Noname manuscript No.
(will be inserted by the editor)

Cross-categorization of legal concepts across boundaries of legal systems

in consideration of inferential links

Received: date / Accepted: date

Abstract This work contrasts Giovanni Sartor's view of inferential semantics of legal concepts (Sartor 2009) with a probabilistic model of theory formation (Kemp et al. 2010). The work further explores possibilities of implementing Kemp's probabilistic model of theory formation in the context of mapping legal concepts between two individual legal systems. For implementing the legal concept mapping, we propose a cross-categorization approach that combines three mathematical models: the Bayesian Model of Generalization (BMG) (Tenenbaum and Griffiths 2001), the probabilistic model of theory formation, i.e. the Infinite Relational Model (IRM) first introduced by Kemp et al. (2006, 2010) and its extended model, i.e. the normal-Infinite Relational Model (n-IRM) proposed by Herlau et al. (2012). We apply our cross-categorization approach to datasets where legal concepts related to educational systems are respectively defined by the Japanese- and the Danish authorities according to the International Standard Classification of Education (ISCED). The main contribution of this work is the proposal of a conceptual framework of the cross-categorization approach that, inspired by (Sartor 2009), attempts to explain reasoner's inferential mechanisms.

Keywords Cross-categorization · Knowledge structuring · Ontology alignment · Generalization · Inference · Relational model

1 Introduction

An ontology is traditionally considered as a means for standardizing knowledge represented by different parties involved in communications (Gruber 1992; Masolo et al. 2003; Declerck et al. 2010). Kemp et al. (2010) also points out that some scholars (Block 1986; Field 1977; Quilian 1968) have argued the importance of knowledge structuring, i.e. ontologies, where concepts are organized

Address(es) of author(s) should be given

into systems of relations and the meaning of a concept partly depends on its relationships to other concepts. However, real human to human communication cannot be absolutely characterized by such standardized representations of knowledge. In Kemp et al. (2010), two challenging issues are raised against such idea of systems of concepts. First, as Fodor and Lepore (1992) originally pointed out, it is beyond comprehension that the meaning of any concept can be defined within a standardized single conceptual system. It is unrealistic that two individuals with different *beliefs* have common concepts. This issue has also been discussed in semiotics (Peirce 2010; Durst-Andersen 2011) and in cognitive pragmatics (Sperber and Wilson 1986). For example, Sperber and Wilson (1986) discuss how mental representations are constructed diversely under different environmental and cognitive conditions. Second point which Kemp et al. (2010) specifically deals in their framework is the concept acquisition problem. According to Kemp et al. (2010) (see also: Hempel (1985); Woodfield (1987)):

> *if the meaning of each concept depends on its role within a system of concepts, it is difficult to see how a learner might break into the system and acquire the concepts that it contains.* (Kemp et al. 2010)

Interestingly, the similar issue is also discussed by legal information scientists. Sartor (2009) argues that:

> *legal concepts are typically encountered in the context of legal norms, and the issue of determining their content cannot be separated from the issue of identifying and interpreting the norms in which they occur, and of using such norms in legal inference.* (Sartor 2009)

This argument implies that if two individuals who are respectively belonging to two different societies having different legal systems, a legal term could be interpreted differently by these two individuals, since the norms in which the two individuals belong are not identical. The argument also implies that these two individuals must have difficulties in learning a concept contained in the other party's legal system without interpreting the norms in which the concept occurs.

These arguments motivate us to contrast the theoretical work presented by Sartor (2009) with the probabilistic model of theory formation by Kemp et al. (2010) in the context of mapping legal concepts between two individual legal systems. Although Sartor's view addresses the inferential mechanisms within a single legal system, we argue that his view is applicable in a situation where a concept learner (reasoner) is, based on the norms belonging to his or her own legal system, going to interpret and adapt a new concept introduced from another legal system. In Sartor (2009), *the meaning of a legal term results from the set of inferential links.* The inferential links are defined based on the theory of Ross (1957) as:

1. the links stating what conditions determine the qualification Q (Q-conditioning links), and

2. the links connecting further properties to possession of the qualification Q (Q-conditioned links.)

These definitions can be seen as causes and effects in Kemp et al. (2010). If a reasoner is learning a new legal concept in his or her own legal system, the reasoner is supposed to seek causes and effects identified in the new concept that are common to the concepts which the reasoner already knows. This way, common-causes and common-effects existing within a concept system, i.e. underlying relationships among domain concepts, are identified by a reasoner. The probabilistic model in Kemp et al. (2010) is supposed to learn these *underlying relationships* among domain concepts and identify a system of legal concepts from a view where a reasoner acquires new concepts in contrast to the concepts already known by the reasoner.

The probabilistic model of theory formation in Kemp et al. (2010) has originally been introduced in Kemp et al. (2006) where it is named the Infinite Relational Model (IRM). Kemp et al. (2006) in their original work applies this model to the ontology learning problem. The model has further been extended and applied to, e.g., the area of neuroimaging data analysis (Mørup et al. 2010), and of collaborative filtering and topic modeling (Xu et al. 2006; Hansen et al. 2011). The present work applies the IRM to analyze structures of an educational system in a specific legal district, which are represented by legal terms and their characteristic features as inferential links: i.e. i) links stating what conditions determine the qualification of an educational concept Q (Q-conditioning links); and ii) links connecting further properties to possession of the qualification Q (Q-conditioned links).

The problem which we address in our work is, however, to map legal concepts between two individual legal systems. This requires an additional reasoning mechanism, i.e., analogical reasoning, which is out of scope in (Kemp et al. 2010). On the other hand, Sartor (2009) applies Thagard's view of conceptual roles (Thagard 1992) to legal reasoning and emphasizes the importance of generalization process and analogical inference. More specifically, we argue that, when a reasoner is learning a new concept introduced from an alien system, the analogical inference is used for comparing a concept introduced from the alien system with the concepts existing in its own legal system. Accordingly, in our work we employ a model, the so-called Bayesian Model of Generalization (BMG) proposed by Tenenbaum and Griffiths (2001) for the analogical inference. The BMG computes a relation between two legal concepts based on characteristic features in a special way: i) it only considers features that are already known by a reasoner, when the reasoner compares a new object with a referent concept (a Danish reasoner who knows nothing about the Japanese educational system cannot know all the characteristic features of all concepts existing in the Japanese system, so only features known by the Danish reasoner should be considered) and ii) it distinguishes importance of features and assigns weights to each feature according to the degree of importance (in the BMG, a feature possessed by few concepts are considered important, and vice versa). In other words, the analogical inference employing the BMG is supposed to seek causes and effects identified in the new concept that are common to the concepts which the reasoner already knows, and prioritizes the causes and effects that are more important than others. After computing the degree of relations between all possible combinations of concepts existing in the two legal systems, we apply an extended version of the IRM, the so-called *normal Infinite Relational Model* (n-IRM) proposed by Herlau et al. (2012) for cross-categorizing the educational concepts existing in the two legal systems. The application of the n-IRM expects to identify number of categories, i.e., groups of educational concepts for the respective legal systems and the degree of the relations between categories in the two legal systems. Finally, the IRM is applied to the original data consisting of legal concepts and their characteristic features for identifying underlying relationships, i.e., feature structures, behind the specified concept system, i.e., the identified categories. We have in our previous work (anonymous) () identified that the usefulness of the IRM for mapping sparse datasets representing two different knowledge systems. However, the previous approach has not been suitable to dense datasets employed in this work. The implementation of the n-IRM realizes the cross-categorization approach by fully utilizing information contained in different types of input data (similarity scores).

The main contribution of our work is the proposal of a conceptual framework of the cross-categorization approach that, inspired by (Sartor 2009), attempts to explain reasoner's inferential mechanisms. We demonstrate the principles of the proposed approach by applying it to datasets where legal concepts related to educational systems are respectively defined by the Japanese- and the Danish authorities according to the International Standard Classification of Education (ISCED).

In the next section, we first review relevant parts of Sartor's work that strengthen the argument for our cross-categorization approach followed by a review of the probabilistic model of theory formation in Section 3. Section 4 briefly summarizes the idea of generalization and analogical inference, followed by the overall idea of cross-categorization approach in light of Sartor's work and in comparison with other related works in Section 5. Sections 6 and 7 respectively review the generalization model and the relational models employed in our cross-categorization approach. After introducing the data source used in this work in Section 8, we present our quantitative results, qualitative analysis and visualization in Sections 9, 10 and 11, respectively. Section 12 discusses the issues to be taken consideration as well as future perspectives, and finally our conclusions follows in Section 13.

2 Inferential semantics of legal concepts

In the view of Sartor (2009) *the meaning of a term* results from the set of inferential links that are extracted from the sentences where the term occurs. The set of inferential links constitute the meaning of the corresponding concept. Sartor gives the concept of *citizenship* as an example. The concept *citizenship* consists of set of inferential links such as *birth in Italy, having Italian parents, and so on* that refer to preconditions determining the *citizenship* and *permission to stay in Italy, the right to vote, etc.* that refer to further consequences derived by possessing the *citizenship*. From this points, he argues that the *intermediate legal concepts*, e.g., *citizenship*, play a role as mediator where *legal norms convey both legal consequences and preconditions of further legal effects* (Sartor 2009). In other words, he expresses that inferential semantics of an intermediate legal concept is fully determined by legal norms.

Sartor itemizes several implications from the idea that legal concepts have an inferential semantics. There are two implications that are relevant to our work:

- Legal concepts are determined by legal systems, which implies that legal meanings are determined by inferential connections. *Each legal norm using a concept contributes to characterizing the meaning of that concept, and different legal norms exist in different legal systems. Hence different systems have different concepts*; and
- *Legal semantics are determined (among other things) by legal doctrine, to the extent that doctrine determines. An inferential approach to the meaning of legal terms seems to entail that the understanding of legal concepts involves a doctrinal commitment with regard to the considered legal system.* (Sartor 2009)

Sartor also contrasts the legal concepts with Ramsey and Carnap's views on theoretical concepts (Ramsey 1991; Psillos 2000). Here, a *theoretical term* refers to non-observable entities such as *energy*. In Ramsey's theory, *theoretical concepts connect possible empirical data and forecasts of further empirical observations* (Boghossian 2003). According to Sartor, the same principle also applies to the legal concepts where *intermediate legal concepts are used within normative systems that connect observable facts and normative qualifications of human actions* (Sartor 2009). Sartor further applies Carnap's view of categories stating that *if there exists some category satisfying the concept's inferential links, then these links hold with regard the concept at issue* (Psillos 2000). These views of Ramsey and Carnap employed by Sartor are considerably relevant when arguing our approach based on the probabilistic model of theory formation by Kemp et al. (2010).

Finally, Sartor adds a notion of cognitive value of concepts as follows:

> *The mere applicability of a concept does not imply that the concept has cognitive value and thus deserves to be explicitly recognized and denoted by a specific term. A concept deserves such a recognition only to the extent that it collects a set of inferential links whose collection is cognitively useful, since it reflects the nature of the domain to be examined.* (Sartor 2009)

He contrasts this notion with the view of epistemology (Haack 2003; Peirce 2006) stating that *concepts should approximate generals or real types, meant as clusters of properties co-occurring because they are lawfully connected* and the view of Lindahl (2004) stating that *a legal concept should unify a bundle of normative links sharing not only the same legal effects, but also the same normative justification*.

Once again, his contrast implies that the study of legal concepts requires the approximation meant as clusters of common-causes and common-effects co-occurring, which can be well explained by the cognitive model explained in the next section.

3 A probabilistic model of theory formation

Kemp et al. (2010) employs the meaning of the term *theory* as *a system that specifies a set of concepts and relationships between these concepts* based on the views from psychology (Carey 1985) and artificial intelligence (Davis 1990). Kemp et al. (2010) emphasizes that *everyday knowledge is organized into intuitive theories that are similar to scientific theories in many respects* (Carey 1985; Murphy and Medin 1985). More specifically, Kemp et al. (2010) point out:

- *theories help to individuate concepts*;
- *many kinds of concepts derive their meanings from the roles they play in theories*;
- *theories allow learners to explain existing observations, and to make predictions about new observations*; and
- *theories guide inductive inferences by restricting a learner's attention to features and hypothesis that are relevant to the task at hand.* (Kemp et al. 2010)

Based on the aforementioned definitions, Kemp et al. (2010) categorize *theory* into two forms: a *framework theory* (Wellman and Gelman 1992) and a *specific theory*. While *framework theories specify the fundamental concepts that exist in a domain and the possible relationships between these concepts*, the specific theory is *a more detailed account of the phenomena in some domain and is typically constructed from concrete instances of the abstract categories provided by the framework theory*. Kemp et al. (2010) explain the relation between them with the domain of medicine. In their view, the framework theory indicates fundamental concepts, e.g., *chemicals*, *diseases* and *symptoms*. In this framework theory of medicine, chemicals cause diseases and diseases can cause symptoms. The specific theory, for example, indicates more specific instance of the abstract categories indicated by the framework theory, i.e., asbestos can cause lung cancer, and lung cancer causes coughing. Kemp et al. (2010) define that *framework theories can be represented as a probabilistic model which includes a set of categories and a matrix of parameters specifying relationships between those categories*.

Returning to the example of legal concepts given by Sartor (2009), the concept of *citizenship* is represented by preconditions that qualify *citizenship* and consequences which *citizenship* qualifies. From this points, Sartor argues that the *intermediate legal concepts*, e.g., *citizenship*, play as a mediator role where *legal norms convey both legal consequences and preconditions of further legal effects*. Contrasting Sartor's example with the aforementioned theory of medicine, the legal concepts and their inferential links can also be represented as the probabilistic model including a set of categories and a matrix, in order to structure a system of concepts that explain some existing set of preconditions and consequences. In other words, the probabilistic model (the IRM) identifies *underlying relationships* among fundamental concepts (abstract categories) and concrete instances of the abstract categories by analyzing co-occurrences of preconditions and consequences identified among specific instances. In this way, the model is expected to explain roles of intermediate legal concepts and position them in the system of concepts representing a specific legal system. These are our motivations for employing the IRM as a basis of our approach for dealing with the problems of legal inference.

4 Generalization and analogical inference

In cognitive sciences, Murphy (2004) illustrates the role of concepts in the following way:

> *concepts are the glue that holds our mental world together. (...) If we have formed a concept (mental representation) corresponding to that category (class of objects in the world), then the concept will help us understand and respond appropriately to a new entity in that category. Concepts are a kind of mental glue, then, in that they tie our past experiences to our present interactions with the world, and because the concepts themselves are connected to our knowledge structures.* (Murphy 2004)

As briefly mentioned in the previous section, the problem which we address in our work is to map legal concepts between two individual legal systems. Especially, our focus is on the issue of how a reasoner B belonging to the legal system B perceives, interprets and positions a legal concept belonging to the legal system A. In order to explain this complex issue, we employ Sartor's interpretation of Thagard's view describing cognitive roles of concepts (Thagard 1992). Sartor uses legal concepts such as *ownership*, *citizenship*, or *drunkenness* as example for explaining the cognitive roles of legal concepts as follows (Sartor 2009):

1. ***Categorization.*** *Things are categorized according to who owns them (for instance, by tax officers), and people are categorized according to their citizenship (by immigration officers) or are qualified as being drunk or sober (by police officers).*
2. ***Learning.*** *We learn the concepts of ownership, citizenship, or drunkenness when we are children, and if we become lawyers we refine and expand our understanding in law school, where we merge our intuitive understanding with knowledge of how such properties are legally determined, what they determine, and for what purpose.*

3. ***Memory.*** *We use these legal concepts for storing and synthesizing information, which may have been extracted from specific authoritative sources or experiences, though we cannot tell which ones.*
4. ***Inference.*** *We use the information that is linked to legal concepts for drawing inferences, for example, about the rights and duties of owners, citizens, and drunk people.*
5. ***Explanation.*** *We make explanations according to our conceptual model of the law. For instance, we may explain judicial decisions concerning certain people by pointing out the fact that these people are owners, citizens, or drunk drivers.*
6. ***Problem solving.*** *Our knowledge of what it means (for the law) to own something, to be a citizen, or to be drunk provides us with clues on how to approach situations where ownership, citizenship, or drunkenness are at issue.*
7. ***Generalization.*** *Our notions of ownership, citizenship, or drunkenness, enable us to consider at a glance all different situations falling under such concepts, perceiving their commonalities and differences.*
8. ***Analogical inference.*** *Our notions of ownership, citizenship, or drunkenness, enable us to make analogies. For instance, they enable us to speculate whether intellectual property or privacy rights over one's data are so similar to ownership that some of the ownership-related normative positions can be extended to them, or whether drug addiction may be likened to drunkenness.*

The aforementioned explanation of Sartor (2009) describes a situation where a reasoner B belonging to the legal system B acquires a new legal concept by generalizing a case/an instance observed in the legal district B. When considering a situation where the two legal districts A and B, e.g., Denmark and Japan, are negotiating a treaty regarding a specific domain, e.g., pension system, tax system, education system etc., the authorities in Denmark and Japan need to agree on how the legal concepts existing in the respective countries are mapped in order to facilitate the negotiation in question. Although the final destination is the mutual agreement between the two parties, the negotiation process involves generalizations and analogical inferences made by the two parties. The Danish authority interprets Japanese concepts in contrast with the Danish legal system and may even need to know how the Danish legal concepts are contrasted with the Japanese legal concepts based on the interpretation made by the Japanese authority. Thagard's view of cognitive roles explained in Sartor (2009) are also applicable to such situation involving two legal systems.

To be more specific, the mapping of legal concepts existing in the two legal systems can be explained in the following way. The reasoner B belonging to the legal system B is supposed to possess solid background knowledge of legal concepts in a specific domain (pension system, tax system, education system etc.) belonging to the legal district B. Hence, the reasoner B is able to generalize and interpret *a new legal concept A* introduced from the legal district A by comparing it with his or her background knowledge of legal concepts B belonging to the legal system B. The perception of commonalities and differences between them enables him or her to perform analogical inferences required to position the new concept A in the background knowledge of legal concepts belonging to the legal system B. This implies that, if we are able to define domain knowledge (set of legal concepts) existing in the respective legal districts and if we have a model that can consider commonalities and differences between them based on a legal reasoner's viewpoint, the model enables us to predict the interpretation made by a reasoner belonging to one of the two districts (either the Japanese or Danish authority in the aforementioned example).

Accordingly, we present a principle of mapping domain knowledge existing in two legal districts based on the interpretation possibly made by a reasoner belonging to one of the districts in the following sections. Our approach is closely related to the ontology mapping discipline. Hence we briefly review the principle of our approach in contrast to relevant works archived in the ontology mapping discipline in the next section.

5 Cross-categorization approach

In computer sciences, ontologies have often been employed for defining domain knowledge for the purpose of achieving common understandings among members of a specific knowledge community. When it comes to the interaction across communities, diverse methods for matching ontologies have been introduced in recent years such as (Cheng et al. 2008; Euzenat and Valtchev 2004; Ichise et al. 2004; Ehrig 2007; Mitra et al. 2005; Stumme and Mädche 2001).

Euzenat and Shvaiko (2007) define classification of ontology matching techniques from two viewpoints, i.e. *Granularity/Input interpretation* and *Kind of Input*. The *Granularity/Input interpretation* dimension is further classified according to the following criteria: *element-level vs. structure-level* and *syntactic vs. external vs. semantic*. Euzenat and Shvaiko (2007) state that *element-level* matching refers to techniques that compute correspondences between entities (objects) or instances (features) of those entities in isolation, while *structure-level* techniques considers how entities or their instances appear together in a structure. *Syntactic* approaches are, according to Euzenat and Shvaiko (2007), the techniques that interpret the input with regard to its sole structure based on a well-defined algorithm, while *external* techniques utilizes external resources of a domain and common knowledge for interpreting the input and *semantic* techniques employ some formal semantics for interpreting the input and justifying their results. From this aspect, we position our cross-categorization approach under the category of *structure-level techniques that employ a syntactic technique to interpret the input*. From the viewpoint of *Kind of Input*, Euzenat and Shvaiko (2007) define the classifications based on kind of data the algorithms handle, i.e. strings *(terminological)*, structure *(structural)*, models *(semantics)*, or data instances *(extensional)*. In this view, the algorithms employed in our approach deals with data instances, i.e. the algorithms employ *extensional* techniques that analyze *the actual population of an ontology* (Euzenat and Shvaiko 2007). The existing works that employ similar techniques to our approache are, e.g., (Berlin and Motro 2002; Bilke and Neumann 2005; Cheng et al. 2008; Doan et al. 2004; Euzenat 1994; Ichise et al. 2003, 2004; Lacher and Grog 2001; Stumme and Mädche 2001; Wang et al. 2004).

The traditional ontology matching is based on the prerequisite that well-organized and hierarchically-structured domain specific ontologies exist. Accordingly, the focus of the ontology matching (the *extensional* ontology matching techniques) is primarily on the relevancy analysis, i.e., similarity computation between concepts existing in two ontologies. The similarity computation usually employs algorithms that compute semantic distance between two concepts in question based on semantic information extracted from the existing ontologies. A work that is most relevant to our approach is the work presented by Cheng et al. (2008), which compares three similarity measures: Cosine similarity (Salton 1989); Jaccard similarity coefficient (Jaccard 1901); and market basket model (Hastie et al. 2001). The uniqueness of their work is to apply the market basket model to identify association rules between a concept in ontology A and a set of concepts in ontology B by computing conditional probabilities. While Cosine similarity and Jaccard similarity coefficient compute symmetric similarities, the market basket model results in asymmetric similarities between concepts existing in ontology A and in ontology B.

The framework introduced in our approach can accommodate these symmetric and asymmetric similarities, too. However, in this work we attempt to argue that our approach theoretically accommodate the aforementioned generalization and analogical inferences pointed out by Sartor (2009). More specifically, one of the interesting natures of our approach is that it tries to explain a situation where a reasoner generalizes and interprets a new legal concept introduced from an alien legal district by comparing it with his or her background knowledge of legal concepts belonging to his or her own legal system. This function of analogical inference is, in this work, tested by employing the Bayesian Model of Generalization (BMG) proposed by Tenenbaum and Griffiths (2001). The BMG has been used for modeling concept learning problem (Tenenbaum and Griffiths 2001), but also its principle has been used for inductive reasoning (Kemp and Tenenbaum 2009) and inductive generalization model (Kemp et al. 2012). The BMG is, in a way, the extended version of Tversky's Ratio Model (Tversky 1977) which computes similarities based on commonalities and differences of features possessed by concepts. In the ontology matching discipline, Tversky's Ratio Model is employed in, e.g., Huang and Kuo (2010); de Souza and Davis (2004). While in Tversky's model the feature weights is assigned heavier to features possessed by the new concept (i.e., object subject to comparison), the BMG considers features perceived only by a reasoner. To be more specific, the BMG computes similarities by dividing common features (features possessed by both the new concept and a referent concept) with the sum of the common features and features of the referent concept that is known by a reasoner. This is quite different from the traditional Jaccard similarity coefficient where the sum of common features and features of both the new and the referent concepts are considered as denominator. In addition to this, the BMG distinguishes importance of features by assigning weights to each feature. In other words, all features appearing in the BMG equation explained in the next section are weighted based on the degree of importance computed by a principle where a feature possessed by fewer concepts within the knowledge possessed by the reasoner in question is weighted higher, and vice versa. Thus, the weights are supposed to provide clearer indications of what concepts are more similar, and vice versa, compared to the Jaccard similarity coefficient and Tversky's model.

Another uniqueness of the proposed approach is that our method rather addresses the problem of identifying latent ontological structures from two independent semi- or unstructured data sources while analyzing interactive relations between the respective domain knowledge in question. Hence our approach can be distinguished from the traditional ontology matching approaches which primarily address issues on integrating already existing ontologies.

This paper introduces our approach, i.e., the cross-categorization approach, for aligning similar semi-structured domain knowledge existing in two heterogeneous legal systems (i.e., the educational systems belonging to legal districts, Japan and Denmark). We employ the Jaccard similarity coefficient and the BMG for computing the degree of relations between all possible combination of concepts existing in the two legal systems. To the obtained similarity scores, we apply an extended version of the IRM, the so-called *normal Infinite Relational Model* (n-IRM) proposed by Herlau et al. (2012) in order to cross-categorize the concepts existing in the two legal systems. The application of the n-IRM expects to identify number of categories, i.e., groups of concepts for the respective legal systems, and to compute the degree of the relations between the categories in the two legal systems. Finally, the IRM is applied to the original data consisting of legal concepts and their features (inferential links), in order to identify underlying relationships, i.e., structures of inferential links, behind the specified concept system, i.e., the identified categories. The workflow of our approach is depicted in Fig. 1.

In this way, the proposed approach simultaneously categorizes legal concepts existing in two legal systems and from there to structure two independent concept systems that are inter-operable in the most efficient manner. By employing the generalization model, i.e., the BMG, we argue that our approach theoretically explains the inferential mechanism of legal concept mapping where a reasoner B generalizes and interprets *a new legal concept A* introduced from the legal district A by comparing it with his or her background knowledge of legal concepts B belonging to the legal system B. To sum up, what our proposed approach expects to achieve are: i) the BMG is, from the reasoner B's viewpoint, identifies causes and effects in the new concept A which co-occur with a referent concept B; ii) the BMG distinguishes more important- and less important causes and effects by assigning weights in each feature; iii) the degree of relations between all possible combinations of

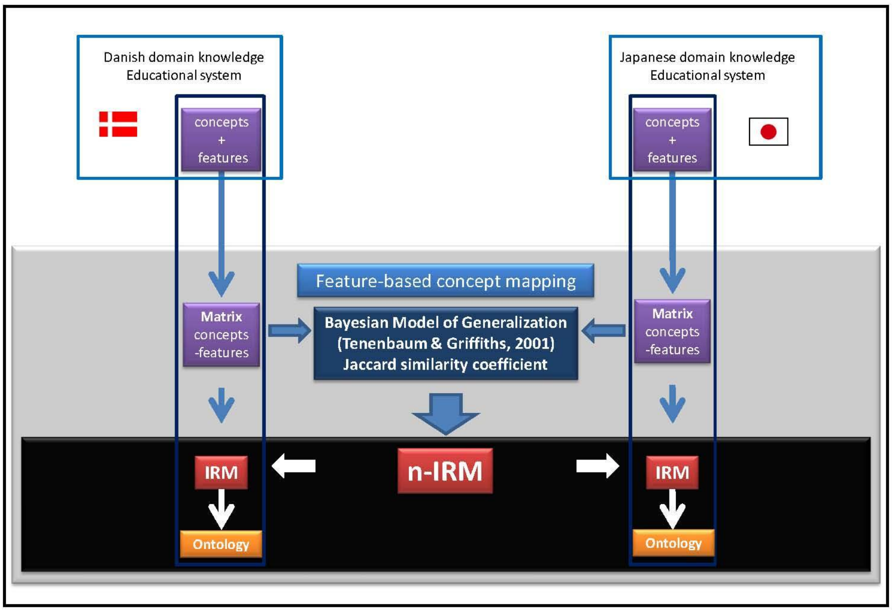

Fig. 1: Flowchart of the cross-categorization approach.

concepts between the two legal systems are cross-categorized by the n-IRM; iv) the n-IRM is supposed to bundle groups of concepts having higher degree of relations between the two systems as categories; v) by fixing the obtained categories, the IRM is supposed to analyze what causes and effects are commonly shared by members of each obtained category; and vi) the structure of commonly shared causes and effects obtained by the IRM is visualized as a hierarchical graph (ontology) [1]. In the following sections, we review each element of our approach in more details.

6 Bayesian Model of Generalization

The Bayesian Model of Generalization (BMG) (Tenenbaum and Griffiths 2001) can be used to measure the similarity of objects based on their features, as discussed in previous work (anonymous) (). Tenenbaum and Griffiths (2001) state that the BMG is considered as an extension of Tversky's set-theoretic model of similarity. Tversky's model can be formulated as

$$\mathrm{sim}(y, x) = \frac{|Y \bigcap X|}{|Y \bigcap X| + \alpha|Y - X| + \beta|X - Y|}. \tag{1}$$

[1] Inspired by our original work presented this paper, we have applied our approach to two identical datasets for constructing a hierarchical graph reprsenting a single knowledge system. The extended work will be presented in (anonymous) ()

Equation 1 computes the degree of similarity between two feature sets X and Y relating to objects x and y respectively. The similarity is computed by dividing the number of common features between the two objects $|Y \cap X|$ with the sum of three terms: the number of common features $|Y \cap X|$; the number of features that exist in y but not in x multiplied by a factor α, $\alpha|Y - X|$; and the number of features that exist in x but not in y multiplied by a factor β, $\beta|X - Y|$. The similarity scores fall in the range between 0 and 1. The free parameters α and β can arbitrary be defined (Tversky 1977). For example, when the parameters are set as $\alpha = \beta = 1$, Equation 1 corresponds to Jaccard's similarity measure (Jaccard 1901). We employ Jaccard's measure as benchmark indicating symmetric similarity relations between legal concepts. Tversky's argument is that *if $sim(y, x)$ is interpreted as the degree to which y is similar to x, then y is the subject of the comparison* (a new concept which a reasoner is suppose to learn from outside of the legal district) *and x is the referent* (knowledge possessed by a reasoner). Hence in his view, the features of the new concept get heavier weights than those of the referent concept (i.e., $\alpha > \beta$). In contrast with Tversky's argument, Tenenbaum and Griffiths (2001) claim that, under the condition where the parameters are set as $\alpha = 0$ and $\beta = 1$, Tversky's model corresponds to Equation 2 of the BMG (Tenenbaum and Griffiths 2001). This implies that the BMG computes similarities by dividing a number of common features (features possessed by both a new concept y and a referent concept x) with the sum of the common features and features of the referent concept x that is known by a reasoner.

$$P(y \in C|x) = \left[1 + \frac{\sum_{h:x \in h, y \notin h} P(h, x)}{\sum_{h:x,y \in h} P(h, x)}\right]^{-1}. \quad (2)$$

From the probabilistic viewpoint, Equation 2 computes conditional probabilities that a new concept y (legal concept existing outside the reasoner's legal district) comes to belong to an undefined *categorical region C* provided by the condition of a referent concept x already known by the reasoner. In the definition in (Tenenbaum and Griffiths 2001), a hypothesized subset h is a region *where a concept belongs to h if and only if it possesses a feature k*. Hence, $\sum_{h:x,y \in h} P(h, x)$ in Equation 2 corresponds to a sum of $P(h, x)$ where a k_i feature is commonly possessed by a referent concept x and the new concept y, while $\sum_{h:x \in h, y \notin h} P(h, x)$ refers to a sum of $P(h, x)$ where a k_i feature is possessed only by a referent concept x. In other words, Equation 2 explains a scenario where the reasoner compares this new object y with an already known concept x which is part of his/her background knowledge attached to the legal district he/she belongs to. The function $P(h, x)$ in Equation 2 is considered as a feature weight assigned to each k_i feature and defined as $P(h, x) = P(x|h)P(h)$ where we consider the prior $P(h)$ as $P(h) = 1$ (the prior $P(h)$ can accommodate *arbitrary flexibility* across contexts according to Tenenbaum and Griffiths (2001)). $P(h, x) = P(x|h)P(h)$ is therefore influenced by the computation of

$P(x|h)$ defined as:

$$P(x|h) = \begin{cases} \frac{1}{|h|} & \text{if } x \in h, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

In this work, $|h|$ in Equation 3, i.e. the size of the region h, is computed based on a number of concepts possessing a specific feature k_i within the reasoner's entire domain knowledge. To simplify the explanation of Equation 3, the weight $P(h, x) = P(x|h)P(h)$ is assigned to every features considered in Equation 2. Since the prior is set as $P(h) = 1$, the weight $P(h, x)$ is directly reflected by the size of the region h in Equation 3, i.e., a feature possessed by fewer objects gets higher weights, and vice versa.

7 Relational Models

In our work we use two types of relational models: i) To co-cluster the similarity scores between two sets of concepts we use the normal Infinite Relational Model (n-IRM) Herlau et al. (2012). In the n-IRM analysis, the data consists of the similarity scores between concepts, which are jointly clustered in each legal system. The output of the analysis is two mutually compatible clusterings of the concepts in each system. ii) To analyze the relation between the extracted concept clusters and their features (inferential links) we use the Infinite Relational Model (IRM) Kemp et al. (2006) to cluster features based on the concept clusters extracted by the n-IRM.

Both the IRM Kemp et al. (2006) and its variant the n-IRM Herlau et al. (2012) are able to automatically infer the number of clusters by using a distribution that is defined for all conceivable partitions based on the Chinese Restaurant Process (CRP) (Aldous 1985; Pitman 2002). Let $\boldsymbol{z}$ denote the partitioning of entities to clusters, then the distribution over partitions formed by the CRP is given by

$$P(\boldsymbol{z}|\gamma) = \frac{\gamma^K \Gamma(\gamma)}{\Gamma(J+\gamma)} \prod_{a=1}^{K} \Gamma(n_a), \tag{4}$$

where K is the number of clusters for the given partition $\boldsymbol{z}$, $\Gamma(x) = (x-1)!$ is the Gamma function, n_a the size of cluster a, and J the total number of entities (i.e. concepts or features). γ is a parameter that is used to specify how favorably it is to form many as compared to few clusters. The above distribution can be derived by the following analogy to a Chinese Restaurant where J customers are to be seated successively such that the first customer sits at the first table and each subsequent customer sits at occupied tables proportionally to how many are already seated at the table and chooses a new table at the restaurant proportional to γ. Let n_a denote how many are seated at table a. According to this process the i^{th} customer will sit at a new table with probability $\frac{\gamma}{i-1+\gamma}$ and at the occupied table a with probability $\frac{n_a}{i-1+\gamma}$. By generating partitions according to this procedure the above distribution is recovered. As the distribution only depends on the size of the clusters (and not the order in which the customers were seated) the distribution is exchangeable (Aldous 1985; Pitman 2002).

7.1 Co-clustering similarity scores by the normal Infinite Relational Model

The n-IRM model proposed in (Herlau et al. 2012) is an extension of the IRM Kemp et al. (2006) designed to co-cluster real valued relational data. For similarity scores R_{ij} defining the similarity between the i^{th} concept in the one legal system (i.e., Japanese educational system) to the j^{th} concept of the other legal system (i.e. Danish educational system) the n-IRM clusters the Japanese and the Danish concepts jointly by partitions induced by mode (i.e. legal system) specific CRP's. The similarity between the extracted groups in the Japanese and Danish systems are then parameterized by a mean intensity m_{ab} with precision (i.e., inverse variance) λ_{ab} such that the real valued similarity $R_{ij} \sim \text{Normal}(m_{z_i^{(1)} z_j^{(2)}}, \lambda^{-1}_{z_i^{(1)} z_j^{(2)}})$, where $z_i^{(1)}$ and $z_j^{(2)}$ denotes the groups that the i^{th} Japanese and j^{th} Danish concepts respectively are assigned. To simplify the inference in the model conjugate Gamma and Normal priors are invoked for the mean intensities and precisions leading to an efficient sampling procedure in which these parameters are analytically integrated (i.e., collapsed) such that inference in the model reduces to sampling the assignment of concepts to clusters in each system. To summarize, the n-IRM presently considered for co-clustering similarity scores is defined by the following generative process (Herlau et al. 2012)

$$
\begin{aligned}
& & \boldsymbol{z}^{(1)} &\sim \text{CRP}(\gamma^{(1)}), && \text{Japanese concept clusters,} \\
& & \boldsymbol{z}^{(2)} &\sim \text{CRP}(\gamma^{(2)}), && \text{Danish concept clusters,} \\
&\text{for all clusters,} & \lambda_{ab} &\sim \text{Gamma}(\alpha_0, \text{rate} = \beta_0), && \text{precision,} && (5) \\
&\text{for all clusters,} & m_{ab} &\sim \text{Normal}\left(m_0, (\kappa_0 \lambda_{ab})^{-1}\right), && \text{mean,} && (6) \\
&\text{for all objects,} & R_{ij} &\sim \text{Normal}(m_{z_i^{(1)} z_j^{(2)}}, \lambda^{-1}_{z_i^{(1)} z_j^{(2)}}), && \text{links (real valued).} && (7)
\end{aligned}
$$

7.2 Analyzing the relationship between the concept clusters and their features by the Infinite Relational Model

To further investigate the relationship between the concept clusters extracted by the n-IRM and their features in each of the legal systems we employ the IRM to the concept by feature binary matrix of legal system c given by $A_{ij}^{(c)}$ where $A_{ij}^{(c)} = 1$ if concept i posses feature j and $A_{ij}^{(c)} = 0$ otherwise. We fix the concept clusters extracted by the n-IRM (i.e., $\boldsymbol{z}^{(c)}$) and investigate the relationship between concept clusters and their features by clustering only features by the IRM. Following the IRM Kemp et al. (2006) links are drawn from a Bernoulli distribution parameterized by the strength of interaction $\eta_{ab}^{(c)}$

that specifies the probability of observing links between concepts in concept cluster a and features in feature cluster b. Thus, our clustering of features $\boldsymbol{z}^{(f_c)}$ in each legal system c according to the concept clusters identified by the n-IRM is defined by the following generative process

$$\mathbf{z}^{(f_c)} \sim \mathrm{CRP}(\gamma^{(f_c)}), \qquad \text{clustering of features,} \quad (8)$$

$$\text{for all clusters,} \quad \eta_{ab}^{(c)} \sim \mathrm{Beta}(\beta_0^+, \beta_0^-), \qquad \text{strength of interactions,} \quad (9)$$

$$\text{for all objects,} \quad A_{ij}^{(c)} \sim \mathrm{Bernoulli}\left(\eta^{(c)}_{z_i^{(c)} z_j^{(f_c)}}\right), \qquad \text{links (binary valued).} \quad (10)$$

As the Beta distribution is conjugate to the Bernoulli distribution the strengths of interactions can be analytically integrated (i.e., collapsed) and inference reduces to sampling the assignment of features to clusters $\boldsymbol{z}^{(f_c)}$.

For inference in the n-IRM and IRM we used the sampling procedures described in Mørup et al. (2010); Herlau et al. (2012) with a total of 1000 iterations where we discarded the first 500 samples to allow the n-IRM and IRM samplers to burnin. We set the hyper-parameters $\gamma^{(\alpha)} = \log(J_\alpha)$ where J_α is the size of mode α, and for the n-IRM we set $\kappa_0 = 1$, $\alpha_0 = 15$ and $\beta_0^{-1} = 10$ whereas for the IRM $\beta_0^+ = \beta_0^- = 1$. The solutions used for displaying the results in the following sections are given by the realization identified with highest value of the joint probability density. The displayed mean intensities m_{ab} and standard deviations $1/\sqrt{\lambda_{ab}}$ are in the following sections calculated disregarding the priors and are thereby defined by the mean and standard deviation of each of the extracted blocks of the n-IRM.

8 Data source

This work employs a database prepared by a third party, the UNESCO Institute for Statistics (UIS) as the data source. The database consists of datasets that have been reported by each UNESCO Member States. The reported datasets are in accordance with the International Standard Classification of Education (ISCED), which enables the contrast of educational systems across different Member States. The datasets collected from all over the world are downloadable from the UIS web-site[2]. In this work, we use datasets representing the Japanese and the Danish educational systems. As shown in Fig 2, the datasets consist of educational terms (concepts) defined by several pre-defined feature dimensions such as *"the ISCED level"*, *"programme destination and orientation"*, *"starting age"*, *"cumulative duration of education"*, and *"entrance requirements"*. For each pre-defined feature dimension column in Fig 2, feature dimension values are displayed such as (A, B, C) for the *"programme destination"* dimension and (G, V) for the *"programme orientation"* dimension. A matrix for each educational system is made in a way that educational

[2] http://www.uis.unesco.org/education/ISCEDmappings/Pages/default.aspx

Concept ID	ISCED level	Programme destination (A/B/C)	Programme orientation (G/P/V)	Theoretical cumulative duration at ISCED 5	Position in the national degree / qualification structure (intermediate, first, second, etc...)	Position in the tertiary education structure (Bachelor-Master-PhD)	Minimum entrance requirement (ISCED level or other)	Theoretical starting age	Theoretical duration of the programme	Theoretical cumulative years of education at the end of the programme	Does the programme have a work based element? Y/N	Programme specifically designed for adults Y/N	Programme specifically designed for part-time attendance Y/N
Danish concepts													
D1	0		G					2-5 y.o	4 years		No	No	No
D2	0		G					5-6 y.o	1 year		No	No	No
D3	1		G					6-7 y.o	6 years	6 years	No	No	No
D4	2	A	G				1	12-13	3-4	9-10	No	No	No
D7	3	C	V				2A	16-30	3-5	14	Yes	No	No
D19	5	B		Short	1st		3A, 3C	18-50	0,5-4	13-15	No	Yes	Yes
D20	5	B		Short	1st		3A, 3C	20-30	2-3	14	No	No	No
D21	5	A		Medium	1st	Bachelor	3A	18-50	2-4	13-15	No	Yes	Yes
D22	5	A		Medium	1st	Bachelor	3A	20-30	3-5	16	Yes	No	No
D23	5	A		Medium	1st	Bachelor	3A	20-30	3	15-16	No	No	No
Japanese Concepts													
J35	5	B		Short	Intermediate		3 ABC	18	2-3	14-15	No	No	No
J36	5	B		Short Medium	Intermediate		5B	20	1+	15+	No	No	No
J37	5	B		Short	Intermediate		3	18	2-3	14-15	No	No	Yes
J38	5	B		Short	Intermediate		3	18	2	14	No	No	No
J40	5	B					3	18	1+	13+	No	No	No
J41	5	A		Medium	1st	Bachelor	3	18	4	16	No	No	No
J42	5	A		Long	1st	Bachelor	3	18	6	18	No	No	No
J44	5	A		Long	Intermediate		5A 1st,M	22	1+	17+	No	No	No

Fig. 2: Example of UIS database

terms and feature dimension values are respectively listed in rows and columns. One difficulty in creating a matrix is the treatment of the numeric feature values refering, e.g., *"theoretical starting age"* and *"cumulative duration of the programme"* in Fig 2. For example, *"theoretical cumulative years of education at the end of the programme"* of Danish term $D23$ in Fig 2 has a value [15-16 years]. On the other hand, Japanese term $J35$ has a value [14-15 years]. In order to deal with this problem, the procedure defined in our previous works (anonymous)() is employed:

1. If a feature value in one dataset is part of a feature value in the other dataset (e.g., the feature [14-15 years] in the Japanese dataset is part of the feature [13-15 years] in Denmark), a concept having the feature that covers the other feature (a concept having [13-15 years]) should also have a feature [14-15 years], and
2. If two features from the respective countries partly overlaps (e.g., [14-15 years] in the Japanese dataset and [15-16 years] in the Danish dataset), a pseudo feature referring to the exact overlapping range (i.e., [15 years]) is created. In this example, a Japanese concept having the feature [14-15 years] should also have the pseudo feature [15 years]. In the same way, a Danish concept having the feature [15-16 years] should also have the pseudo feature [15 years].

Based on this procedure, a Japanese matrix consisting of 54 Japanese objects and 113 features and a Danish matrix consisting of 27 Danish objects and 113 features are created. In these matrices, the educational terms and their characteristic features are respectively considered as legal concepts and inferential links (i.e., Q-conditioning links or Q-conditioned links) existing in each of the two legal systems (Japan and Denmark) as explained in the previous sections. In each matrix, a link between a feature and a concept possesing the feature is considered as *true* value. In other words, a feature vector considered in Section 6 is a set of features having *true* values in each row of the binary matrices, representing each educational concept.

9 Implementation of the cross-categorization approach

Fig. 3 depicts an overview of the results obtained from the n-IRM applied to similarity scores computed by Jaccard similarity coefficient and the BMG. The three plots (1-a; 1-b; 1-c) in the upper row illustrate similarity scores computed in all combinations of concepts between the Japanese- and the Danish educational systems, while the three plots (2-a; 2-b; 2-c) in the second row show the cross-categorization results obtained by the n-IRM. The plots (3-a; 3-b; 3-c) and the plots (4-a; 4-b; 4-c) in Fig. 3 respectively contrast mean values and standard deviation of each intersection of clusters in the three plots (2-a; 2-b; 2-c) obtained by the n-IRM.

The three columns from the left to the right in Fig. 3 illustrate the results respectively obtained by the following data:

1. First column (1-a; 2-a; 3-a; 4-a): similarity scores computed by Jaccard similarity coefficient;
2. Second column (1-b; 2-b; 3-b; 4-b): similarity scores computed by the BMG when the Japanese educational concepts are set as reasoner's background knowledge; and
3. Fourth column (1-c; 2-c; 3-c; 4-c): similarity scores computed by the BMG when the Danish educational concepts are set as reasoner's background knowledge;

The plot (1-a) depicts the results obtained from Jaccard similarity coefficient. Intuitively, the plot is darker and more even than the other two plots (1-b and 1-c). The cross-categorization results of the Jaccard scores in 2-a (i.e., the number of clusters obtained) are coarse especially in the Japanese concepts. The mean values (m_{ab}) in the plot (3-a) are also darker and more even compare to the other two plots (2-b and 2-c). The standard deviations ($1/\sqrt{\lambda_{ab}}$) in the plot (4-a) are relatively uniform and the intersections of clusters are dominated by gray colores compared to the other plots (4-b and 4-c) in Fig. 3.

On the other hand, the results obtained from the BMG shown in the plots 1-b and 1-c are substantially different, since the BMG computes similarities by dividing common features (features possessed by both a new concept and a referent concept) with the sum of the common features and features of the referent concept that is known by a reasoner, then weights features of concepts existing in reasoner's background knowledge. These differences in similarity scores clearly influence the cross-categorization results obtained by the n-IRM. The number of clusters obtained in 2-b is 13 for the Japanese system and 12 for the Danish system, while 8 for the Japanese and 9 for the Danish in 2-c. The obtained numbers of clusters are correlative to the number of concepts that are considered as reasoner's background knowledge. To be more specific, when the Japanese educational system consisting of 54 educational concepts is considered as reasoner's background knowledge in 2-b, the number of the obtained clusters are larger compared to the situation where the Danish educational system consisting of 27 educational concepts is considered as reasoner's background knowledge in 2-c.

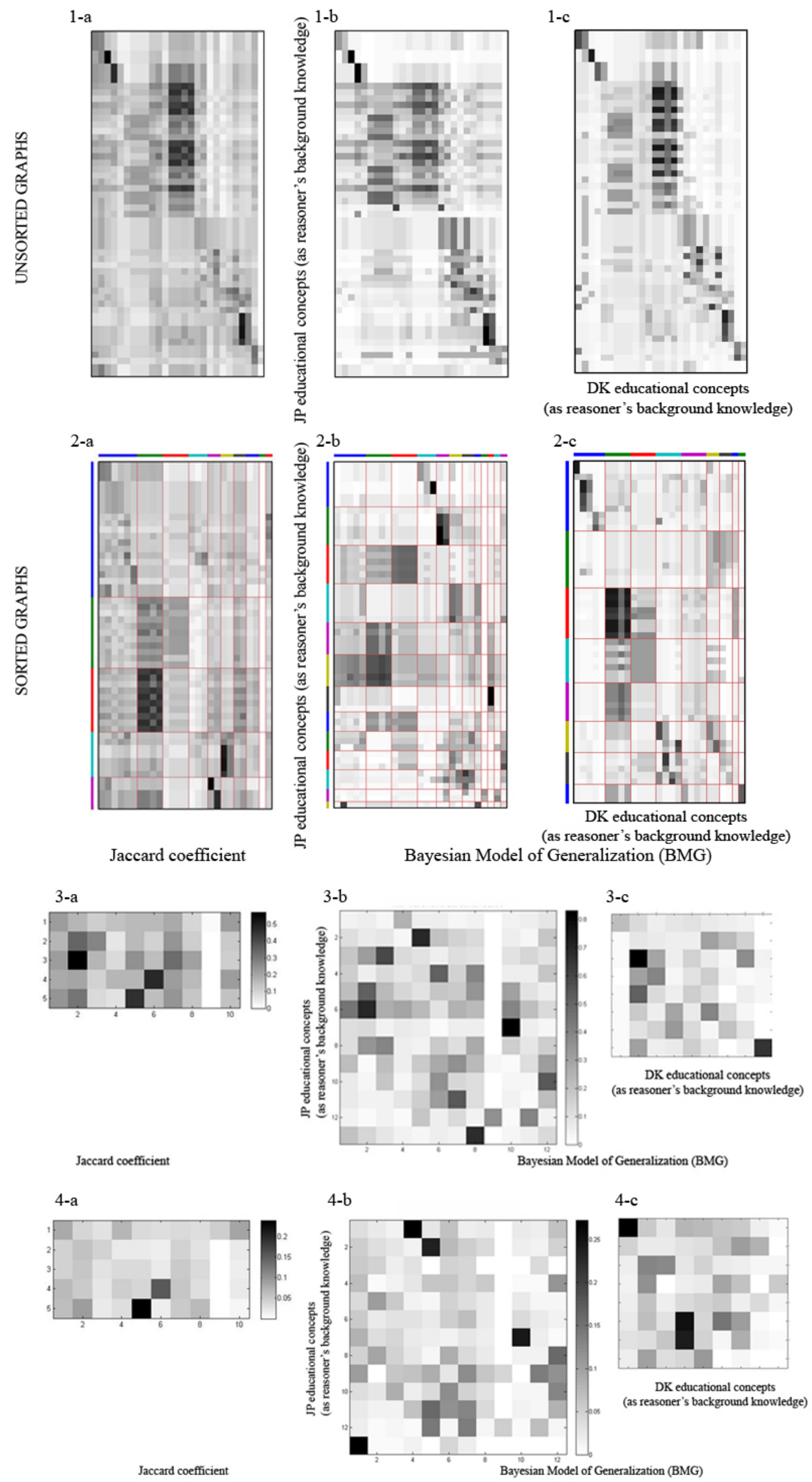

Fig. 3: Results of cross-categorization obtained by the n-IRM

The differences in the obtained number of clusters in 2-b and 2-c are likely caused by the distributions of similarity scores. When the 54 Japanese concepts are considered as reasoner's background knowledge, the similarity computation is based on the features possessed only by these 54 concepts. Hence, differentiations in the similarity scores are stronger across the 54 Japanese concepts rather than across the 27 Danish concepts. This phenomenon is identified in 1-b where the horizontal lines are more visible compared to 1-c. Accordingly, the partition of the Japanese concepts results in fine-grained clusters, which also affect the partition of the Danish concepts in 2-c by the n-IRM. In the same way, when the Danish educational system is considered as reasoner's background knowledge, differentiations in the similarity scores are stronger across the 27 Danish concepts, which can be seen by the stronger vertical lines in 1-c. Thus, the partition of the Danish domain knowledge consisting of only 27 concepts affects the partition of the 54 Japanese concepts in 2-c, which results in fewer number of clusters.

The plots from 3-a to 3-c of Fig. 3 show mean values (m_{ab}) of the intersection of the clusters obtained from the n-IRM. The gray scale indicates that, when a cluster is close to black, a mean value is close to one, and vice versa. While the plots 3-a from the Jaccard scores are more uniform, i.e. the majority of clusters are gray colored, the gray colors in the plots 3-b and 3-c are more differentiated due to the weight assigned to each feature during the computation of the conditional probabilities (i.e. similarities) in the BMG. Hence, the interactivity between Japanese- and Danish clusters is clearly explained with the results obtained from the BMG, whereas it is rather ambiguous with the results obtained from the Jaccard scores.

The standard deviations $(1/\sqrt{\lambda_{ab}})$ shown in the plots from 4-a to 4-c of Fig. 3 explain the uniformity within each intersecton between clusters. If an intersection of two clusters is completely uniform, the intersection is indicated with the white color, and vice versa. The plots in Fig. 3 does not indicate substantial differences in the gray color distributions among the three plots (4-a; 4-b; 4-c). Only implication identified here is that, when the clusters are fine-grained, the proportion of the light gray colored clusters are slightly dominant as is indicated in 4-b. This also implies that the fine-grained uniform clusters obtained from the BMG may potentially be more effective for interactively uncovering latent hierarchical structures of respective domain knowledge.

	BMG JP		BMG DK		Jaccard	
	Japanese concept clusters	Danish concept clusters	Japanese concept clusters	Danish concept clusters	Japanese concept clusters	Danish concept clusters
NMI (means)	0.86	0.93	0.8	0.86	0.94	0.96
NMI (SD)	0.02	0.02	0.02	0.03	0.04	0.02

Fig. 4: Stabilities of the cross-categorization performances

Finally, we have quantified the stabilities of the cross-categorization performances in Fig. 4. The stabilities of the cross-categorization performances are computed by a commonly used measure called Normalized Mutual Information (NMI) (Danon et al. 2005) . The NMI indicates a number between 0 and 1 such that 1 describes identical assignments to clusters. The results generally indicate that the performances achieved by the n-IRM are stable.

10 Qualitative assessment of clusters

Figs. 5, 7 and 9 all illustrate results obtained from the cross-categorization approach. Each Fig. consists of three plots. The plots at the upper-left corner of Figs.5, 7 and 9 respectively correspond to the n-IRM plots 2-a, 2-b and 2-c in Fig. 3, i.e., the n-IRM results computed based on the scores obtained by the Jaccard similarity coefficient, by the BMG when the Japanese educational concepts are set as reasoner's background knowledge, and by the BMG when the Danish educational concepts are set as reasoner's background knowledge. In each of the n-IRM plots, concept cluster IDs for the Japanese and the Danish educational systems are assigned as $[J1, J2, ...J_i]$ and $[D1, D2, ...D_j]$. The plots at the right side of the n-IRM plots represent the feature structures of the Japanese educational concept clusters $[J1, J2, ...J_i]$. In other words, the plots represent the structures of inferential links (causes and effects) that are commonly shared by each Japanese concept cluster J_i. The obtained feature clusters are displayed as $[Jf1, Jf2, ...Jf_j]$. The plots at the bottom of the n-IRM plots refer to the feature structures of the Danish educational concept clusters $[D1, D2, ...D_j]$. The obtained feature clusters for the Danish educational system are displayed as $[Df1, Df2, ...Df_j]$.

Tables in Figs. 6, 8 and 10 respectively correspond to the graphical results shown in Figs. 5, 7 and 9. More specifically, the numbers in the upper-left tables respectively corresponding to the n-IRM plots are the mean values (m_{ab}) of each intersection connecting a Danish concept cluster and a Japanese concept cluster. The numbers in the upper-right tables corresponding to the Japanese feature structures are the $\boldsymbol{\eta}$ values of each intersection connecting a concept cluster $[J1, J2, ...J_i]$ and a feature cluster $[Jf1, Jf2, ...Jf_i]$ computed by the IRM. In the same way, the numbers in the bottom-left tables corresponding to the Danish feature structures are the $\boldsymbol{\eta}$ values of each intersection connecting a concept cluster $[D1, D2, ...D_i]$ and a feature cluster $[Df1, Df2, ...Df_i]$. Figs. 18, 20 and 22 in Appendix I of this paper list the members of the Japanese- and the Danish legal concept clusters $[J1, J2, ...J_i]$ and $[D1, D2, ...D_i]$. Figs. 19, 21 and 23 list members of the feature clusters. The members of each feature cluster have common feature IDs $[f1, f2, ...f113]$ in all Figs. 19, 21 and 23.

These results demonstrate interesting differences among the similarity measures employed in this work. In Figs. 5 and 6 representing the results based on the Jaccard similarity scores, a Japanese concepts *"junior college - regular course"*, *"junior college - correspondence course"*, *"college of technology - reg-*

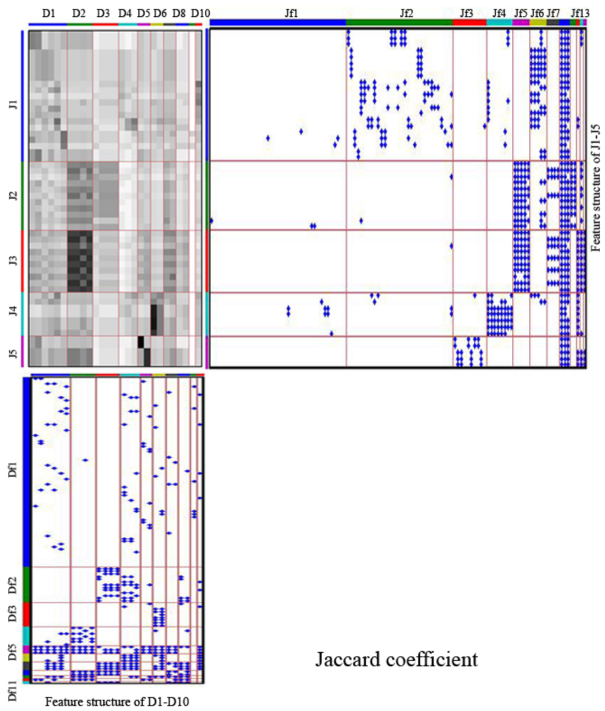

Fig. 5: n-IRM clustering and feature structure: symmetric relations

Avr. 0.18	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Jf1	Jf2	Jf3	Jf4	Jf5	Jf6	Jf7	Jf8	Jf9	Jf10	Jf11	Jf12	Jf13
J1	0.22	0.17	0.11	0.14	0.16	0.15	0.19	0.09	0.02	0.22	0.05	0.08	0.01	0.39	0.01	0.01	0.94	0.02	0.55	0.04	0.17	0.17	0.13
J2	0.17	0.35	0.27	0.06	0.15	0.11	0.23	0.11	0.00	0.12	0.01	0.01	0.01	0.01	0.91	0.28	0.94	0.88	0.42	0.08	0.62	0.15	0.08
J3	0.22	0.57	0.17	0.09	0.24	0.14	0.32	0.21	0.00	0.11	0.00	0.01	0.01	0.02	0.94	0.50	0.94	0.05	0.05	0.92	0.92	0.17	0.92
J4	0.19	0.20	0.11	0.17	0.16	0.48	0.22	0.16	0.00	0.22	0.02	0.79	0.03	0.09	0.03	0.03	0.87	0.06	0.19	0.11	0.22	0.33	0.89
J5	0.24	0.31	0.08	0.07	0.46	0.15	0.24	0.13	0.00	0.14	0.00	0.02	0.50	0.03	0.04	0.05	0.94	0.08	0.08	0.57	0.57	0.14	0.57
Df1	0.07	0.01	0.00	0.09	0.08	0.05	0.05	0.01	0.10	0.07													
Df2	0.01	0.02	0.52	0.39	0.07	0.04	0.04	0.25	0.07	0.20													
Df3	0.02	0.03	0.03	0.07	0.10	0.70	0.05	0.10	0.09	0.18													
Df4	0.11	0.50	0.03	0.04	0.06	0.13	0.06	0.13	0.11	0.11													
Df5	0.95	0.93	0.57	0.45	0.88	0.88	0.88	0.38	0.20	0.80													
Df6	0.25	0.07	0.07	0.91	0.13	0.63	0.13	0.13	0.20	0.80													
Df7	0.50	0.07	0.93	0.91	0.13	0.75	0.50	0.88	0.20	0.80													
Df8	0.07	0.90	0.90	0.13	0.17	0.17	0.67	0.67	0.25	0.25													
Df9	0.50	0.83	0.17	0.60	0.50	0.75	0.75	0.75	0.33	0.33													
Df10	0.50	0.83	0.17	0.20	0.75	0.25	0.50	0.75	0.33	0.33													
Df11	0.25	0.17	0.83	0.20	0.25	0.25	0.75	0.75	0.33	0.33													

Intersection between Japanese legal concepts (J1-J5) and Danish legal concepts (D1-D10)	
If mean value is equal or over 0.4	0.43
If mean value is equal or over 0.3	0.34
If mean value is equal or over 0.2	0.25
If mean value is equal or over average of all mean values	0.12
Japanese or Danish feature structures for each legal concept cluster	
If likelihood is equal of over 0.5	0.50

Fig. 6: n-IRM clustering and feature structure: symmetric relations (Jaccard: mean and $\boldsymbol{\eta}$ values)

ular course" are included in a very big concept cluster *J1* and does not have any strong relations with specific Danish concept clusters.

In Figs. 7 and 8 representing the results based on the BMG, two Japanese concept clusters *J10* (consisting of *"junior college - regular course"*, *"junior college - correspondence course"*, *"college of technology - regular course"*) and *J11* (consisting of *"university undergraduate"*, *"university undergraduate - medicine etc."*, *"university undergraduate - correspondence course"*), respectively have stronger relations with *D12* (consisting of *"short cycle tertiary edu-*

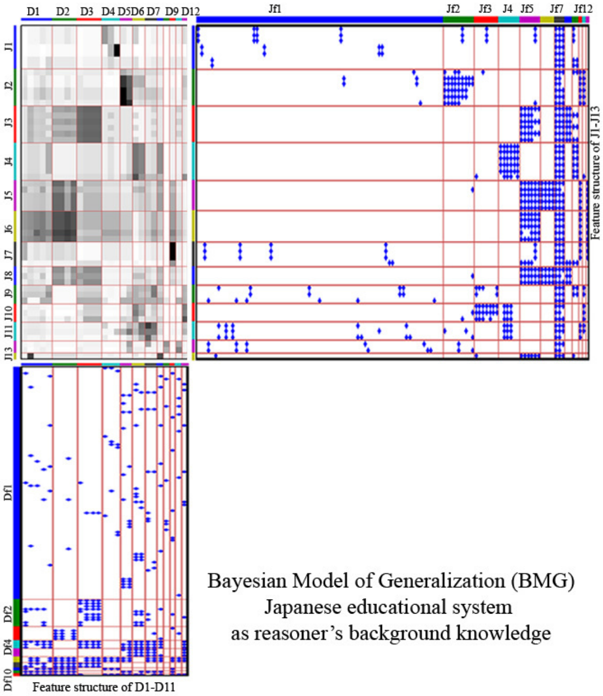

Fig. 7: n-IRM clustering and feature structure: JP as background knowledge

Avr. 0.15	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	Jf1	Jf2	Jf3	Jf4	Jf5	Jf6	Jf7	Jf8	Jf9	Jf10	Jf11	Jf12
J1	0.05	0.06	0.03	0.26	0.06	0.05	0.04	0.05	0.00	0.05	0.06	0.08	0.04	0.06	0.08	0.02	0.09	0.03	0.96	0.07	0.47	0.11	0.11	0.11
J2	0.11	0.07	0.09	0.03	0.70	0.20	0.15	0.12	0.00	0.05	0.03	0.19	0.01	0.73	0.05	0.03	0.05	0.04	0.85	0.07	0.21	0.88	0.88	0.13
J3	0.19	0.32	0.59	0.12	0.10	0.06	0.11	0.10	0.00	0.10	0.14	0.10	0.00	0.02	0.02	0.03	0.81	0.04	0.95	0.93	0.50	0.13	0.13	0.13
J4	0.16	0.06	0.12	0.08	0.12	0.51	0.15	0.40	0.00	0.06	0.06	0.19	0.00	0.02	0.05	0.92	0.03	0.04	0.95	0.07	0.92	0.13	0.25	0.13
J5	0.22	0.51	0.23	0.13	0.10	0.19	0.17	0.35	0.00	0.30	0.10	0.05	0.00	0.04	0.03	0.03	0.97	0.95	0.88	0.08	0.08	0.86	0.14	0.86
J6	0.44	0.69	0.27	0.30	0.23	0.09	0.25	0.23	0.00	0.40	0.22	0.13	0.00	0.02	0.03	0.03	0.91	0.05	0.94	0.08	0.08	0.86	0.14	0.86
J7	0.11	0.12	0.05	0.07	0.06	0.02	0.06	0.06	0.00	0.83	0.05	0.03	0.05	0.03	0.03	0.04	0.31	0.06	0.93	0.22	0.22	0.67	0.17	0.67
J8	0.11	0.35	0.40	0.06	0.05	0.18	0.11	0.28	0.00	0.16	0.09	0.16	0.00	0.07	0.04	0.05	0.95	0.93	0.82	0.88	0.13	0.20	0.20	0.20
J9	0.21	0.02	0.22	0.05	0.18	0.28	0.36	0.23	0.03	0.02	0.12	0.32	0.07	0.03	0.30	0.05	0.05	0.07	0.91	0.13	0.63	0.20	0.60	0.20
J10	0.07	0.03	0.19	0.04	0.11	0.36	0.14	0.11	0.00	0.03	0.03	0.54	0.00	0.10	0.83	0.50	0.05	0.07	0.82	0.13	0.13	0.20	0.80	0.20
J11	0.07	0.04	0.05	0.08	0.16	0.32	0.54	0.11	0.00	0.04	0.02	0.22	0.08	0.14	0.04	0.50	0.05	0.07	0.82	0.13	0.13	0.80	0.60	0.20
J12	0.04	0.01	0.05	0.01	0.18	0.06	0.12	0.05	0.36	0.01	0.40	0.05	0.08	0.20	0.06	0.07	0.07	0.10	0.75	0.17	0.17	0.25	0.25	0.25
J13	0.20	0.13	0.09	0.05	0.04	0.06	0.05	0.68	0.00	0.04	0.06	0.13	0.03	0.09	0.22	0.13	0.88	0.17	0.80	0.25	0.50	0.33	0.33	0.50
Df1	0.04	0.01	0.01	0.05	0.12	0.10	0.08	0.07	0.08	0.08	0.09	0.08												
Df2	0.19	0.02	0.67	0.03	0.05	0.41	0.14	0.17	0.08	0.17	0.08	0.17												
Df3	0.04	0.59	0.05	0.06	0.08	0.08	0.17	0.29	0.14	0.29	0.29	0.14												
Df4	0.75	0.07	0.93	0.09	0.75	0.88	0.88	0.80	0.20	0.20	0.20	0.80												
Df5	0.06	0.07	0.07	0.09	0.63	0.88	0.88	0.40	0.20	0.20	0.20	0.80												
Df6	0.58	0.89	0.70	0.86	0.80	0.17	0.83	0.75	0.25	0.75	0.75	0.75												
Df7	0.58	0.90	0.90	0.13	0.17	0.17	0.17	0.25	0.25	0.25	0.25	0.33												
Df8	0.71	0.83	0.17	0.80	0.25	0.25	0.25	0.33	0.33	0.67	0.33	0.33												
Df9	0.86	0.83	0.17	0.20	0.75	0.50	0.75	0.67	0.33	0.67	0.33	0.33												
Df10	0.83	0.83	0.33	0.80	0.75	0.75	0.50	0.67	0.33	0.67	0.67	0.67												

Intersection between Japanese legal concepts (J1-J8) and Danish legal concepts (D1-D9)
If mean value is equal or over 0.4 — 0.43
If mean value is equal or over 0.3 — 0.34
If mean value is equal or over 0.2 — 0.25
If mean value is equal or over average of all mean values — 0.15

Japanese or Danish feature structures for each legal concept cluster
If likelihood is equal of over 0.5 — 0.50

Fig. 8: n-IRM clustering and feature structure: JP as background knowledge (BMG: mean and $\boldsymbol{\eta}$ values)

cation") and *D7* (consisting of *"medium cycle tertiary education"* and *"Bachelor's program"*). These clusters are obtained by the BMG considering the background knowledge of the Japanese educational system from a Japanese reasoner's viewpoint. When inspecting the feature clusters colored[3] in Fig. 7,

[3] $\boldsymbol{\eta}$ values equal to or over 0.5

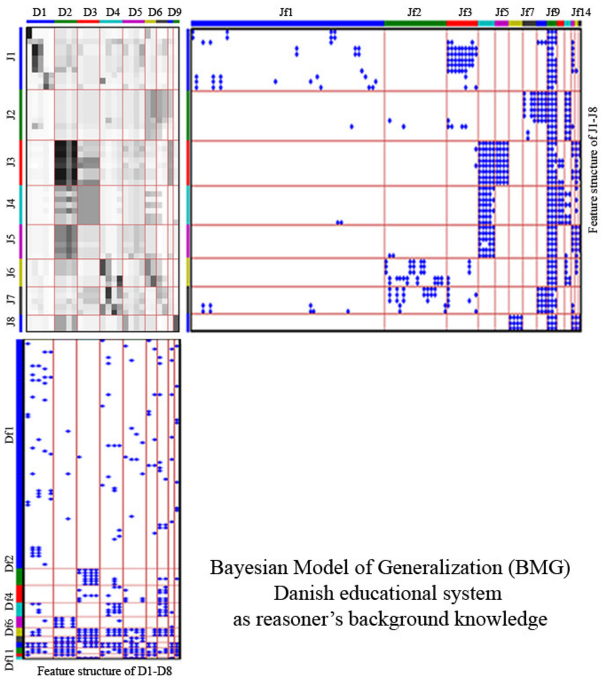

Fig. 9: n-IRM clustering and feature structure: DK as background knowledge

Avr. 0.12	D1	D2	D3	D4	D5	D6	D7	D8	D9	Jf1	Jf2	Jf3	Jf4	Jf5	Jf6	Jf7	Jf8	Jf9	Jf10	Jf11	Jf12	Jf13	Jf14
J1	0.19	0.11	0.06	0.08	0.09	0.07	0.06	0.05	0.02	0.06	0.01	0.46	0.02	0.02	0.02	0.02	0.12	0.89	0.04	0.04	0.54	0.15	0.08
J2	0.03	0.08	0.07	0.06	0.06	0.26	0.20	0.13	0.02	0.00	0.02	0.06	0.02	0.03	0.03	0.49	0.66	0.96	0.05	0.95	0.18	0.09	0.10
J3	0.03	0.69	0.25	0.06	0.13	0.07	0.04	0.32	0.05	0.00	0.01	0.04	0.98	0.97	0.03	0.03	0.04	0.88	0.39	0.06	0.60	0.90	0.60
J4	0.02	0.26	0.34	0.04	0.09	0.14	0.05	0.03	0.04	0.01	0.01	0.02	0.86	0.03	0.03	0.03	0.04	0.95	0.94	0.56	0.11	0.38	0.11
J5	0.03	0.43	0.11	0.07	0.14	0.11	0.02	0.06	0.04	0.00	0.03	0.02	0.90	0.04	0.04	0.04	0.05	0.95	0.07	0.08	0.75	0.88	0.71
J6	0.03	0.15	0.06	0.25	0.07	0.34	0.11	0.04	0.02	0.00	0.32	0.13	0.04	0.05	0.05	0.05	0.06	0.94	0.08	0.67	0.14	0.57	0.14
J7	0.03	0.09	0.14	0.29	0.12	0.05	0.18	0.09	0.02	0.03	0.24	0.17	0.04	0.05	0.05	0.05	0.94	0.82	0.08	0.08	0.43	0.14	0.14
J8	0.03	0.26	0.06	0.07	0.15	0.09	0.01	0.05	0.55	0.01	0.02	0.03	0.06	0.07	0.93	0.07	0.09	0.91	0.13	0.13	0.80	0.80	0.80
Df1	0.10	0.01	0.01	0.05	0.03	0.06	0.07	0.07	0.10														
Df2	0.03	0.04	0.81	0.15	0.04	0.07	0.21	0.13	0.25														
Df3	0.03	0.04	0.27	0.15	0.23	0.07	0.71	0.13	0.13														
Df4	0.19	0.05	0.05	0.59	0.05	0.08	0.83	0.29	0.14														
Df5	0.09	0.72	0.06	0.11	0.06	0.10	0.10	0.33	0.17														
Df6	0.38	0.07	0.93	0.69	0.71	0.50	0.88	0.75	0.20														
Df7	0.08	0.90	0.90	0.10	0.70	0.17	0.17	0.25	0.25														
Df8	0.73	0.90	0.70	0.90	0.44	0.80	0.17	0.75	0.67														
Df9	0.58	0.90	0.20	0.56	0.90	0.67	0.67	0.75	0.75														
Df10	0.29	0.83	0.17	0.33	0.67	0.75	0.25	0.33	0.67														
Df11	0.14	0.17	0.83	0.17	0.83	0.25	0.25	0.67	0.33														

Intersection between Japanese legal concepts (J1-J8) and Danish legal concepts (D1-D9)

If mean value is equal or over 0.4	0.43
If mean value is equal or over 0.3	0.34
If mean value is equal or over 0.2	0.25
If mean value is equal or over average of all mean values	0.13

Japanese or Danish feature structures for each legal concept cluster

If likelihood is equal of over 0.5	0.50

Fig. 10: n-IRM clustering and feature structure: DK as background knowledge (BMG: mean and $\boldsymbol{\eta}$ values)

the feature clusters that differentiate between the Japanese concept clusters *J10* and *J11* are *Jf3* possessed by *J10*, and *Jf10* possessed by *J11*. The feature cluster *Jf3* consists of features among others *"programme destination B"* and *"ISCED 5 - short"*, and the feature cluster *Jf10* consists of a feature *"programme destination A"* as listed in Fig. 21 in the Appendix I. Interestingly, the Danish concept clusters *D12* and *D7* that respectively have stronger relations with these Japanese concept clusters *J10* and *J11* are differentiated by whether the feature cluster *Df9* consisting of a feature *'programme destination B'* is possessed or not.

On the other hand, in Fig. 9, the Japanese concept cluster *J7* to which *"junior college - regular course"*, *"junior college - correspondence course"*, *"college of technology - regular course"* belong also includes *"university undergraduate"* and *"university undergraduate - correspondence course"* (see also Fig. 22 in the Appendix I). This Japanese *J7* concept cluster is, in Fig. 9, cross-categorized with the Danish concept cluster *D4* considering the background knowledge of the Danish educational system from a Danish reasoner's viewpoint. One notable point is that, at the upper-left table in Fig. 10, the intersection connecting *J7* and *D4* has slightly high mean value, 0.29. On the other hand, the n-IRM plot (the upper-left plot) in Fig. 9 indicates that the intersection of *J7* and *D4* is uneven, i.e., the standard deviation is high as identified in the plot 4-c in Fig. 3. This implies that these two concept clusters are rather linked based on individual concept-concept relations.

	Elementary school	School for Special Needs Education, elementary department	University, undergraduate of pharmacy (practical course)	University, graduate school, Master's correspondence course	University, professional graduate school, correspondence	University, graduate school, Master's course	University, professional graduate school	University, professional graduate law school	University, graduate school, Doctor's course	University, graduate school, Doctor''s course of pharmacy (practical course)	University, graduate school, Doctor's correspondence course
Primary level 1st-6th grade	0.77	0.77	0.24	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.02
Master	0.02	0.02	0.20	0.76	0.76	0.62	0.62	0.57	0.03	0.03	0.02
Undivided Master degrees	0.02	0.02	0.27	0.46	0.46	0.31	0.31	0.31	0.09	0.03	0.11
Doctoral programmes	0.02	0.02	0.02	0.01	0.01	0.02	0.02	0.05	0.42	0.42	0.57
Doctorate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.13	0.26

Fig. 11: Similarity relations between members of J1 and D1 clusters: BMG - DK as background knowledge (BMG)

Elaborating the issue of higher standard deviations, some of the intersections between the Japanese- and the Danish concept clusters indicate lower mean values with higher standard deviations as shown in, e.g., the intersection between *J1* and *D1* in the n-IRM (the upper-left) plot in Fig. 9. More specifically, the mean value of the intersection between *J1* and *D1* is 0.19 in the upper-left table of Fig. 10 and the standard deviation of this intersection is very high in the plot 4-c of Fig. 3. It means that this intersection consists of similarity scores that are substantially uneven influenced by the combinations of members in *J1* and *D1*. For instance, Fig. 11 shows the similarity relations between the members of the concept clusters *J1* (column) and *D1* (row), each computed by the BMG when Danish legal concepts are considered as reasoner's background knowledge. The members of both *J1* and *D1* consist of different types of concept such as *"primary school"*, *"Master's degree"* and *"Doctor's degree"* concepts. This indicates that, by analyzing the individual similarity relations between members of *J1* and *D1*, our approach also identifies more fine-grained one-to-one (or sub-group) relations within the two concept clusters.

11 Visualization

As demonstrated in the previous section, the cross-categorization approach identifies an interactive structure between the Japanese and the Danish educational systems according to the different types of similarity measures, from there uncovers underlying feature structures of the respective educational systems. For the purpose of visual inspection, we developed a hierarchical graph (ontology) based on the principles of the Terminological Ontology (TO) (Madsen et al. 2004)[4]. Fig. 12 represents the Japanese educational system developed based on the Japanese reasoner's viewpoint where the $\boldsymbol{\eta}$ values (links between concept clusters [J1-J13] and feature clusters [Jf1-Jf12]) equal to or above 0.5 are extracted from the upper-right table in Fig. 8 and considered as true values.

The ontology contains all the Japanese concept clusters (categories) obtained from the n-IRM computation, which are displayed in the colored boxes. More specifically, the concept clusters $[J1, J2, ...J13]$ are displayed in the colored boxes and their corresponding Danish concept clusters among $[D1, ...D12]$ are listed above the J_i labels in the colored boxes. These corresponding Danish concept clusters are identified based on the strength of the links between a Japanese and a Danish concept clusters, i.e., when the mean values (m_{ab}) are equal to or over 0.3 in the upper-left table of Fig. 8, they are considered as corresponding pairs. For example, it is possible to view in Fig. 12 that the Japanese concept cluster $J10$ consisting of members referring *junior college* and *college of technology* strongly corresponds to the Danish concept cluster $D12$ referring to *short cycle tertiary education* and $D6$ referring to *short cycle open tertiary education* in Denmark.

The ontology is structured based on the feature inheritance rules. It means that each category (colored box) in the ontology is represented in the form of feature structure listed under each colored box. For example, the concept cluster *"J10: junior college, college of technology"* is represented in the form of feature structure *(Jf7, Jf11, Jf4, Jf3)*. This means that the concept cluster $J10$ consists of the feature clusters $Jf7, Jf11, Jf4, Jf3$. Among these feature clusters, $Jf7, Jf11, Jf4$ are inherited from the superordinate concept clusters: *"category share Jf11"* and *"category share Jf4"*. $Jf3$ is only the non-inherited feature cluster. Accordingly, the contents of $Jf3$ [destination B, ISCED5 short] is specified just below the gray box of $J10$.

[4] In this work, we consider that the ontology construction is out of the main focus . In other words, the ontology has been developed solely for the purpose of visualization. The theoretical background of the TO method is therefore explained in Appendix II.

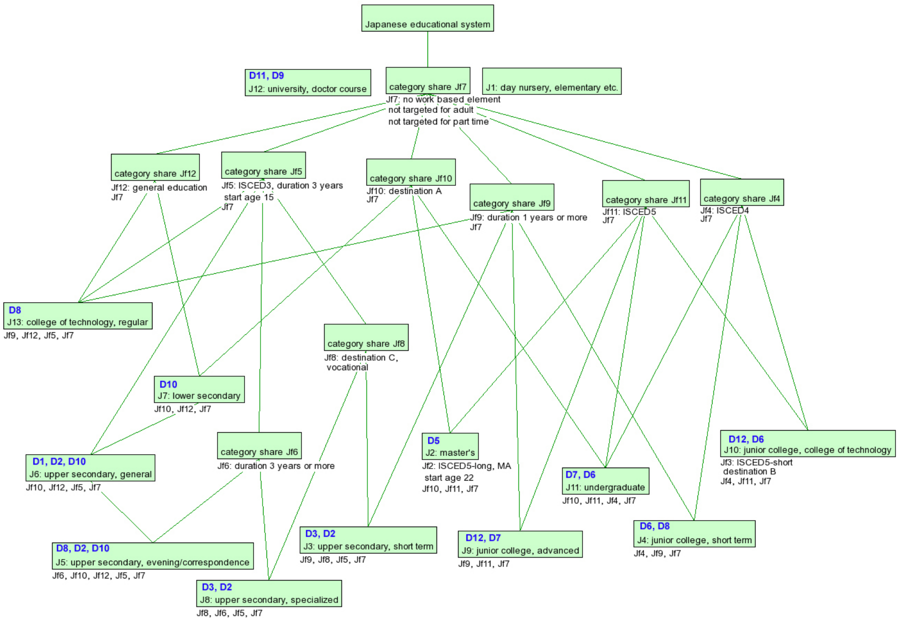

Fig. 12: Japanese educational system ontology: BMG - JP as background knowledge (NOTE: J1-J13; D1-D12; and Jf1-Jf12 in this ontology refer to clusters shown in Fig11)

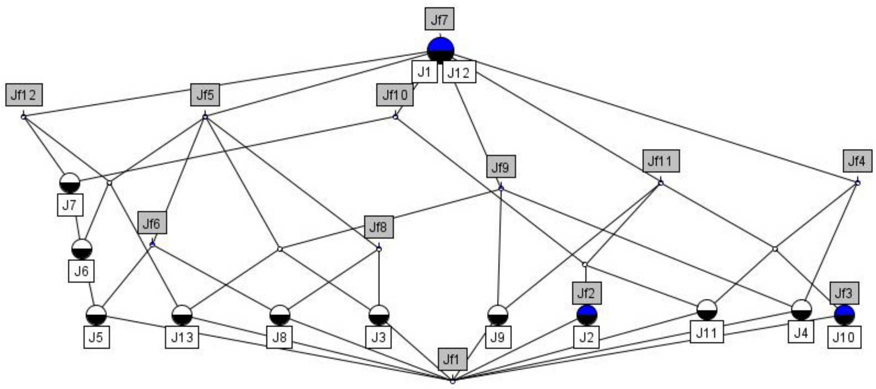

Fig. 13: Formal Concept Analysis of Japanese educational system: BMG - JP as background knowledge (NOTE: J1-J13 and Jf1-Jf12 in this graph refer to clusters shown in Fig11)

The principle for creating nodes such as *"category share Jf11"* and *"category share Jf4"* are very similar to the theory of Formal Concept Analysis (FCA) (Ganter and Wille 1997)[5]. Fig. 13 is automatically drawn by a FCA

[5] The theory of FCA is reviewed in Appendix II

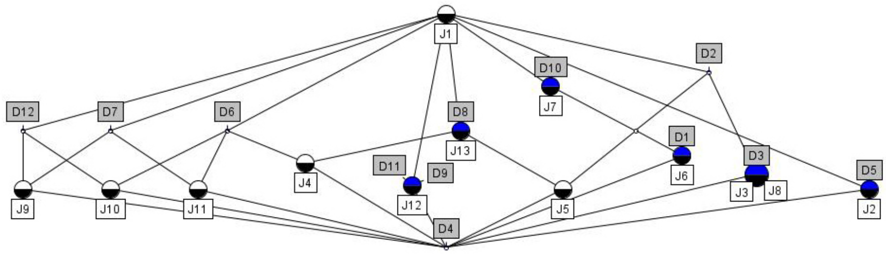

Fig. 14: Concept mapping made by the FCA: BMG - JP as background knowledge (NOTE: J1-J13 and D1-D12 in this graph refer to clusters shown in Fig11)

tool called Concept Explorer (ConExp) (Yevtushenko 2000) [6] from the same data source used for constructing the ontology in Fig. 12. More specifically, Fig. 13 is created from the binary object-feature matrix consisting of the concept clusters [J1-J13] and the feature clusters [Jf1-Jf12] where the $\boldsymbol{\eta}$ values equal to or above 0.5 are considered as true values. The white labels and the gray labels in Fig. 13 are the concept and the feature clusters, respectively. A big round nodes consisting of two colors are the so-called *formal concept*. The blue and black color indicates that both concept and feature clusters are directly attached to a formal concept in question. When a node is white and black color, it means that only concept clusters are attached to a formal concept in question. Each formal concept inherits gray-labeled features from ascendent edges connected to a formal concept in question. For example, the concept node $J10$ possesses a feature cluster $Jf3$ directly attached to the node $J10$, feature clusters $Jf11andJf4$ that are connected via the intersection between $Jf11$ and $Jf4$, and feature cluster $Jf7$ that is the top node where all concepts are eventually connected in Fig. 13. Notable point is that the edges where only feature clusters, e.g. $Jf11$ and $Jf4$, are attached are corresponding to the nodes (e.g. *"category share Jf11"* created in Fig. 12). It is observable that the structure of the ontology in Fig. 12 and the FCA graph in Fig. 13 are substantially similar to each other.

Finally, this FCA method can be used for visualizing how the Danish concept clusters are corresponding to the Japanese concept clusters based on the information extracted from the n-IRM computation. For example, Fig. 14 is drawn from the binary matrix consisting of the Japanese concept clusters [J1-J13] and the Danish concept clusters [D1-D12] where the mean values equal to or over 0.3 in the upper-left table of Fig. 8 are considered as true values. In Fig. 14, a node (either blue/black or white/black) indicates that a Japanese concept cluster is directly or indirectly connected to one or more Danish concept cluster(s). This indicates that the Japanese concept cluster $J10$ corresponds to the two Danish concept clusters $D12$ and $D6$ in Fig. 14. In

[6] http://conexp.sourceforge.net/

this way, the ontology in Fig. 12 overviews the relations automatically drawn by the FCA method in a more specific manner.

12 Discussion

To summarize, what we have demonstrated in this work are:

- the cross-categorization approach co-clusters the legal concepts belonging to the two different legal systems;
- the cluster distributions resulted from the n-IRM are highly dependent on different types of similarity scores, i.e., conditional probabilities computed by the generalization model, BMG;
- by closer looking into an intersection of clusters having a higher standard deviation obtained from the n-IRM, it is possible to identify clusters that include concept-concept relations within the intersection in question;
- the cross-categorization approach analyzes feature structures of each obtained concept cluster by combining the IRM applied to the original concept-feature matrices;
- by defining a threshold on the $\boldsymbol{\eta}$ values obtained from the IRM, the extract feature structures can be visualized as hierarchical graphs; and
- by defining a threshold on the mean values (m_{ab}) obtained from the n-IRM, it is possible to map concept clusters between the two different legal systems;

In terms of the issues of the knowledge alignment, it has been indicated that the application of the n-IRM enables us not only to map legal concepts between the two legal systems, but also to structure two independent concept systems based on the analysis of the interactivity between them. For example, by observing interactive relations extracted based on the mean values, e.g., in Fig. 8, it is possible to align knowledge into several layers, i.e. individual concept level, cluster (category) level, and more abstract category levels. The knowledge structuring process can further be supported by the IRM analysis of the original binary concept-feature matrices that have been used for computing similarity scores. Such information additionally supports the structuring of individual knowledge systems, since feature clusters and concept clusters also inter-relate to each other as shown in Fig. 6, Fig. 8 and Fig. 10.

One of the technical limitations in this work is the seamlessness of the entire process due to the combination of several independent algorithms. The combination of the n-IRM and the IRM computations could, in the future, be seamlessly integrated as a joint model based on e.g. Infinite Hidden Relational Model (Xu et al. 2006). This could optimize the involved cluster partitions by reflecting the two types of input data, i.e. similarity scores and binary data. Another technical limitation is that the alignment of knowledge structures depends, to a certain extent, on the subjective judgments. To be more specific, the threshold on $\boldsymbol{\eta}$ values obtained from the IRM has arbitrary been defined in order to approximate the structures of the respective domain knowledge identified by the cross-categorization approach. In the same way, the threshold on mean values computed by the n-IRM has also been defined subjectively, in order to determine which Danish concept clusters are corresponding to each Japanese concept cluster in Fig. 12 and Fig. 14. Here, an optimal thresholds could be estimated based on a statistical method in the future. When the thresholds are identified in a statistical manner, a method such as FCA could efficiently integrated to automatically draw an hierarchical graph of knowledge while aligning the two knowledge structures. Such an integrated method could be useful for not only aligning the two different knowledge systems as shown in this work, but also for constructing an ontology representing a single knowledge system, when the cross-categorization approach is applied to two identical knowledge systems (concept-feature matrices) representing a single legal system.

From the viewpoint of the human reasoning, we have argued that the cross-categorization approach attempts to explain reasoner's inferential mechanisms. As Sartor (2009) employs Ross's theory defining *Q-conditioning links* and *Q-conditioned links* for an intermediate legal concept, each educational concept consists of features, i.e. *Q-conditioning links* such as *"minimum entrance qualification"* defined by age or by *"previous educational degree"*, and *Q-conditioned links* such as *"degree obtained upon the completion of the education"*. Sartor (2009) claims that *intermediate legal concepts are used within normative systems that connect observable facts and normative qualifications of human actions* (Sartor 2009). From this point of view, our approach applying the IRM have grouped the co-occurring features (the members of feature clusters listed in Appendix I: Fig. 19, Fig. 21 and Fig. 23) as feature clusters. In other words, the approach has identified underlying relationships among concepts and categories by analyzing co-occurrences of preconditions and consequences identified among specific instances. More specifically, the generalization model, i.e. the BMG, has theoretically been considered as the *analogical inference* described in (Sartor 2009) after (Thagard 1992). The mechanism of our approach attempted to explain the following situation: i.e., if a reasoner only knows about legal concepts used within a legal system which he or she belongs to, a concept belongs to the other legal system is considered as a new concept which a reasoner has to learn. The reasoner then learns the new concept by identifying features in the new concept which co-occur with features possessed by the already known concepts, concept clusters, or abstract categories - his or her background knowledge. The co-occurrence of features within the respective feature clusters have further been influenced by the BMG in a way that a feature possessed by fewer concepts within reasoner's background knowledge has been considered more important.

In order to argue reasoner's reasoning mechanisms, we need to design and implement thorough experiments involving human evaluators. Our standpoint for the human assessment of this work is therefore that it should be considered as out of scope in this paper. However, we attempt, as a preliminary experiment, to test our hypothesis that the tendencies demonstrated with three types of similarity scores in this work may possibly reflect a way evaluators align the educational concepts based on his/her background knowledge.

Accordingly, two versions of questionnaires are respectively sent out to two Japanese and three Danish evaluators. In this preliminary experiment, no specific criteria have been specified for the selection of the evaluators. We have randomly asked several Danish and Japanese individuals. This implies that the evaluators are not specialized in the evaluation of the foreign educational qualifications. The questionnaires are designed in a way that the Japanese evaluators are asked to first read through the definitions of the 54 Japanese educational concepts (background knowledge). Each of the 27 Danish educational terms expressed in Danish (a new concept introduced from the other system) and its definitions (features) are displayed on one page and the evaluators are asked to select as many as Japanese concepts based on the two criteria defined below:

– to select IDEAL Japanese concept(s) - IDEAL means here option(s) in which the Danish concept in question is suitably corresponding to. Note: there is an option not to select any of them in this criterium.
– to select ACCEPTABLE Japanese concept(s) - ACCEPTABLE means here option(s) in which the Danish concept can be categorized as a category member together with the option(s). Note: At least one or more option(s) should be selected in this criterium.

For the Danish evaluators, the questionnaire is made in a way that they are asked to read through the definitions of the 27 Danish educational terms (background knowledge) and to select one or more Danish concept(s) that correspond to each of the 54 Japanese educational concepts (a new concept). When the evaluators select an IDEAL concept (background knowledge) corresponding to a concept newly encountered, a value *"1"* is assigned. If an ACCEPTABLE concept is selected, *"0.3"* is assigned, otherwise *"0"* is assigned. The overview of the human evaluations are illustrated in Figs. 15 and 16.

Fig. 15 contrasts the five human assessments with the similarity scores computed for all the combinations between the Japanese and the Danish concepts by the three measures: Jaccard similarity coefficient (Jaccard), the BMG when Japanese concepts are considered as reasoner's background knowledge (BMG JP) and the BMG when Danish concepts are considered as reasoner's background knowledge (BMG DK). One interesting finding is that the Japanese evaluator's results are rather similar to the plot (BMG JP) and the Danish evaluator's results are the plot (BMG DK) in Fig. 15. The explanation of this phenomenon is that the Japanese evaluators were asked to select IDEAL (black colors in the plot) and ACCEPTABLE (gray colors) Japanese concepts when each of the 27 Danish concept was introduced as a new concept. Typically, the Japanese evaluators selected zero to one Japanese concept for the IDEAL options, and one or more Japanese concepts for the ACCEPTABLE options. Therefore, the horizontal lines become clearer for the Japanese evaluations. On the other hand, the Danish evaluators were asked to select IDEAL and ACCEPTABLE Danish concepts when each Japanese of the 54 Japanese concept was introduced as a new concept. Fig. 15 shows that Danish evaluators

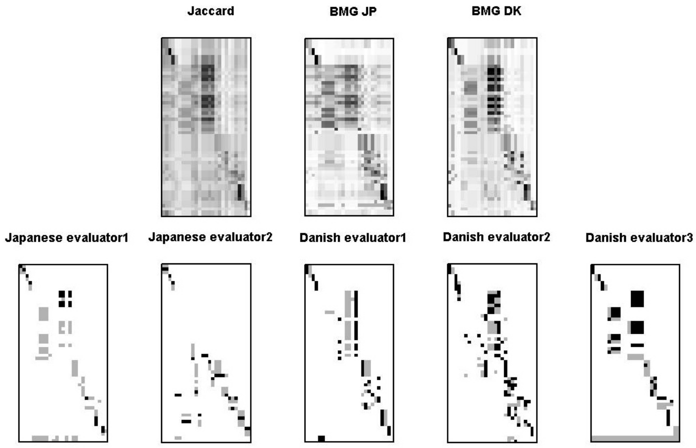

Fig. 15: Human assessments contrasted to the similarity scores

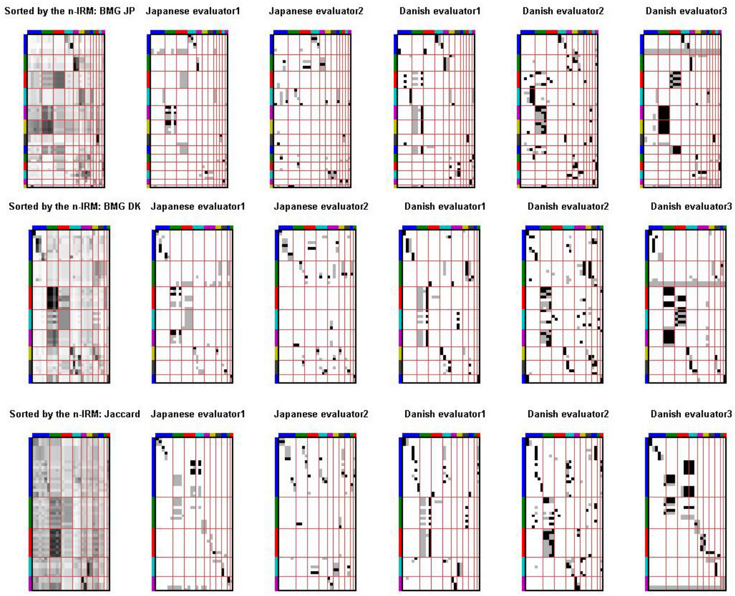

Fig. 16: Human assessments contrasted to the cross-categorization results

generally selected more options than Japanese has done. Since the selection criteria between IDEAL and ACCEPTABLE highly depend on individual Danish evaluators, the plots in Fig. 15 are not identical. However, when observing the Danish evaluator 1 who has, like the Japanese evaluator 1, selected zero to one Danish concept for the IDEAL options and one ore more Danish concepts for the ACCEPTABLE options, the vertical lines become clearer for the Danish evaluations. In Fig. 16, the human evaluation data has been re-ordered according to the cluster assignments obtained in this work (i.e. the results obtained in Fig. 3). In this way, the evaluations made by the fine evaluators can be contrasted to the results obatined by the n-IRM.

The collected data is further contrasted to the cross-categorization results based on the two criteria defined as: i) strict mapped pairs, and ii) relaxed mapped pairs. The strict mapped pairs refer to the number of Japanese- and Danish concept pairs that fall under an intersection of the Japanese and the Danish concept clusters with the mean value equal to or above 0.30 computed by the n-IRM computation, while the relaxed mapped pairs refer to the number of Japanese- and Danish concept pairs that fall under an intersection with the mean value equal to or above the average mean values identified in the upper-left plots in Figs. 6, 8 and 10. Based on these definitions, precision, recall and F-measures scores are computed based on the equations defined as:

$$precision = correspond.\ pairs \bigcap mapped\ pairs\ /\ mapped\ pairs \tag{11}$$

$$recall = correspond.\ pairs \bigcap mapped\ pairs\ /\ correspond.\ pairs \tag{12}$$

$$F - measure = (2 * precision * recall)\ /\ (precision + recall) \tag{13}$$

Here, $correspond.\ pairs \bigcap mapped\ pairs$ are the number of corresponding pairs that fall under the mapped intersection of the Japanese and the Danish concept clusters defined in the above. Fig. 17 below summarize the obtained results .

The results shown in Fig. 17 indicate that one of the major limitations of this preliminary experiment is the way the mapped pairs (strict- and relaxed mapped pairs) have been defined. As already pointed out as a technical limitation, the threshold for defining the *mapped relationships* should in the future statistically be estimated, because the distributions of scores obtained by the three different similarity measures are substantially different. This may severely affect the judgment of the performances achieved by each similarity measures shown in Fig. 17. Another limitation is that the precision, recall and F-measures cannot consider the differences in granularities of the concept clusters obtained by the three similarity measures. To be concrete, the categorization result obtained by the BMG (BMG JP) is fine-grained so that it is more difficult to perform good mapping results compared to the more abstract categorization results achieved by the Jaccrd coefficient or the BMG (BMG DK). In addition, the possibility to construct a hierarchical graph (ontology) from the more fine-grained clusters should be taken into consideration when assessing the mapping performances. Another limitation is how to control evaluators' background knowledge of the respective educational systems. In this preliminary experiment, the evaluators are exposed to each of the newly encountered concepts and its definition (list of features) introduced from the other educational system. It means that the more a new concept is introduced,

	strict mapped pairs	relaxed mapped pairs (incl. Strict)		Corresponding pairs	
				IDEAL	ACCEPTABLE
BMG_JP	221	664	JP evaluator1	22	121
BMG_DK	105	449	JP evaluator2	27	89
Jaccard	148	634	DK evaluator1	56	121
			DK evaluator2	88	151
			DK evaluator3	85	201

Japanese evaluator1	IDEAL						ACCEPTABLE (incl. IDEAL)					
	strict mapped pairs			relaxed mapped pairs (incl. Strict)			strict mapped pairs			relaxed mapped pairs (incl. Strict)		
BMG_JP	18			21			80			96		
BMG_DK	47			21			55			102		
Jaccard	17			21			38			101		
	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.
BMG_JP	0,08	0,82	0,146	0,03	0,95	0,058	0,36	0,66	0,466	0,14	0,79	0,238
BMG_DK	0,13	0,64	**0,216**	0,05	0,95	**0,095**	0,52	0,45	**0,482**	0,23	0,84	**0,361**
Jaccard	0,11	0,77	0,193	0,03	0,95	0,058	0,26	0,31	0,283	0,16	0,83	0,268

Japanese evaluator2	IDEAL						ACCEPTABLE (incl. IDEAL)					
	strict mapped pairs			relaxed mapped pairs (incl. Strict)			strict mapped pairs			relaxed mapped pairs (incl. Strict)		
BMG_JP	10			17			37			64		
BMG_DK	4			16			8			55		
Jaccard	6			18			15			56		
	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.
BMG_JP	0,05	0,37	**0,088**	0,03	0,63	0,057	0,17	0,42	**0,242**	0,1	0,72	0,176
BMG_DK	0,04	0,15	0,063	0,04	0,59	**0,075**	0,08	0,09	0,085	0,12	0,62	**0,201**
Jaccard	0,04	0,22	0,068	0,03	0,67	0,057	0,1	0,17	0,126	0,09	0,63	0,158

Danish evaluator1	IDEAL						ACCEPTABLE					
	strict mapped pairs			relaxed mapped pairs (incl. Strict)			strict mapped pairs			relaxed mapped pairs (incl. Strict)		
BMG_JP	34			48			83			109		
BMG_DK	17			47			49			104		
Jaccard	27			40			48			104		
	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.
BMG_JP	0,15	0,61	0,241	0,07	0,86	0,129	0,38	0,69	**0,490**	0,16	0,9	0,272
BMG_DK	0,16	0,3	0,209	0,1	0,84	0,179	0,47	0,4	0,432	0,23	0,86	**0,363**
Jaccard	0,18	0,48	**0,262**	0,45	0,55	**0,495**	0,08	0,86	0,146	0,16	0,86	0,270

Danish evaluator2	IDEAL						ACCEPTABLE					
	strict mapped pairs			relaxed mapped pairs (incl. Strict)			strict mapped pairs			relaxed mapped pairs (incl. Strict)		
BMG_JP	48			80			85			133		
BMG_DK	29			66			58			109		
Jaccard	40			64			72			110		
	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.
BMG_JP	0,22	0,55	0,314	0,12	0,91	0,212	0,38	0,56	0,453	0,2	0,88	0,326
BMG_DK	0,28	0,33	0,303	0,15	0,75	**0,250**	0,55	0,38	0,449	0,24	0,72	**0,360**
Jaccard	0,27	0,45	**0,338**	0,1	0,73	0,176	0,49	0,48	**0,485**	0,17	0,73	0,276

Danish evaluator3	IDEAL						ACCEPTABLE					
	strict mapped pairs			relaxed mapped pairs (incl. Strict)			strict mapped pairs			relaxed mapped pairs (incl. Strict)		
BMG_JP	77			80			111			134		
BMG_DK	55			80			71			144		
Jaccard	46			72			56			136		
	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.	precision	recall	F-meas.
BMG_JP	0,35	0,91	0,506	0,12	0,94	0,213	0,5	0,55	**0,524**	0,2	0,67	0,308
BMG_DK	0,52	0,65	**0,578**	0,18	0,94	**0,302**	0,68	0,35	0,462	0,32	0,72	**0,443**
Jaccard	0,31	0,54	0,394	0,11	0,85	0,195	0,38	0,28	0,322	0,21	0,68	0,321

Fig. 17: Precision, recall and F-measure scores of the human assessment results

the more the evaluators will be exposed to their features existing in the other educational system. The critical issue is how to investigate the influence of reasoner's background knowledge and its impact on the categorization. It is not enough to claim that the way the human evaluators have been asked to select IDEAL and ACCEPTABLE corresponding concepts reflects the way they categorize a newly learned alien concept based on their background knowledge. Here, we need to further explore an suitable method for implementing a thorough human assessment in the future.

An interesting open question is whether reasoner's ontology will be updated and therefore should be reorganized, or unchanged, when a new concept is introduced from an alien system. Our cross-categorization framework could contribute to the alignment of domain knowledge existing in two legal systems. However, it is too optimistic to claim that the ontologies constructed in this work can represent reasoner's categorical knowledge structure, i.e. ontology. Thus, it is wise to claim that the ontologies constructed in this work is merely limited for the purpose of visual inspection. It is obvious that thorough experimental designs and much deeper investigations are required for explaining these open issues. One thing we can claim here is that the conceptual framework presented in this work could be useful for conducting such thorough investigations in the future. It should particularly noted that the analogical inference algorithms employed in the framework would be one example among other possible generalization models. Whether the employed model, i.e. the BMG by Tenenbaum and Griffiths (2001) is optimal or not is still open question and requires further investigation in line with the human-based assessment of cross-categorization performances. For example, the generalization model (i.e. similarity measures) could be replaced by some advanced models such as (Goodman et al. 2008; Kemp et al. 2008), or alternatively by a new model reflecting the performances made by reasoners. Accordingly, the ambitious future perspective would be to design thorough experiments on investigating such knowledge alignment and structuring mechanism performed by expert reasoners belonging to different legal systems and contrast their performances with the cross-categorization approach employing several different generalization models.

Finally, this work has employed the data sets prepared by the UIS. Hence, the features representing the two different knowledge systems have been well standardized in advance. An issue to be considered in the future is how to obtain standardize features existing in different systems, when the proposed framework is applied to other legal domains such as taxation systems, pension systems etc. but also to more general domains (e.g. e-commerce, linked-data and semantic web) that require the alignment and integration between heterogeneous knowledge systems.

13 Conclusion

In this work, we proposed a cross-categorization approach that combines three mathematical models: the Bayesian Model of Generalization (BMG) by Tenenbaum and Griffiths (2001), the probabilistic model of theory formation, i.e. the Infinite Relational Model (IRM) first introduced by Kemp et al. (2006, 2010) and its extended model, i.e. the normal-Infinite Relational Model (n-IRM) proposed by Herlau et al. (2012). Our proposed approach that employs the principles of probabilistic model of theory formation (Kemp et al. 2010), has been contrasted to Giovanni Sartor's view of inferential semantics of legal concepts (Sartor 2009). The results indicate that our proposed approach, which optimally utilizes all available information contained in a dataset, demonstrates an advantage for directly reflecting similarity scores into the cross-categorization analyses between two knowledge systems. We also demonstrated that our proposed approach simultaneously categorizes concepts existing in two legal systems and from there structures two independent knowledge systems. From the view of legal reasonings, we argued that the generalization model, i.e. the BMG, employed as similarity measures in the aforementioned process could be considered as *analogical inference* described in Sartor (2009) after Thagard (1992). This implies that, although thorough investigations on knowledge alignment and structuring mechanism performed by expert reasoners are needed, the proposed framework would be useful for conducting such thorough investigations in the future, because the framework *theoretically* accommodate reasoner's analogical inferential mechanisms by combining the generalization model and the probabilistic model of theory formation.

References

D. Aldous. Exchangeability and related topics. In *École dÉté de Probabilités de Saint-Flour XIII-1983 (Lecture Notes in Mathematics)*, pages 1–198. Berlin, Germany: Springer, 1985.

J. Berlin and A. Motro. Database schema matching using machine learning with feature selection. In *Proc. 14th International Conference on Advanced Information Systems Engineering, Vol. 2348 of Lecture Notes in Computer Science*, pages 452–466. 2002.

A. Bilke and F. Neumann. Schema matching using duplicates. In *Proc. 21st International Conference on Data Engineering*, pages 69–80. 2005.

N. Block. Advertisement for a semantics for psychology. In *University of Minnesota Press*. Midwest studies in philosophy. Studies in the Philosophy of Mind, 1986.

P. Boghossian. Epistemic analyticity: A defense. *In: Glock, H-J, Glür, K., Keil, G. (eds) Grazer Philosophische Studien, Fifty Years of Quine's two dogmas, Rodopi, Amsterdam*, pages 15–35, 2003.

C. M. T. Cabré. Elements for a theory of terminology, towards an alternative paradigm. In *Terminology / 2000, Vol.6 , No.1*. Amsterdam, Netherlands: John Benjamins Publishing Company, 2000.

S. Carey. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press, 1985.

C. P. Cheng, G. T. Lau, K. H. Law, J. Pan, and A. Jones. Regulation retrivial using industry specific taxonomies. In *Artificial Intelligence and Law, vol. 16*, pages 277–303. 2008.

L. Danon, Díaz-Guilera A., J. Duch, and A. Arenas. Comparing community structure identification. Theory and Experiment, Journal of Statistical Mechanics, 2005.

E. Davis. *Representation of Commonsense Knowledge.* Morgan Kaufmann, 1990.

K. X. S. de Souza and J. Davis. Aligning ontologies and evaluating concept similarities. Lecture Notes in Computer Science. Springer, On the Move to Meaningful Internet Systems 2004, 2004.

T. Declerck, H. U. Krieger, S. M. Thomas, P. Buitelaar, S. O'Riain, T. Wunner, G. Maguet, J. McCrae, D. Spohr, and E. Montiel-Ponsoda. Ontology-based multilingual access to financial reports for sharing business knowledge across europe. In *Internal Financial Control Assessment Applying Multilingual Ontology Framework, J. Rooz,and J. Ivanyos(Eds.).* HVG Press, Budapest, 2010.

A. H. Doan, J. Madhavan, P. Domingos, and A. Halevy. Ontology matching: A machine learning approach. In S. Staab and R. Studer, editors, *Handbook on Ontologies, Chapter 18*, pages 385–404. Springer Verlag, Berlin, 2004.

P. Durst-Andersen. *Linguistic Supertypes: A Cognitive-semiotic Theory of Human Communication.* Berlin-New York: De Gruyter Mouton, 2011.

M. Ehrig. *Ontology Alignment: Bridging the Semantic Gap.* Springer, New-York, US, 2007.

J. Euzenat. Brief overview of t-tree: the tropes taxonomy building tool. In *Proc. 4th ASIS SIG/CR Workshop on Classification Research*, pages 69–87, 1994.

J. Euzenat and P. Shvaiko. *Ontology Matching.* Springer-Verlag, Berlin Heidelberg, 2007.

J. Euzenat and P. Valtchev. Similarity-based ontology alignment in owl-lite. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI),Valencia, Spain*, pages 333–337, 2004.

H. Field. Logic, meaning and conceptual role. *the Journal of Philosophy, vol. 69*, pages 379–408, 1977.

J. Fodor and E. Lepore. *Holism: A shopper's guide.* Cambridge, MA: Blackwell, 1992.

B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1997. ISBN 3540627715.

N. D. Goodman, J. B. Tenenbaum, J. Feldman, and T. L. Griffiths. A rational analysis of rule-based concept learning. In *Cognitive Science.*, pages 108–154. 1, 32, 2008.

T. Gruber. Toward principles for the design of ontologies used for knowledge sharing. In *International Journal Human-Computer Studies, vol. 43*, pages 907–928. 1992.

S. Haack. *Defending Science Within Reason.* Prometheus, Amherst, NY, 2003.

T. J. Hansen, M. Mørup, and L. K. Hansen. Non-parametric co-clustering of large scale sparse bipartite networks on the gpu. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), IEEE.* 2011.

T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2001.

C. G. Hempel. Thoughts on the limitations of discovery by computer. In K. Schaffner, editor, *Logic of discovery and diagnosis in medicine*, pages 115–122. University of California Press, Berkeley, 1985.

T. Herlau, M. Mørup, M. N. Schmidt, and L. K. Hansen. Modelling dense relational data. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Santander, Spain.* 2012.

H. H. Huang and Y. H. Kuo. Cross-lingual document representation and semantic similarity measure: A fuzzy set and rough set approach. In *IEEE Transaction on Fuzzy Systems. vol.18, 6.* 2010.

R. Ichise, H. Takeda, and S. Honiden. Integrating multiple internet directories by instance-based learning. In *Proc. 18th International Joint Conference on Artificial Intelligence*, pages 22–30, 2003.

R. Ichise, H. Takeda, and S. Honiden. Discovering relationships among catalogs. In *Proceedings of the 7th International Conference on Discovery Science, in Lecture Booktitles in Computer Science, vol.3245, Padova, Italy*, pages 371–379, 2004.

P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques regions voisines. In *Bulletin de la societe vaudoise des sciences naturelles,vol.37*, pages 241–272. 1901.

K. Kageura. *Dynamics of Terminology.* 2002.

C. Kemp and J. B. Tenenbaum. Structured statistical models of inductive reasoning. In *Psychological Review. 116(1)*, pages 20–58. 2009.

C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *The Twenty-First National Conference on Artificial Intelligence*. 2006.

C. Kemp, N. D. Goodman, and J. B. Tenenbaum. Learning and using relational theories. In *Advances in Neural Information Processing Systems 20.* 2008.

C. Kemp, J. B. Tenenbaum, S. Niyogi, and T. L. Griffiths. A probablistic model of theory formation. In *Cognition, vol. 114*, pages 165–196. 2010.

C. Kemp, P. Shafto, and J. B. Tenenbaum. An integrated account of generalization across objects and features. In *Cognitive Psychology, vol.64*, pages 35–73. 2012.

M. Lacher and G. Grog. Facilitating the exchange of explicit knowledge through ontology mappings. In *Proc. International Florida Artificial Intelligence Research Society Conference*, pages 305–309, 2001.

L. Lindahl. Deduction and justification in the law: The role of regal terms and concepts. *Ration Juris*, pages 182–201, 2004.

B. N. Madsen, H. E. Thomsen, and C. Vikner. Principles of a system for terminological concept modelling. In *Proceedings of the 4th International Conference on Language Resources and Evaluation, ELRA*, pages 15–19, 2004.

C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. WonderWeb deliverable D18 ontology library (final). Technical report, IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web, 2003.

P. Mitra, N. Noy, and A. Jaiswal. Ontology mapping discovery with uncertainty. In *Proceedings of the 4th International Semantic Web Conference (ISWC), In Lecture Booktitles in Computer Science,vol.3729, Galway, Ireland*, pages 537–547, 2005.

M. Mørup, K. H. Madsen, A. M. Dogonowski, H. Siebner, and L. K. Hansen. Infinite relational modeling of functional connectivity in resting state fmri. In *Proceedings of Neural Information Processing Systems*, 2010.

G. L. Murphy. *The Big Book of Concepts.* Cambridge, Massachusetts: The MIT Press, 2004.

G. L. Murphy and D. L. Medin. The role of theories in conceptual coherence. pages 289–316. 1985.

C. S. Peirce. The categories. In *Haack S (ed) Pragmatism, old and new. Prometheus, Amherst*, pages 177–208. 2006.

C. S. Peirce. *Collected papers of Charles Sanders Peirce I-VIII.* Cambridge University Press, 2010.

J. Pitman. Combinatorial stochastic processes. In *Booktitles for Saint Flour Summer School*. 2002.

S. Psillos. Rudolf carnap's theoretical concepts in science. *Stud Hist Philos Sci*, pages 151–172, 2000.

M. R. Quilian. Semantic memory. MIT Press, Semantic Information Processing, Cambridge, MA, 1968.

F. P. Ramsey. *On Truth: Original Manuscript Materials, 1927-1929 from the Ramsey Collection At the University of Pittsburgh.* Kluwer, Dordrecht, 1991.

A. Ross. Tu-tu. *Scand Stud Law*, pages 139–153, 1957.

G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Addison-Wesley Longman Publishing Co., Inc, Boston, MA, USA, 1989.

G. Sartor. Legal concepts as inferential nodes and ontological categories. In *Artificial Intelligence and Law, vol. 17*, pages 217–251. 2009.

D. Sperber and D. Wilson. *Relevance: Communication and Cognition.* Oxford: Blackwell, 1986.

G. Stumme and A. Mädche. Fca-merge: Bottom-up merging of ontologies. In *Proceedings of 17th International Joint Conference on Artificial Intelligence (IJAI)*, pages 225–234, 2001.

J. B. Tenenbaum and T. L. Griffiths. Generalization, similarity, and bayesian inference. In *Behavioral and Brain Sciences*, number 4, pages 629–640. 2001.

P. Thagard. *Conceptual Revolutions.* Princeton Unviersity Press, Princeton, 1992.

A. Tversky. Features of similarity. *Psychological Review*, (4):327–352, 1977.

J. Wang, J. R. Wen, F. Lochovsky, and W. Y. Ma. Instance-based schema matching for web databases by domain-specific query probing. In *Proc. 30th International Conference on Very Large Data Bases*, pages 408–419, 2004.

H. M. Wellman and S. A. Gelman. Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, pages 227–375, 1992.

A. Woodfield. On the very idea of acquiring a concept. Philosophical perspectives on developmental psychology. Basil Blackwell, Oxford, 1987.

E. Wüster. *Das worten der weld, schaubildlich und terminologisch dargestellt.* 1959.

Z. Xu, V. Tresp, K. Yu, and H. P. Kriegel. Infinite hidden relational models. In *Proceedings of 22nd Conference on Uncertainty in Artificial Intelligence, Cambridge, MA*, 2006.

S. A. Yevtushenko. System of data analysis "concept explorer" (in russian). In *Proceedings of the 7th national conference on Artificial Intelligence KII-2000, Russia*, pages 127–134, 2000.

A Appendix: Members of the concept clusters and the feature clusters

J1	Day nursery
	Kindergarten
	School for Special Needs Education, kindergarten department
	Upper secondary school, (full day, day/evening), advanced course (general, integrated, specialized)
	Secondary education school (Upper division), (full day, day/evening), advanced course (general, integrated, specialized)
	School for Special Needs Education, upper secondary department, advanced course (general, specialized)
	Junior college, short-term course
	University, short-term course
	Junior college, regular course
	Junior college, advanced course
	Junior college, correspondence course
	College of technology, regular course
	College of technology, advanced course
	Specialised training college, post-secondary course
	University, undergraduate
	University, undergraduate, correspondence course
	University, graduate school, Doctor's course
	University, graduate school, Doctor's course of pharmacy (only practical course), medicine, dentistry and veterinary medicine
	University, graduate school, Doctor's course correspondence course
	Specialised training college, general course
	Miscellaneous schools
J2	Upper secondary school, full day specialized course
	Upper secondary school, day/evening specialized course
	Upper secondary school, correspondence specialised course
	Upper secondary school, (full day/evening school), short-term course (general, integrated, specialized)
	Secondary education school (upper division), full day specialized course
	Secondary education school (Upper division), day/evening specialized course
	Secondary education school (Upper division), full day, day/evening short-term course (general, integrated, specialized)
	School for Special Needs Education, upper secondary department, Specialized course
	School for Special Needs Education, upper secondary department short-term course (general, specialized)
	College of technology, regular course
	Specialized training college, upper secondary course
J3	Upper secondary school, full day general course
	Upper Secondary school, day/evening general course
	Upper secondary school, correspondence general course
	Upper secondary school, full day integrated course (general)
	Upper secondary school, day/evening integrated course (general)
	Secondary education school (upper division), full day general course
	Secondary education school (upper division), day/evening general course
	Secondary education school (upper division), full day integrated course (general)
	Secondary education school (upper division), day/evening integrated course (general)
	School for Special Needs Education, upper secondary department, general course
J4	University, undergraduate of pharmacy (only practical course) medicine, dentistry and veterinary medicine
	University, advanced course
	University, graduate school, Master's course correspondence course
	University, professional graduate school, Professional course correspondence course
	University, graduate school, Master's course
	University, professional graduate school, Professional course
	University, professional graduate school, Graduate law school
J5	Elementary school
	School for Special Needs Education, elementary department
	Lower secondary school
	Secondary education school(lower division)
	School for Special Needs Education, lower secondary department

D1	Kindergarten
	Pre-school class in primary school
	Practical admitance courses for programmes at 5B
	Admittance courses for programmes at 5A and 5B
	Bachelor
	PhD programmes
D2	Upper sec. Higher commercial ex.
	Upper sec. Higher tecnical ex.
	HF higher prepatory examination
	Upper secondary school leaving examination
D3	Upper secondary, open vocational education
	health service assistent
	Carpenter, blacksmith, electrian upper secondary
	agriculture, horticulture, foresty, vocational education
D4	Tertiery ed., short cycle, Open education
	Tertiery ed. Open education, post secondary, open education
	Tertiery ed., medium cycle
D5	Primary level 1st-6th grade
	Lower secondary level 7th-10th grade
D6	Master
	Undivided Master degrees
D7	Home economics end needlework
	Folk- and youth high-school
D8	Generel adult education 9th-10th grade
	Higher prepatory examination, single subject education
D9	Doctorate
D10	Tertiery ed. Short cycle

Fig. 18: Labels for concept clusters obtained by the n-IRM: Jaccard

Jf2	'f6:ISCED 5' 'f17:theoretical cumulative at ISCED 5: long' 'f21:qualification structure at ISCED 5: 2nd degree' 'f24:qualification structure at ISCED 5: MA' 'f32:Minimum entrance requirement (ISCED level or other) 5A 1st' 'f56:starting age: 22' 'f66:duration: 2' 'f100:cumulative duration: 18'
Jf3	'f2:ISCED 1' 'f3:ISCED 2' 'f26:Minimum entrance requirement (ISCED level or other) 1' 'f38:starting age: 6' 'f40:starting age: 12' 'f80:duration: 6' 'f81:cumulative ducatiion: 6' 'f82:cumulative ducation: 9'
Jf4	'f29:Minimum entrance requirement (ISCED level or other) 3A' 'f30:Minimum entrance requirement (ISCED level or other) 3B' 'f47:starting age: 18' 'f68:duration: 2-3' 'f90:cumulative ducatiion: 13-15' 'f91:cumulative ducation: 13+'
Jf5	'f4:ISCED 3' 'f27:Minimum entrance requirement (ISCED level or other) 2A' 'f28:Minimum entrance requirement (ISCED level or other) 2B' 'f42:starting age: 15' 'f86:cumulative ducatiion: 12'
Jf6	'f71:duration 3 +' 'f87:cumulative ducatiion: 12.5' 'f88:cumulative ducatiion: 12-13' 'f89:cumulative ducation: 12+'
Jf7	'f109:no work based element ' 'f111: not specifically designed for adults ' 'f113:not specifically designed for part-time attendance'
Jf8	'f11:programme destination C' 'f14:programme orientation vocational'
Jf9	'f63:duration: 1' 'f64:duration: 1+'
Jf10	'f12:programme orientation General'
Jf11	'f70:duration: 3'
Jf12	'f112:specifically designed for part-time attendance'
Jf13	'f9:programme destination A'

Df2	'f11:programme destination C' 'f14:programme orientation vocational' 'f45:starting age: 16-30' 'f51:starting age: 18-50' 'f60:duration: 0.2-1' 'f67:duration: 2+' 'f71:duration 3 +' 'f73:duration: 3-5' 'f92:cumulative ducatiion: 14' 'f93:cumulative ducation: 14-15' 'f108:work based element ' 'f110: specifically designed for adults ' 'f112:specifically designed for part-time attendance'
Df3	'f3:ISCED 2' 'f10:programme destination B' 'f17:theoretical cumulative at ISCED 5: long' 'f24:qualification structure at ISCED 5: MA' 'f78:duration: 5-7' 'f100:cumulative duration: 18' 'f101:cumulative ducatiion: 18-20' 'f102:cumulative ducation: 18+' 'f103:cumulative ducation: 19'
Df4	'f43:starting age: 16-17' 'f66:duration: 2' 'f70:duration: 3' 'f86:cumulative ducatiion: 12' 'f88:cumulative ducatiion: 12-13' 'f89:cumulative ducation: 12+' 'f106:cumulative ducatiion: 21-24'
Df5	'f109:no work based element ' 'f111: not specifically designed for adults ' 'f113:not specifically designed for part-time attendance'
Df6	'f6:ISCED 5' 'f19:qualification structure at ISCED 5: 1st degree' 'f29:Minimum entrance requirement (ISCED level or other) 3A'
Df7	'f53:starting age: 20' 'f56:starting age: 22' 'f57:starting age: 24'
Df8	'f4:ISCED 3' 'f27:Minimum entrance requirement (ISCED level or other) 2A'
Df9	'f9:programme destination A'
Df10	'f12:programme orientation General'
Df11	'f47:starting age: 18'

Fig. 19: Labels for feature clusters obtained by the IRM: Jaccard (BMG)

J1	Day nursery Kindergarten School for Special Needs Education, kindergarten department Elementary school School for Special Needs Education, elementary department Specialised training college, general course Miscellaneous schools
J2	University, advanced course University, graduate school, Master's course correspondence University, professional graduate school, correspondence University, graduate school, Master's course University, professional graduate school University, professional graduate law school
J3	Upper secondary school, full day specialized course Upper secondary school, (full day/evening school), short-term course Secondary education school (upper division), full day specialized course Secondary education (Upper division), full day, day/evening short-term School for Special Needs Education, upper secondary specialized course School for Special Needs Education, upper secondary short-term course
J4	Upper secondary school, (full day, day/evening), advanced course Secondary education (Upper division), (full day, day/evening), advanced School for Special Needs Education, upper secondary, advanced course Junior college, short-term course University, short-term course Specialised training college, post-secondary course
J5	Upper Secondary school, day/evening general course Upper secondary school, correspondence general course Upper secondary school, day/evening integrated course (general) Secondary education school (upper division), day/evening general course Secondary education school (upper division), day/evening integrated course
J6	Upper secondary school, full day general course Upper secondary school, full day integrated course (general) Secondary education school (upper division), full day general course Secondary education school (upper division), full day integrated course School for Special Needs Education, upper secondary, general course
J7	Lower secondary school Secondary education school(lower division) School for Special Needs Education, lower secondary department Specialized training college, upper secondary course
J8	Upper secondary school, day/evening specialized course Upper secondary school, correspondence specialised course Secondary education school (Upper division), day/evening specialized course
J9	Junior college, advanced course College of technology, advanced course University, graduate school, Doctor's course of pharmacy, medicine etc.
J10	Junior college, regular course Junior college, correspondence course College of technology, regular course
J11	University, undergraduate University, undergraduate of pharmacy, medicine etc. University, undergraduate, correspondence course
J12	University, graduate school, Doctor's course University, graduate school, Doctor's correspondence course
J13	College of technology, regular course

D1	Generel adult education 9th-10th grade Home economics end needlework Folk- and youth high-school Higher prepatory examination, single subject education Admittance courses for programmes at 5A and 5B
D2	Upper secondary higher commercial examination Upper secondary higher tecnical examination HF higher prepatory examination Upper secondary school leaving examination
D3	Upper secondary, open vocational education Upper secondary, health service assistent, vocational education Upper secondary, carpenter, blacksmith, electrian, vocational education Upper secondary, agriculture, horticulture, foresty, vocational education
D4	Kindergarten Pre-school class in primary school Primary level 1st-6th grade
D5	Master's programme Undivided Master degrees
D6	Short cycle tertiery education, Open education Post secondary tertiery education, open education
D7	Medium cycle tertiery education Bachelor's programme
D8	Practical admitance courses for programmes at 5B
D9	Doctorate degree
D10	Lower secondary level 7th-10th grade
D11	Doctoral programmes
D12	Short cycle tertiery education

Fig. 20: Labels for concept clusters obtained by the n-IRM: BMG - JP as background knowledge (BMG)

Jf2	'f17:theoretical cumulative at ISCED 5: long' 'f21:qualification structure at ISCED 5: 2nd degree' 'f24:qualification structure at ISCED 5: MA' f32:Minimum entrance requirement (ISCED level) 5A 1st' 'f56:starting age: 22' 'f66:duration: 2' 'f100:cumulative duration: 18' 'f102:cumulative ducation: 18+' 'f112:specifically designed for part-time attendance'
Jf3	'f10:programme destination B' 'f15:theoretical cumulative at ISCED 5: short' 'f20:qualification structure at ISCED 5: intermediate' 'f68:duration: 2-3' 'f92:cumulative ducatiion: 14' 'f93:cumulative ducation: 14-15' 'f94:cumulative ducatiion: 15'
Jf4	'f5:ISCED 4' f29:Minimum entrance requirement (ISCED level) 3A' f30:Minimum entrance requirement (ISCED level) 3B' 'f47:starting age: 18' 'f90:cumulative ducatiion: 13-15' 'f91:cumulative ducation: 13+'
Jf5	'f4:ISCED 3' f27:Minimum entrance requirement (ISCED level) 2A' f28:Minimum entrance requirement (ISCED level) 2B' 'f42:starting age: 15' 'f70:duration: 3' 'f86:cumulative ducatiion: 12'
Jf6	'f71:duration 3 +' 'f87:cumulative ducatiion: 12.5' 'f88:cumulative ducatiion: 12-13' 'f89:cumulative ducation: 12+'
Jf7	'f109:no work based element ' 'f111: not specifically designed for adults ' 'f113:not specifically designed for part-time attendance'
Jf8	'f11:programme destination C' 'f14:programme orientation vocational'
Jf9	'f63:duration: 1' 'f64:duration: 1+'
Jf10	'f9:programme destination A'
Jf11	'f6:ISCED 5'
Jf12	'f12:programme orientation General'

Df2	'f11:programme destination C' 'f14:programme orientation vocational' 'f45:starting age: 16-30' 'f47:starting age: 18' 'f51:starting age: 18-50' 'f71:duration 3 +' 'f92:cumulative ducatiion: 14' 'f108:work based element ' 'f110: specifically designed for adults ' 'f112:specifically designed for part-time attendance'
Df3	'f41:starting age: 12-13' 'f43:starting age: 16-17' 'f70:duration: 3' 'f88:cumulative ducatiion: 12-13' 'f89:cumulative ducation: 12+'
Df4	'f53:starting age: 20' 'f56:starting age: 22' 'f57:starting age: 24'
Df5	'f6:ISCED 5' 'f19:qualification structure at ISCED 5: 1st degree' f29:Minimum entrance requirement (ISCED level) 3A'
Df6	'f111: not specifically designed for adults ' 'f113:not specifically designed for part-time attendance'
Df7	'f4:ISCED 3' f27:Minimum entrance requirement (ISCED level) 2A'
Df8	'f12:programme orientation General'
Df9	'f9:programme destination A'
Df10	'f109:no work based element '

Fig. 21: Labels for feature clusters obtained by the IRM: BMG - JP as background knowledge (BMG)

	Elementary school
	School for Special Needs Education, elementary department
	University, undergraduate of medicine, dentistry and veterinary medicine
	University, graduate school, Master's correspondence course
	University, professional graduate school, correspondence
J1	University, graduate school, Master's course
	University, professional graduate school
	University, professional law graduate school
	University, graduate school, Doctor's course
	University, graduate school, Doctor's course pharmacy, medicine etc.
	University, graduate school, Doctor's course correspondence
	Upper secondary school, (full day, day/evening), advanced
	Secondary education (Upper division), (full day, day/evening), advanced
	School for Special Needs Education, upper secondary, advanced
	Junior college, short-term course
J2	University, short-term course
	Specialised training college, post-secondary course
	University, advanced course
	Specialised training college, general course
	Miscellaneous schools
	Upper Secondary school, day/evening general course
	Upper secondary school, correspondence general course
	Upper secondary school, day/evening integrated course (general)
	Upper secondary school, day/evening specialized course
J3	Upper secondary school, correspondence specialised course
	Secondary education school (upper division), day/evening general course
	Secondary education school (upper division), day/evening integrated course
	Secondary education school (Upper division), day/evening specialized course
	Upper secondary school, full day specialized course
	Upper secondary school, (full day/evening school), short-term course
	Secondary education school (upper division), full day specialized course
J4	Secondary education (Upper division), full day, day/evening short-term course
	School for Special Needs Education, upper secondary , specialized course
	School for Special Needs Education, upper secondary short-term course
	Specialized training college, upper secondary course
	Upper secondary school, full day general course
	Upper secondary school, full day integrated course (general)
J5	Secondary education school (upper division), full day general course
	Secondary education school (upper division), full day integrated course
	School for Special Needs Education, upper secondary, general course
	College of technology, regular course
	Day nursery
	Kindergarten
J6	School for Special Needs Education, kindergarten department
	Junior college, advanced course
	College of technology, advanced course
	Junior college, regular course
	Junior college, correspondence course
J7	College of technology, regular course
	University, undergraduate
	University, undergraduate, correspondence course
	Lower secondary school
J8	Secondary education school(lower division)
	School for Special Needs Education, lower secondary department

	Primary level 1st-6th grade
	Master programmes
D1	Undivided Master degrees
	Doctoral programmes
	Doctorate degree
	Upper secondary higher commercial examination
D2	Upper secondary higher tecnical examination
	HF higher prepatory examination
	Upper secondary school leaving examination
	Upper secondary, open vocational education
D3	Upper secondary, health service assistent, vocational education
	Upper secondary, carpenter, blacksmith, electrian, vocational education
	Upper secondary, agriculture, horticulture, foresty, vocational education
	Kindergarten
D4	Short cycle tertiery education
	Medium cycle tertiery education
	Bachelor programme
	Generel adult education 9th-10th grade
D5	Home economics end needlework
	Folk- and youth high-school
	Higher prepatory examination, single subject education
D6	Pre-school class in primary school
	Admittance courses for programmes at 5A and 5B
D7	Short cycle tertiery education, open education'
	Post secondary tertiery education, open education
D8	Practical admitance courses for programmes at 5B
D9	Lower secondary level 7th-10th grade

Fig. 22: Labels for concept clusters obtained by the n-IRM: BMG - DK as background knowledge (BMG)

Jf2	'f1:ISCED 0' 'f10:programme destination B' 'f13:programme orientation Pre vocational' 'f15:theoretical cumulative at ISCED 5: short' 'f16:theoretical cumulative at ISCED 5: medium' 'f20:qualification structure at ISCED 5: intermediate' 'f34:Minimum entrance requirement (ISCED level or other) 5B' 'f35:starting age: 2-5' 'f36:starting age: 3-5' 'f53:starting age: 20' 'f65:duration: 1-3' 'f68:duration: 2-3' 'f92:cumulative ducatiion: 14' 'f93:cumulative ducation: 14-15' 'f94:cumulative ducatiion: 15' 'f95:cumulative ducatiion: 15-16' 'f96:cumulative ducation: 15+' 'f97:cumulative ducatiion: 16'
Jf3	'f6:ISCED 5' 'f17:theoretical cumulative at ISCED 5: long' 'f21:qualification structure at ISCED 5: 2nd degree' 'f24:qualification structure at ISCED 5: MA' 'f32:Minimum entrance requirement (ISCED level or other) 5A 1st' 'f56:starting age: 22' 'f66:duration: 2' 'f100:cumulative duration: 18' 'f112:specifically designed for part-time attendance'
Jf4	'f4:ISCED 3' f27:Minimum entrance requirement (ISCED level) 2A' f28:Minimum entrance requirement (ISCED level) 2B' 'f42:starting age: 15' 'f86:cumulative ducatiion: 12'
Jf5	'f71:duration 3 +' 'f87:cumulative ducatiion: 12.5' 'f88:cumulative ducatiion: 12-13' 'f89:cumulative ducation: 12+'
Jf6	'f3:ISCED 2' f26:Minimum entrance requirement (ISCED level) 1' 'f40:starting age: 12' 'f82:cumulative ducation: 9'
Jf7	'f5:ISCED 4' 'f8:ISCED NC' 'f90:cumulative ducatiion: 13-15' 'f91:cumulative ducation: 13+'
Jf8	f29:Minimum entrance requirement (ISCED level) 3A' f30:Minimum entrance requirement (ISCED level) 3B' 'f47:starting age: 18'
Jf9	'f109:no work based element ' 'f111: not specifically designed for adults ' 'f113:not specifically designed for part-time attendance'
Jf10	'f11:programme destination C' 'f14:programme orientation vocational'
Jf11	'f63:duration: 1' 'f64:duration: 1+'
Jf12	'f9:programme destination A'
Jf13	'f70:duration: 3'
Jf14	'f12:programme orientation General'

Df2	'f11:programme destination C' 'f14:programme orientation vocational' 'f45:starting age: 16-30' 'f71:duration 3 +' 'f73:duration: 3-5' 'f108:work based element '
Df3	'f23:qualification structure at ISCED 5: BA' 'f51:starting age: 18-50' 'f90:cumulative ducatiion: 13-15' 'f92:cumulative ducatiion: 14' 'f110: specifically designed for adults ' 'f112:specifically designed for part-time attendance'
Df4	'f6:ISCED 5' 'f16:theoretical cumulative at ISCED 5: medium' 'f19:qualification structure at ISCED 5: 1st degree' f29:Minimum entrance requirement (ISCED level) 3A' 'f67:duration: 2+'
Df5	'f43:starting age: 16-17' 'f70:duration: 3' 'f88:cumulative ducatiion: 12-13' 'f89:cumulative ducation: 12+'
Df6	'f53:starting age: 20' 'f56:starting age: 22' 'f57:starting age: 24'
Df7	'f4:ISCED 3' f27:Minimum entrance requirement (ISCED level) 2A'
Df8	'f111: not specifically designed for adults ' 'f113:not specifically designed for part-time attendance'
Df9	'f9:programme destination A' 'f109:no work based element '
Df10	'f12:programme orientation General'
Df11	'f47:starting age: 18'

Fig. 23: Labels for feature clusters obtained by the IRM: BMG - DK as background knowledge (BMG)

B Appendix II: Ontology construction methods

B.1 Formal Concept Analysis

The FCA (Ganter and Wille 1997) is the method that analyzes a relation connecting objects and their features. A *formal concept* in a context C is defined as $C = (G, M, I)$. In this definition, G and M represent a set of objects and a set of features, respectively. I refers to relations between G and M. In Fig. 13, the context *Japanese educational system* is represented as G: *(J1, J2, J3, ... J13)*, M: *(Jf1, Jf2, Jf3,..... Jf12)*, and their relations I. Each element g (e.g., $J2$) of G is expressed as $g \in G$ and. If this $g \in G$ has a feature m (e.g.,$Jf2$) that is a member of M (expressed as $m \in M$), this relation is represented as gIm. When all members of a set of objects A *(J5, J8)* that is part of G ($A \subseteq G$) shares a set of features *(Jf5, Jf6, Jf7)* in Fig. 13, it is defined as $\acute{A} = \{m \in M \mid gIm\ for\ all\ g \in A\}$. In the same way, when a set of objects *(J5, J6, J7)* are shared by all members of a set of features B (*Jf7, Jf10, Jf12*) that is part of M ($B \subseteq M$), it is expressed as $\acute{B} = \{g \in G \mid gIm\ for\ all\ m \in B\}$. A *formal concept* existing in the context (G, M, I) is expressed as (A, B) defined by $A \subseteq G, B \subseteq M, \acute{A} = B, \acute{B} = A$. Here, A and B are respectively called the extent and the intent of the concept (A, B). The set of all concepts existing in the context (G, M, I) is drawn as a *Gallois lattice* as shown in Figs. 13 and 14.

B.2 Terminological Ontology

The method of Terminological Ontology (TO) (Madsen et al. 2004) is originated from the theory of terminology. The theory of terminology was first introduced by (Wüster 1959). The original objective of terminology by (Wüster 1959) was to eliminate ambiguity from technical languages by means of standardization of terminology in order to make the terms efficient tools of communication (Cabré 2000). The traditional theory of terminology thus addresses the relation between concepts and terms, starting from concepts and focusing on the present state of the conceptual structure and its representation (Kageura 2002).

The uniqueness of TO is its *feature specifications* and *subdivision criteria.* The principles and constrains defined for the applications of feature specifications are described in detail in (Madsen et al. 2004). The most important principle is that a concept must inherit all feature specifications (i.e., features) of its superordinate concepts (Madsen et al. 2004). Another important key point is that *subdivision criteria* are strictly defined as *dimensions* and *dimension values.* It means that a given dimension can only occur for specifying features on sister concepts and a given dimension value can only appear on one of these sister concepts (Madsen et al. 2004). A dimension and its dimension values are registered as (DIMENSION : [value1, value2, ...]). In the case of Fig 24, one dimension specification under the concept *"JP education"* can be represented as (PHASE : [under school age, compulsory]). This dimension specification subdivides the concept *"JP education"* into two sub concepts *"preschool education"* and *"compulsory"* which respectively possess the features, [PHASE: under school age] and [PHASE: compulsory]. Finally, a concept must be distinguished from each of its nearest superordinate concepts as well as from each of its sister concepts by at least one feature specification (Madsen et al. 2004).

These strict principles, however, generate some difficulties in constructing an ontology when some feature specifications are considered as very important in several places in the ontology. For example, features such as [FOUNDATION: self governing] [FOUNDATION: municipality] might occur in two different occasions such as under *"elementary education"* and under *"lower secondary education"* in Fig 24. As a solution, Madsen et al. (2004) argue that this problem is solved by creating nodes, *"e: private"* and *"e: municipality"* respectively possessing the features [FOUNDATION: self governing] and [FOUNDATION: municipality] at a higher level of the ontology as depicted in Fig 24. Accordingly, subordinate concepts, *"public elementary school"* and *"public lower secondary school"* can both inherit [FOUNDATION: municipality] because of the polyhierarchical structure.

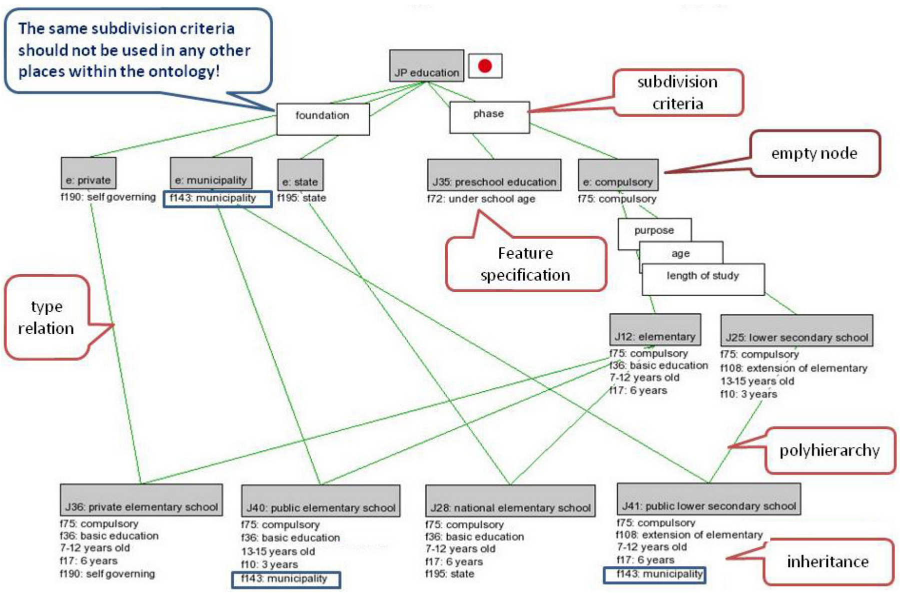

Fig. 24: Terminological Ontology principles

These strict rules are not directly applicable to the present work, since the information extracted from the cross-categorization approach only consists of the concept clusters and the feature clusters, which are rather fuzzy sets of concepts and features. In order to apply these strict principles of TO to the present work, the rules have been modified as follows:

1. A feature cluster inherited from a superordinate concept cluster can only occur on its descendant concept clusters
2. A non-inherited feature cluster can only occur in one concept cluster and its descendant concept clusters in an ontology
3. A concept cluster must be distinguished from each of its nearest superordinate concept clusters as well as from each of its sister concept clusters by at least one feature cluster.
4. For fulfilling these rules, polyhierarchical inheritance of feature clusters and generation of pseudo concept clusters are allowed.