

Empirical Validation of Concept Maps: Preliminary Methodological Considerations

Dietrich Albert and Christina M. Steiner

Cognitive Science Section - Department of Psychology
University of Graz
Universitätsplatz 2 / III
8010 Graz, Austria
{dietrich.albert, chr.steiner}@uni-graz.at

Abstract

For their usage in the semantic web, valid ontologies are required for a given domain. Here we focus on ontologies represented as concept maps (semantic nets). For one and the same domain several alternative concept maps may exist, originating from different world views or purposes. Some of these concept maps may be valid, however not all of them. Thus, efforts for validating empirically and objectively concept maps in the respective context are necessary. We outline two methodological approaches for empirically validating concept maps, one for giving evidence of content validity of a concept map and one for application validity. One procedure is to validate a given concept map with concept maps generated systematically by others. For this, persons of different knowledge level, are prompted to externalise their understanding of the domain through a concept mapping task. For giving evidence of the content validity of a given concept map, the similarity between the given concept map and the collected criterion maps has to be examined. As a second method of validating a given concept map, we suggest to observe the performance or behaviour in a relevant situational context as a validation criterion. In this scope, a method for predicting persons' problem solving behaviour by using a given concept map is outlined. In this case, evidence of the application validity of a given concept map can be found by comparing the predicted answer patterns with empirically obtained answer patterns. In general, the purpose and ultimate use of a given concept map has to be taken into consideration for choosing a validation procedure and interpreting its results.

1 Introduction

An ontology allows to structure a domain with respect to its conceptual organisation. It constitutes a specification of the concepts in a domain and the relations among them and thus, defines a common vocabulary of a knowledge domain (for an overview

see e.g. [Chandrasekaran et al., 1999; Fridman Noy et al 1997; Mizoguchi, 2003, 2004a, 2004b; Noy et al., 2001]. Currently, there is still a growing interest in ontologies and their application in information technology and computer science. Typical ontology applications arise in the context of the semantic web and e-Learning. The Semantic Web in which information is given well-defined meaning, as an extension of the current Web, is better enabling computers and people to work in cooperation [Berners-Lee et al., 2001]. For making the web computer-understandable it is necessary to write metadata for world wide web information and web services. In this regard, ontologies play the key role, in order to provide a vocabulary for metadata descriptions and to allow interoperability among metadata. In e-Learning one demanding issue is that of reuseability and interoperability of learning material, thus on the one hand standardisations (such as LOM) are required. On the other hand, for e-Learning ontologies have to be generated and applied in order to specify the vocabulary of the metadata for learning objects. On the whole, ontologies serve as a way to express a common vocabulary, as a help to information access, as a medium for mutual understanding, as a specification and as a foundation of knowledge systematisation [Mizoguchi, 2004a].

There are several alternatives to represent an ontology. There have been several attempts to create engineering frameworks for constructing ontologies [Chandrasekaran et al., 1999], i.e. formal languages exist for writing an ontology, such as e.g. Ontolingua or Web Ontology Language (OWL). They allow for using a logical notation for writing and sharing ontologies. Aside from such formal languages, concept maps (semantic nets), which will be the focus of this paper, provide a natural way of expressing and presenting ontologies [Hayes et al., 2003]. A concept map is characterised by a set of concepts and a set of relationships (or relations) between them. Most commonly, a concept map is represented as a labelled directed graph with the vertices (nodes) representing the concepts and the directed and labelled edges (arcs) between them representing the relationships that exist between concepts, such as is-a, part-of etc. (for an overview see e.g. [Novak, 2001]). In Figure 1

an example of a concept map is shown, describing what a concept map is.

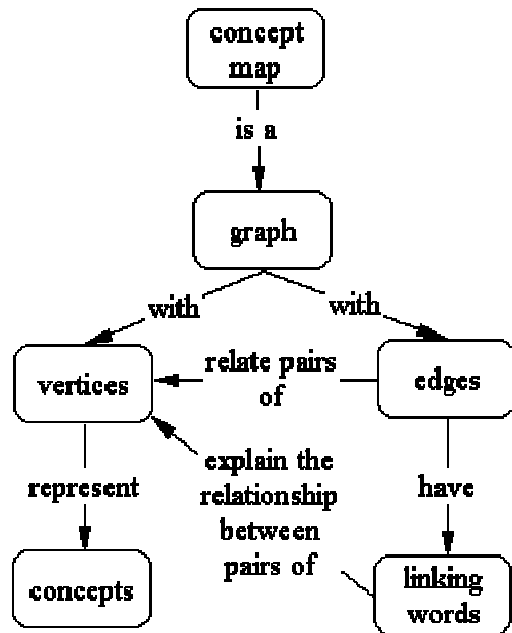


Figure 1. Concept map describing what a concept map is (adapted from [Ruiz-Primo, 2000])

The combination of two concepts and the link relating them constitutes a proposition. Hence, another way to represent a concept map is a list of propositions. Furthermore, the information contained in a concept map may also be presented in form of a matrix, with the set of concepts labelling the columns and rows and the relations depicted in the cells of the matrix. Of course, concept maps can be expressed also in a formal way (e.g. by notations of set theory and logic). On the whole, concept map representations constitute a natural way of displaying ontologies [Bokma et al., 2004; Dicheva et al., 2002; Hayes et al., 2003]. Hence, in the sequel we refer to concept maps as representations of ontologies.

In building a concept map, two different approaches can be distinguished: a normative and a descriptive one. The normative approach aims in obtaining the ‘one and only’ correct concept map of a specific domain, representing knowledge of complete consensus. Hence, constructing a concept map in a normative way assumes, that there exists only one concept map of completely shared understanding for a particular domain. Conversely, according to the descriptive approach, it is assumed that there may exist alternative concept maps for a specific domain, representing e.g. differing views or opinions. Hence, for a given domain different, but not arbitrary concept maps may exist. In general, this approach seems to be more reasonable and one should abandon demanding the ‘one’ correct concept map. A concept map or ontology necessarily entails some sort of world view with respect to a given domain [Uschold et al., 1996]. Different concept maps representing the same

domain, even when representing most general concepts, often differ from each other. Diverging concept maps may result due to different world views or alternate possibilities of sub-categorizations [Chandrasekaran et al., 1999]. Moreover, different concept maps categorising the same domain may differ by reason of their purpose and ultimate use. Therefore, there is not ‘one’ correct concept map for a specific domain.

Summarizing, according to a descriptive approach, not arbitrary concept maps of a domain may exist, but there may exist several valid concept maps representing the same domain. These alternatives may differ due to different views or because of their different purpose. Most likely, alternative concept maps modelling the same domain will match with respect to parts of the contained concepts and relationships or even agree in whole substructures. The interrelation between such concept maps is up to further research.

One general aim in constructing concept maps is to obtain one of possibly several viable and valid concept maps. Thus, a framework for evaluating the adequacy of a concept map or different proposals for a concept map should be provided. The field of evaluating concept maps is only just emerging. Until now, there is a lack of ideas and guidelines for evaluation issues.

[Leclère et al., 2002] claim that the best way for evaluating a concept map is the application for which the ontology was designed. This aspect of validity could be denoted as ‘application validity’ of a concept map. It refers to the practical usability and usefulness of a concept map. Thus, different kinds of applications will require different means of validating a concept map.

A concept map constitutes a model of a part of the current knowledge about the world for the given domain in the given context. Before examining the quality of a concept map by applying it for the purpose it is generated, i.e. application validity, evidence needs to be given regarding its validity as a model of the respective knowledge. This aspect of validity we call ‘content validity’. Content validity is an important issue, as the evaluation of the content of concept maps is critical for using them. It would be unwise to publish a concept map, to reuse an existing concept map, to build a new concept map, and to implement an application that relies on concept maps written by others (or even yourself), without first evaluating it [Gómez-Pérez, 2001]. The content validity of a concept map refers to the correct building of the content of the concept map, with the aim of proving compliance of the world model with the world modelled. According to [Gómez-Pérez, 2001], for the evaluation of a given concept map, the following criteria should be considered: consistency, completeness, conciseness, expandability, and sensitiveness. Of course, the evaluation whether a concept map adequately reflects the respective knowledge will also need to take into consideration its intended purpose and ultimate use.

Subjective evidence for the correctness of a concept map does not suffice for making a statement regarding the validity of the concept map. Even the principle of consensus regarding the correctness of a concept map does not suffice. In fact, objective and empirical criteria are needed for getting evidence of the validity of a given concept map.

To summarize, there has already been done some work regarding the evaluation and validation of ontologies and concept maps, but nevertheless there is still a lack of efficient strategies for validating them [Breitman et al., 2003]. The validity of a concept map can be understood from two perspectives. On the one hand, it may be examined whether a concept map serves the purpose for that it has been designed (application validity). On the other hand, before examining the application validity, it should be evaluated whether a concept map adequately reflects the knowledge in question (content validity).

We concentrate on both types of validity and propose two methods of empirically validating a concept map (in the sequel denoted as 'given concept map'). Regarding the content validity of a concept map, we propose an approach of validating a given concept map with empirically collected concept maps. For examining the application validity of a given concept map, we suggest to validate the concept map through performance. After discussing the issue of reliability in the following section, these two methods will be outlined in more detail.

2 Reliability of Concept Maps

Before considering the validity, the issue of reliability of a concept map should be regarded. In this scope the question arises for a concept map, how it has been developed. Let us assume an expert of a specific knowledge domain who constructed a concept map for this domain. By asking the respective person in a different point of time to construct a concept map for the same domain again, the reliability of the concept map can be examined. If both concept maps correspond to each other, they reliably represent (the domain or at least) the understanding of the respective person regarding the domain. Pre-assuming that no indication of a changing knowledge or understanding of the person exists; if the person generates a different concept map each time, though, it is obvious that none of them will be a reliable model of the knowledge of this person. Another aspect of reliability refers to the consistency of scores assigned to a concept map, when evaluating it according to particular characteristics such as number or accuracy of propositions, number of examples etc. When having two or more judges that independently from each other score a concept map according to a particular scoring system (e.g. [Novak et al., 1984; Ruiz-Primo et al., 1997]), the interrater reliability can be determined (e.g. [Ruiz-Primo, 2000]), i.e. whether different persons consistently score a concept map. In

this regard, interrater reliabilities around .90 could be found which indicates that different raters are able to reliably score concept maps [Shavelson et al., In Press]. By asking to generate a concept map in alternative forms of representation (e.g. as a directed graph and as a list of propositions), parallel-forms reliability could be examined. A reliable concept map should represent the same model of knowledge for a person, regardless in which representation format it is generated.

3 Validating a Concept Map with Concept Maps of Others: Content Validity

One approach of validating a given concept map as a model of knowledge is by taking empirically collected concept maps (in the sequel denoted as 'criterion maps') as a criterion for content validity.

[Cimolino et al., 2002] argue that a concept mapping task is able to capture a person's ontology of a domain. Hence, given a concept map representing the ontology of a specific domain, it may be validated by comparing it with empirically collected concept maps representing personal ontologies of different individuals. For this, persons of different knowledge level, including experts, are prompted to externalise their understanding of the domain through a concept mapping task. The concept map that is to be validated is then compared with the concept maps of individuals.

For posing a concept mapping task, a set of procedures is available, which will be described in more detail below. As concept maps for the same domain may differ due to their intended purpose, a concept mapping task should include some reference to the purpose and context when collecting criterion maps for validation.

3.1 Construct-a-map

One common method is to ask individuals to generate or construct a concept map concerning a specific knowledge domain from scratch. This is called 'construct-a-map' technique [Ruiz-Primo, 2000]. The method may vary considerably in practical application. The concepts and/or relations to be used for drawing the concept map may have to be generated or may also be provided. Moreover, it may either be required to construct a hierarchical or a non-hierarchical map or no information regarding the structure is given. The task of constructing a concept map may be completed individually or in groups. The maps may be drawn by hand, on the computer, or by arranging note cards. For evaluating constructed maps, different scoring systems exist [e.g. Novak et al., 1984; Ruiz-Primo et al., 1997] considering different map characteristics such as number of correct propositions, proposition accuracy, levels of hierarchy, examples etc.

One possibility of validating a given concept map based on collected criterion maps is to have the given concept map as well as the criterion maps scored by several raters. The correlation between the given concept map and the criterion maps can then be used in order to give evidence of the validity of the given concept map.

Another way, which is more direct, is to examine the similarity of the given concept map to the criterion maps. For this purpose, the conformance of propositions, i.e. of pairs of related concepts, can be examined. For examining the similarity between the given concept map and the criterion concept maps suitable measures of association [e.g. Tversky, 1977] may be chosen and utilized. For example 2x2 tables can be established for a given concept map and a criterion map, explicating the number of propositions that are contained either in both or in only one of them. Based on this, similarity or distance measures can be calculated. Moreover, a so-called convergence score¹ can be computed in correspondence to [Ruiz-Primo, 2000], which expresses the proportion of propositions in the given concept map out of the total number of propositions in one criterion map. A high similarity or convergence between the given concept map and the criterion maps indicates the content validity of the given concept map. The same conclusion can be made using well known advanced statistical measures of association developed by L.A. Goodman and W.H. Kruskal.

One prerequisite for the application of the 'construct-a-map' method is that persons that are to undergo this concept mapping task, first need to learn how to construct concept maps, i.e. they need to receive some training [Ruiz-Primo, 2000]. This might be a time-consuming process which probably results in frustration and rejection of the method [Schau et al., 2001]. One general problem of this method is, that there is no universally accepted scoring system for the evaluation of constructed maps. Therefore, when using concept map scores for validation, applying and comparing alternative scoring systems might result in different findings regarding the validity of a concept map. Thus, the direct method is preferable, i.e. validating a concept map by examining the similarity between the given concept map and the criterion map in terms of measures of similarity, convergence or association.

3.2 Fill-in-the-map

Another possibility of posing a concept mapping task is a method called 'fill-in-the-map' or 'map completion', respectively [Ruiz-Primo, 2000; Schau et al., 2001]. For this, a concept map is provided, where all or some of the concepts and/or relations

have been left out. The blanks have to be filled in requiring either the generation of the words to use or by selecting them from a list which may include distractors. When collecting criterion maps with this method, the concept map to be validated then can be compared with the criterion maps. This would be done again, by examining the similarity, e.g. by calculating typical measures of association.

Some researchers claim that this method should be preferred to the 'construct-a-map' technique, as the least is assumed to impose a too high cognitive demand and the generated maps may highly depend on a person's communication skills [Schau, et al. 2001]. Map completion tasks can easily and quickly be administered. However, this method poses the problem that only a part of the knowledge can be queried, as it is necessary to provide at least a small part of the concepts and/or relations in the map to be filled in. As a representation of the domain structure is provided, a person does not have to create an individual structure representing his/her personal ontology of the domain. Therefore, the validity of a whole given concept map is likely to be overestimated, when using a map completion task for collecting concept maps of individuals. On the other hand, the 'fill-in-the-map' technique has the advantage of allowing to validate specific parts of a concept map that are of special interest. This would be useful, for example, if only the relations of a concept map are to be validated. The map completion method would also be of special advantage, if two proposals for a concept map of a domain are to be evaluated. Such two concept maps will most probably overlap in parts of their propositions. For deciding which one of the two maps is the more adequate one, criterion maps could be collected by posing a map completion task querying for those parts that are not in the intersection of the two maps. When the validation of particular parts of a given concept map is intended, this needs to be taken into account for selecting and applying the methods for statistical analysis. Furthermore, the number of words that have to be filled in should also be regarded, as a map completion task implies – in contrast to the 'construct-a-map' technique – a constraint with respect to the number of concepts and relations that are contained or requested in the concept map.

3.3 Relatedness ratings of pairs of concepts

One further alternative that can be classified as a concept mapping task is characterized by a two-stage indirect approach [Schau et al., 1997]. This technique is based on relatedness ratings between pairs of concepts, i.e. individuals are asked to rate the degree of relatedness between pairs of previously defined concepts on a numerical scale. The resulting relatedness matrix then can be used to visually represent the personal ontology of the respective person in form of a concept map. This is done through the application of an algorithm, e.g. by using the Pathfinder software [Schvaneveldt, 1990], which allows to create a network representation based on

¹ In the present context, the convergence C of a given concept map to one criterion map is calculated in the following way: $C = x / y$, whereas x denotes the number of propositions in the given concept map that are also contained in the criterion map and y denotes the total number of propositions in the criterion map.

relatedness ratings. Notice, that the generated maps based on relatedness ratings do not include labelled relations, but only indicate the degree of relatedness through the distance between two given concepts. When collecting criterion maps for validating a given concept map in this way, this fact has to be taken into account. Hence, only a particular aspect of validity of the given concept map can be examined, namely the existence of relationships between concepts. The type of the relations between the concepts of the given concept map, however, can not be validated in this case.

In order to overcome the drawback of this method, [Shavelson et al., In Press] propose that after analysing the relatedness ratings, the resulting network representation can be provided to the individual. The person then will be asked to label the relations between the concepts, and if necessary to add or remove relations. Based on concept maps collected in this way, statements about the content validity of a given concept map could also be made with respect to the type of relations that exist between concepts.

In both cases, i.e. when collecting the criterion concept maps with the original method or with the extended method proposed by [Shavelson et al., In Press], evidence on the validity of a given concept map can be given again by examining the similarity or conformance to the criterion maps, e.g. by utilizing appropriate measures of association.

3.4 Proposition correct-incorrect discrimination task

Another technique for posing a concept mapping task makes use of the representation of a concept map through a list of propositions. The propositions of a concept map representing a particular domain may be presented to individuals as a correct-incorrect discrimination task. This technique has been implemented in the concept mapping software 'Cmap Pro' [Bernd et al., 2000] in order to assess previous knowledge before a learning sequence takes place. In [Steiner, 2004] an extension of this method was proposed and applied, which will be described in the following. In addition to using the propositions from the concept map, distractor items are included in the correct-incorrect discrimination task. In order to reduce the risk of lucky guesses, a confidence rating is required for each proposition judgement. Moreover, a phase of answer checking can be realised. For this, after finishing the correct-incorrect discrimination task, for each subject an individual concept map is drawn with respect to his/her answer pattern. Such an individual concept map represents the personal ontology of the respective person, containing all those propositions that have been judged as being 'correct'. Notice that, of course limited by the presented distractor items, such an individual concept map may also include misconceptions if distractor items are wrongly judged as 'correct'. Based on the individual concept map, each person then has the opportunity to check his/her answers and make corrections, if

required. Applying this technique seems easy and comfortable and does not need any foregoing training. Moreover, it is possible to query all the propositions contained in a concept map, which is not possible when using the fill-in-the-map technique. This method can also be applied for collecting criterion maps for the validation of a given concept map. This technique allows to validate a whole concept map as well as specific parts of a concept map that are of special interest. When using this method for collecting criterion maps, the propositions of the given concept map will be presented as correct-incorrect discrimination task in combination with distractor items, which could be actually incorrect statements but also statements representing an alternative world view. The validity of the given concept map then can be examined by considering its similarity to the collected criterion maps through measures of association etc.

Summarising, there are various ways of presenting a concept mapping task which considerably vary in the extend of constraints imposed and information provided to persons, respectively. Based on these techniques it is possible to empirically collect criterion maps representing the personal ontologies of individuals with different knowledge level. These criterion maps then may be utilised for examining the content validity of a given concept map, by considering the similarity or conformance to the criterion maps. A high association between the given and the criterion maps indicates the validity of the given concept map.

4 Validating a Concept Map with Performance: Application Validity

For examining the application validity of a given concept map, we propose making use of observing situational behaviour or performance as another conceivable and valuable criterion of validation. In the present context, situational performance refers to behaviour in real-world situations which does not consist in performing a concept mapping task. The performance in question is for instance solving problems, answering questions, or even social behaviour in given situations. For this validation approach, we have a concept map that needs to be validated on the one hand. On the other hand we have empirically collected situational performance scores or profiles of people of different knowledge level, including experts. These performance measures are related to the purpose and intended application of the given concept map. In this case the personal ontologies are not elicited directly through a concept mapping task. Indirectly, the performance of a person - e.g. in a problem solving context - is controlled by his/her individual knowledge and understanding regarding the ontology of a domain. Situational performance as indicating indirectly knowledge is

therefore useable as a criterion for examining the validity of a concept map. When using situational performance as a criterion for validating a concept map, the purpose of the given concept map has to be kept in mind. Due to different purposes that are intended for a given concept map, different kinds of situational performance will be appropriate for validation.

Until now there have been conducted a number of investigations that tried to validate individual concept maps on the basis of situational performance. This means, individuals completed a concept mapping task on the one hand and underwent a context of situational performance, as e.g. doing a multiple choice test. Taking the performance scores as a criterion, it was tried to give evidence of the validity of the individual concept maps. When using multiple choice scores as a criterion, the findings of [Schau et al., 2001] and [Rice et al., 1998] indicated a high validity (.75) of individual concept maps. [Ruiz-Primo, 2000] also obtained relatively high validity scores (on average .57). [West et al. 2000], who used standardized tests as a validation criterion, reported indications of low validity of individual concept maps. Conversely, [Rice et al. 1998] received high validity scores (.82-.87) for individual concept maps when using standardized tests as a criterion. [Schultz, 1999, cited from Shavelson et al., In Press] applied different criteria based on performance (e.g. multiple choice test, reading test) and reported moderately high validity scores (ranging from .43 to .60) for individual concept maps. [Jöckel et al., 1999] conducted a study in the context of geometry and spatial ability. Individual concept maps were collected by asking participants to generate maps with note cards. As a criterion for validating these individual concept maps, problem solving performance referring to the knowledge domain was used. The individual concept maps did not prove to be valid in this investigation applying performance as validity criterion.

[Steiner, 2004] also used problem solving performance as a criterion for validating individual concept maps. This validation approach was conducted in the knowledge domain of geometry of right triangles. The individual concept maps were collected by using the concept mapping task outlined in the 'proposition correct-incorrect discrimination task' section. Furthermore, typical geometry problems coming from the respective domain were posed to the participants of the investigation. The data collected from subjects' problem solving performance was then utilised for examining the validity of the individual concept maps. The results indicated a high validity of the individual concept maps (.67-.74). Hence, the individual concept maps validly represented personal ontologies of individuals.

Summarising, it could be shown that measures of situational performance are suitable for validating individual concept maps or at least, that individual concept maps and performance are somehow positively interrelated. It seems natural and obvious,

that an individual's personal understanding of the ontology of a domain is reflected in his/her performance in given situations, at least as declarative knowledge is involved. Hence, we are in the position that a given concept map can not only be validated with concept maps of persons of different knowledge level, but also with their performances. In the sequel we will describe such an approach in more detail.

As in the suggested validation approach, Knowledge Space Theory [Albert et al., 1999; Doignon et al., 1985, 1999; Falmagne et al., 1990] is implemented, let us first give a very short introduction into the basic notions of this theoretical framework. Knowledge Space Theory provides a formal model for structuring a domain of knowledge and for representing the knowledge of individuals based on prerequisite relationships. Within this theory, a domain of knowledge is characterised by a set Q of problems. The knowledge of a learner is represented by the subset of problems that he/she is capable of solving. Due to mutual dependencies between the problems of a domain, from the correct solution of certain problems the mastery of other problems can be surmised. In order to capture the relationships between the problems of a domain the notion of a surmise relation has been introduced. Two problems a and b are in a surmise relation whenever from a correct solution to problem b the mastery of problem a can be surmised. In other words, problem a is a prerequisite problem for problem b . A surmise relation can be illustrated by a so-called Hasse diagram (see Figure 2 for an example), where descending sequences of line segments indicate a surmise relationship.

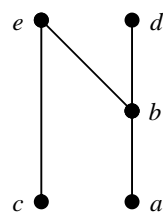


Figure 2: Example of a Hasse diagram illustrating a surmise relation on a knowledge domain $Q = \{a, b, c, d, e\}$ (adapted from [Falmagne et al., 1990])

According to the surmise relation depicted in Figure 2, from a correct solution to problem b the correct solution to problem a can be surmised, while the mastery of problem e implies correct answers to problems a , b , and c . The surmise relation forms a quasi-order on the set of problems and restricts the number of possible knowledge states (i.e. subsets of problems) that are expected to be observable. The collection of all possible knowledge states, including the empty state \emptyset and the whole set Q , constitutes the so-called knowledge structure. The knowledge structure corresponding to the surmise relation shown in Figure 2 is given by

$$K = \{\emptyset, \{a\}, \{c\}, \{a, c\}, \{a, b\}, \{a, b, c\}, \{a, b, d\}, \{a, b, c, e\}, \{a, b, c, d\}, Q\}.$$

Let us now assume a concept map of a particular knowledge domain that has to be validated. As a criterion for validation, situational performance, in particular in a problem solving context, is intended to be applied. A set of typical and representative problems of the domain is chosen. To each problem the subset of propositions of the given concept map is assigned, that represents those semantic knowledge elements that are required for solving the respective problem. Each proposition can be considered as an atomic competency or skill [e.g. Falmagne et al., 1990; Doignon, 1994; Düntsch et al., 1995; Korossy, 1997, 1999] required for solving the problems of the domain. The subsets of propositions assigned to the problems will most likely overlap. Based on the representation of the problems by subsets of the given concept map, dependencies between problems in terms of a surmise relation can be derived. This could be done e.g. by set inclusion, i.e. if the representation of a problem *a* in the concept map is a subset of that of problem *b*, then problem *a* is a prerequisite for problem *b*. From the derived dependencies between the problems a knowledge structure can be established, i.e. the possible knowledge states can be identified. This means, that we are able to predict specific answer patterns out of all possible subsets of items ($2^{|Q|}$). Based on this, we can investigate empirically whether the observed answer patterns correspond to the identified and predicted knowledge states. For this, the problems are posed to individuals of different knowledge level. The answer patterns that are obtained empirically then are compared to the predicted knowledge states. As these knowledge states have been formed in the basis of the problems' representations on the concept map, the answer patterns serve as a criterion for validating the given concept map. Comparing the empirically obtained answer patterns with the knowledge states can be done e.g. by using a discrepancy index, describing the similarity between the knowledge structure and the set of answer patterns. If the answer patterns correspond well to the predicted knowledge states, the given concept map can be considered also to be valid – under the condition that both, the chosen set of problems as well as the sample of persons are adequate and representative.

5 Discussion

In general, it does not suffice to subjectively evaluate a concept map and judge it as valid by subjective evidence or uncontrolled consensus. In fact, objective, empirical measures or criteria are needed to give evidence of a concept map's validity. Efforts for empirically and objectively validating concept maps have proven reasonable and promising.

In this paper we suggested two approaches for empirically validating a given concept map, one for giving evidence of content validity of a given concept map and one for application validity. The suggested approach concerning content validity is to validate a given concept map by taking empirically collected concept maps of others as a criterion. The approach proposed regarding application validity applies situational performances as a criterion. Both approaches constitute useful and valuable procedures that can complement each other in order to build up a coherent picture about the validity of a given concept map. Of course, there are also other possibilities for validating concept maps, such as e.g. consulting the published literature of the given knowledge domain for examining content validity.

According to the suggested validation approaches, it is not necessary to only query experts in a given field for which a concept map is to be validated. In fact, we propose to rather involve persons of different knowledge level, who will probably afterwards work (indirectly) with the validated concept map.

Regarding the validation of a given concept map it is very important to take into account the purpose and ultimate use of the concept map (e.g. predicting problem solving behaviour). This refers to validation issues concerning application validity as well as content validity. The intended purpose of a given concept map has to be regarded when collecting concept maps as validation criterion, i.e. the concept mapping task should refer to the context and purpose of the concept map. The choice of a measure of situational performance will likewise depend on the purpose that a given concept map will serve in the future.

When applying validation methods for a given concept map, it has to be specified which aspects are intended to be validated (e.g. the whole concept map, the concepts, the relations, substructures of the given concept map). Another important issue is to choose appropriate statistical measures, depending on the particular criterion for validity and its constrains.

Finally, we want to point out, that there is little attention paid to the reliability of concept maps. Before considering the validity of a given concept map, it should be examined whether the generated model of knowledge is reliable, i.e. whether the concept map looks the same when it is constructed again or by a parallel method. Hence, the issue of reliability or at least the circumstances under which the concept map originated, should be taken into account when the validity of a concept map is addressed.

We outlined, that for a given knowledge domain there might be several alternative concept maps that validly represent the respective knowledge of the domain. Such alternative concept maps may originate from different world views or opinions, but may also be due to different purposes of the maps. Based on this fact two interrelated questions arise:

- Should alternative ontologies be realised in the context of the semantic web?
- If alternative ontologies are not to be implemented into the semantic web, which one should be realised?

These two questions are going far beyond the scope of validating ontologies in form of concept maps and give raise to future research.

References

- [Albert et al., 1999] Albert, D. & Lukas, J. (1999). *Knowledge Spaces: Theories, Empirical Research, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- [Bernd et al., 2000] Bernd, H., Hippchen, T., Jüngst, K.-L., & Strittmatter, P. (2000). Durcharbeiten von Begriffsstrukturdarstellungen in unterrichtlichen und computergestützten Lernumgebungen. In H. Mandl & F. Fischer (Eds.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken (pp 15-36)*. Göttingen: Hogrefe.
- [Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). *The Semantic Web*. Scientific American 284, 34-43. available at < http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 > (30 March 2005)
- [Bokma et al., 2004] Bokma, A., Solomon, N., Harvey, H., & Jones, S. (2004). NOPIK: Creating Shared Conceptualisations for Information and Knowledge Sharing in the Virtual Enterprise. *Proceedings of the 10th International Conference on Concurrent Enterprising, Seville, June 2004*. available at <http://voyager.sunderland.ac.uk/Downloads/nopik_ICE.pdf> (10 May 2005)
- [Breitman et al., 2003] Breitman, K. & Leite, J. C. S. (2003). Ontology as a Requirement Engineering Product. In *Proceedings of the 11th IEEE International Requirements Engineering Conference. 8-12 Sept. 2003, Monterey Bay, California, USA. (pp. 309-319)*.
- [Chandrasekaran et al., 1999] Chandrasekaran, B., Josephson, J.R., & Benjamins, V.R. (1999). What Are Ontologies and Why Do We Need Them?. *IEEE Intelligent Systems, 14 (1)*, 20-26. available at <http://www.infofusion.buffalo.edu/conferences_and_workshops/ontology_wkshop_2/ont_ws2_working_materials/ChandrasekaranRoleofOntology.pdf> (30 March 2005)
- [Cimolino et al., 2002] Cimolino, L & J Kay, (2002). Verified Concept Mapping for Eliciting Conceptual Understanding. In Aroyo, L & Dicheva, D, (Eds.), *Proceedings ICCE Workshop on Concepts and Ontologies in Web-based Educational Systems, ICCE 2002, International Conference on Computers in Education, CS-Report 02-15 Technische Universiteit Eindhoven (pp. 9-14)*. Proceedings available at <http://www.win.tue.nl/~laroyo/ICCE2002_Workshop/proc-Workshop-ICCE2002.pdf> (30 March 2005)
- [Dicheva et al., 2002] Dicheva, D. & Aroyo, L. (2002). Concept-based Courseware Authoring: An Engineering Perspective. *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 2002) (pp. 46-49)*. available at <http://ltf.ieee.org/icalt2002/proceedings/t108_icalt099_End.pdf> (30 March 2005)
- [Doignon, 1994] Doignon, J. (1994). Knowledge spaces and skill assignments. In G. H. Fischer & D. Laming (Eds.), *Contributions to Mathematical Psychology, Psychometrics and Methodology* (pp. 111–121). New York: Springer-Verlag.
- [Doignon et al., 1985] Doignon, J.-P. & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies, 23*, 175-196.
- [Doignon et al., 1999] Doignon, J.-P. & Falmagne, J.-C. (1999). *Knowledge Spaces*. Berlin: Springer.
- [Dütsch et al., 1995] Dütsch, I., & Gediga, G. (1995). Skills and Knowledge Structures. *British Journal of Mathematical and Statistical Psychology, 48*, 9–27.
- [Falmagne et al., 1990] Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johannesen, L. (1990). Introduction to Knowledge Spaces: How to Build, Test, and Search Them. *Psychological Review, 97*, 201-224.
- [Fridman et al., 1997] Fridman Noy, N. F. & Hafner, C.D. (1997). The State of the Art in Ontology Design. A Survey and Comparative Review. *AI Magazine, 18 (3)*, 53-74.
- [Gómez-Pérez, 2001] Gómez-Pérez, A. (2001). Evaluation of Ontologies. *International Journal of Intelligent Systems, 16*, 391-409.
- [Hayes et al., 2003] Hayes, P., Saavedra, R., Reichherzer, T. (2003). A Collaboration Development Environment for Ontologies. *Proceedings of the Semantic Integration Workshop, Sanibel Island, Florida, October, 20, 2003*. available at <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS//Vol-82/SI_demo_05.pdf> (10 May 2005)
- [Jöckel et al., 1999] Jöckel, S. & Reiss, K. (1999). *Ergebnisse des DFG-Projekts „Problemlösestrategien bei räumlich geometrischen Aufgaben mit konkretem oder*

- computersimuliertem Material“. Universität Oldenburg: Fachbereich Mathematik. available at <<http://www.math.uni-oldenburg.de/didaktik/strategie/bericht.html>> (30 March 2005)
- [Korossy, 1997] Korossy, K. (1997). Extending the theory of knowledge spaces: a competence-performance approach. *Zeitschrift für Psychologie*, 205, 53-82.
- [Korossy, 1997] Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge Spaces: Theories, Empirical Research, Applications* (pp. 103-132). Mahwah, NJ: Lawrence Erlbaum Associates.
- [Leclère et al., 2002] Leclère, M., Trichet, F., & Furst, F. (2002). Operationalizing domain ontologies: towards an ontological level for the SG family. In *Supplementary proceedings of 10th International Conference On Conceptual Structures (ICCS'2002), Borovets, Bulgaria*.
- [Mizoguchi, 2003] Mizoguchi, R. (2003). Tutorial on ontological engineering - Part 1: Introduction to Ontological Engineering. *New Generation Computing*, 21, 365-384. available at <<http://www.ei.sanken.osaka-u.ac.jp/pub/miz/Part1-pdf2.pdf>> (30 March 2005)
- [Mizoguchi, 2004a] Mizoguchi, R. (2004a). Tutorial on ontological engineering - Part 2: Ontology development, tools and languages. *New Generation Computing*, 22, 61-96. available at <<http://www.ei.sanken.osaka-u.ac.jp/pub/miz/Part2V3.pdf>> (30 March 2005)
- [Mizoguchi, 2004b] Mizoguchi, R. (2004b). Tutorial on ontological engineering - Part 3: Advanced course of ontological engineering. *New Generation Computing*, 22, 196-220. available at <<http://www.ei.sanken.osaka-u.ac.jp/pub/miz/Part3V3.pdf>> (30 March 2005)
- [Novak, 2001] Novak, J.D. (2001). The Theory Underlying Concept Maps and How To Construct Them. available at <<http://cmap.coginst.uwf.edu/info/printer.html>> (10 May 2005)
- [Novak et al., 1994] Novak, J.D. & Gowin, B.D. (1984). *Learning how to learn*. New York: Cambridge University Press.
- [Noy et al., 2001] Noy, N. F. & McGuinness, D.L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Technical Report KSL-01-05. Stanford University. available at <<http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>> (30 March 2005)
- [Rice et al., 1998] Rice, D.C., Ryan, J.M., & Samson, S.M. (1998). Using Concept Maps to Assess Student Learning in the Science Classroom: Must Different Methods Compete? *Journal of Research in Science Teaching*, 35, 1103-1127. available at <http://www.clab.edc.uoc.gr/hy302/papers%5Cconcept_map%20assessment.pdf> (30 March 2005)
- [Ruiz-Primo, 2000] Ruiz-Primo, M.A. (2000). On the Use Of Concept Maps As An Assessment Tool in Science: What We Have Learned so Far. *Revista Electrónica de Investigación Educativa*, 2. available at <<http://redie.uabc.mx/vol2no1/contents-ruizpri.html>> (30 March 2005)
- [Ruiz-Primo et al., 1997] Ruiz-Primo, M.A., Schultz, S.E., & Shavelson, R.J. (1997). *Concept map-based assessment in science: two exploratory studies*. CSE Technical Report 436. Los Angeles: University of California, Center for the Study of Evaluation (CSE), National Center for Research on Evaluation, Standards, and Student Testing. available at <http://www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/TECH436.pdf> (30 March 2005)
- [Schau et al., 1997] Schau, C. & Mattern, N (1997). Assessing Students' Connected Understanding of Statistical Relationships. In I. Gal & J.B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 91-104). Amsterdam: IOS Press.
- [Schau et al., 2001] Schau, C., Mattern, N., Zeilik, M., Teague, K. W. & Weber, R. (2001). Select-and-Fill-In Concept Map Scores as a Measure of Students' Connected Understanding of Science. *Educational and Psychological Measurement*, 61, 136-158.
- [Schvaneveldt, 1990] Schvaneveldt, R.W. (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood: Ablex.
- [Shavelson et al., In Press] Shavelson, R.J., Ruiz-Primo, M.A., & Wiley, E. (In Press). Windows into the Mind. *International Journal of Higher Education*.
- [Steiner, 2004] Steiner, C.M. (2004). *Beziehung zwischen Concept Maps und Kompetenzstrukturen*. Unpublished diploma thesis: University of Graz.
- [Tversky, 1977] Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- [Uschold, 1996] Uschold, M & Gruninger, M (1996). Ontologies: Principles, Methods, & Application. *Knowledge Engineering Review*, 11 (2), 93-136.
- [West et al., 2000] West, D., Pomeroy, J.R., Park, J., Gerstenberger, E., & Sandoval, J. (2000) Critical thinking in graduate medical education. A Role for Concept Mapping Assessment? *Journal of the American Medical Association*, 284, 1105-1110.