributes and its relations. The result is the knowledge model of a concept. Traditional concept analysis methods are stepwise. Our proposal is incremental, thus enlarge the model is easier. The UML purpose is to model any type of systems (not just software). This language should be understandable to humans and machines and could be used as a knowledge representation language.

KBCA is very systematic. Further implementation of the result, if needed, will be easier because of UML is used. The resulting diagram could be used to check by end-users or documentmakers and even could be used to integrate in bigger projects, related or not with computer software.

Despite we have reduced as much as possible the subjective component, the requirements gathering are a human task and the method still suffers from a subjective component. Thus, most probably the same text analyzed by several people may easily lead to slightly different outcome.

International Relations field has made, in the last decades, a lot of work on the concept and structure of the security notion. Their main concern are the types of securities, the existing relationship between several securities, security policies and, to a lesser extend the semantic notion of security and its consequences on individuals, entities or nationalities. There are no works available in order to link that security with information security in computer science. A generic framework could benefit both fields.

The security concept is meaningless until all the elements are instantiated and the “definition” of security rely on the values instead of the word on its own.

In the case of complex concepts, the review from just a single source of knowledge is clearly insufficient. Therefore, a further work to obtain a class diagram (formalization of a concept) from many sources (formal or informal) is needed. In order to extend the range, other kind of sources such as written documents, voice recordings, pictures and in general any multimedia documents need to be included.

More research is also needed to discover in which areas this methodology is useful and what changes or improvements would be needed to adapt to these new scenarios.

REFERENCES


Miquel Colobran is a doctoral student at the Department of Information and Communication Engineering. His research is in the field of ontologies, security and Social Computing.

Josep M. Basart is a PhD professor at the Department of Information and Communication Engineering. His research is in the field of Computer Ethics, Engineering Ethics, Applied Ethics and Social Computing.