

# Combining clustering analysis and ontological methods for identifying groups of comorbidities for developmental disorders

**Tsvi Kuflik** (Department of Management Information Systems, the University of Haifa)

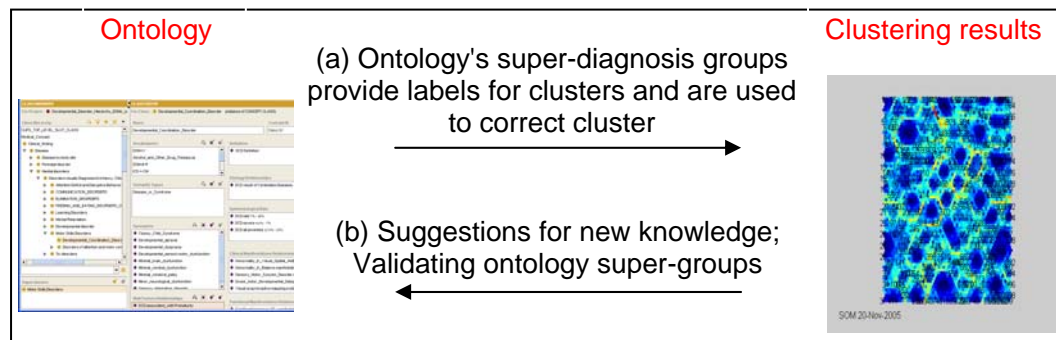
**Mor Peleg** (Department of Management Information Systems, the University of Haifa)

**Nuaman Asbeh** (Department of Statistics, the University of Haifa)

**Mitchell Schertz** (Institute for Child Development, Kupat Holim Meuhedet)

Children with developmental disorders usually exhibit multiple problems (comorbidities) rather than a single problem. For example, as reported by Gilberg [1], about 50% of children who suffer from Attention Deficit Disorder with Hyperactivity (ADHD) also suffer from Developmental Coordination Disorder (DCD) [1]. The same study reports that 60-85% of children who suffer from DCD and ADHD also suffer from Learning Disorders. Another study [2] showed that DCD is an important comorbidity of Specific Language Impairment (SLI). In some cases, a common underlying mechanism causing the different disorders in the group has been argued [1, 3]. Since comorbidities are the norm rather than the exception, diagnosis should identify groups of disorders and should not focus on single disorders. Modern classification systems, such as DSM-IV (American Psychiatric Association 1994) and ICD-10 (World Health Organization 1992), classify children with developmental and psychiatric disorders in terms of their emotional symptoms and behaviors while ignoring the underlying mechanisms that cause comorbid disorders [4], thus they are deeply unsatisfactory for many parents, front-line clinicians, and academic researchers [5]. As a result there is no agreed-upon terminology in the field of developmental disorders.

An objective of our research was to create an ontology that would lead to a consistent and agreed-upon terminology, accounting for groups of comorbid diagnoses. A second related objective was to identify the groups of comorbid diagnoses and represent them as super-diagnoses groups in our ontology that will be suggested as candidates for new terms in the field that reflect the groups of comorbidities in developmental disorders.



**Figure 1** A schematic representation of combining the clustering results and the ontology. (a) The super groups at the ontology provide labels for the clusters and super-diagnosis groups at the ontology are used to interpret and improve clustering results; clusters that contain more than one population of vectors are split; clusters with similar labels are joined. (b) In addition, clusters found by clustering techniques direct future literature searches for comorbidity relationships and super-diagnosis groups and entering such detailed knowledge into the ontology, as well as validating the super-diagnoses defined in the ontology.

To this end, we developed an initial ontology based on a literature survey. We combined the ontology development process with the application of clustering analysis to available clinical data of children with developmental disorders; we defined super-diagnosis groups using the knowledge in the ontology, and we also identified clusters that exist in the clinical data. We propose a methodology that combines clustering results and ontology development, so that each approach complements the other in identifying groups of comorbidity for developmental disorders. We show how the clustering results can be used (1) to validate the groups of super-diagnoses defined in the ontology, by evidencing these groups as clusters in the clinical data, and (2) to suggest directions for future ontology development by identifying possible comorbidity relationships in the clinical data, and we show how the ontology can be used (1)

to interpret the clustering results by labeling clusters using the groups defined at the ontology and classifying these clusters based on evidence from the ontology, and (2) to improve the clustering results by applying ontology-based corrective operations on the clustering results, as illustrated in Figure 1. Demonstrating the ontology's potential to interpret and improve the clustering results can establish the credibility of our ontology, by making explicit the way in which the ontology reflects the comorbidity phenomena through the super-diagnosis representation in the ontology.

We evaluated our methodology and showed the improvement of the clustering results after applying knowledge contained in the ontology, using a measure of agreement with a clinical expert partition. In addition, our methodology allowed us to identify and interpret more groups of comorbidities in the final clustering results after applying the ontological knowledge as compared to the original clustering results. These groups were defined as Actual Groups in the ontology and are suggested as candidates for new terms in the field that reflect the groups of comorbid developmental disorders. In this way, we demonstrated the ontology's potential of defining literature-based and data-evidenced terminology that reflects the comorbidity phenomena in the domain.

We also reported a framework for taking an inconsistent, manually-constructed set of diagnoses and forming a consistent set of diagnostic terms that was based on established medical vocabularies. The framework that we propose showed its potential to identify the most comprehensive vocabulary (SNOMED-CT in our study) of the clinical vocabularies that covered a medical sub-domain, suggesting that it may serve as the main vocabulary used in the clinical sub-domain.

The importance of this research is manifold. It contributes to medical informatics research and to clinical practice. An ontology of developmental disorders will (1) increase the consistency of terminology by which clinical diagnoses in the field are described, (2) allow better communication between experts in the domain and ease collaboration between researchers, and (3) yield consistent sets of data that can be analyzed to discover diagnoses clusters, as done in this research. Once completed, the ontology would (4) serve as an online aid for reminding clinicians of the definitions of particular diagnoses as well as for directing clinicians in examining relevant comorbidities. The systematic examination of naturally occurring clusters will provide researchers with much more homogenous groups of subjects. This should (5) significantly contribute to our ability to investigate etiology as well as to evaluate treatment and intervention programs, which that are sorely lacking in the field. In this research we applied unsupervised machine learning methodology on clinical data in exploratory way and combined it with domain knowledge, it (6) sets the stage for future studies applying supervised machine learning methodologies to clinical data of developmental disorders; once the groups of comorbid developmental disorders are identified, classification methods can be applied to patient data. This will enable the development of classifier systems that can assign patients to the accurate group to which he/she belongs based on the diagnoses and finding in his/her patient's record, and also can direct the clinicians to examine additional diagnoses that are usually common in this group. In addition, classification methods can contribute to the study of the underlying mechanisms that cause the disorders in each group, by training the system (using the already known groups and patient data) to learn the attributes (e.g., neurodevelopmental findings) that are shared by members of the same group or that most effectively discriminate among members of different groups.

[1] Gillberg C. Deficits in attention, motor control, and perception: a brief review. *Arch Dis Child*. 2003; 88: 904-910.

[2] Webster RI, Majnemer A, Platt RW, Shevell MI. Motor function at school age in children with a preschool diagnosis of developmental language impairment. *J Pediatr*. 2005 Jan;146(1):80-5.

[3] Piek JP, Dyck MJ. Sensory-motor deficits in children with developmental coordination disorder, attention deficit hyperactivity disorder and autistic disorder. *Hum Mov Sci*. 2004; 23(3-4): 475-88.

[4] Wiener JM, Shaffer D. Concepts of Diagnostic Classification. In: Wiener JM, editor. *Textbook of Child and Adolescent Psychiatry*. Washington, DC: American Psychiatric Press; 1997: 23-32.

[5] Szatmari P. The classification of autism, Asperger's syndrome, and pervasive developmental disorder. *Can J Psychiatry*. 2000; 45(8): 731-8.