Digital Healthcare Empowering Europeans
R. Cornet et al. (Eds.)

doi:10.3233/978-1-61499-512-8-221

Enrich classifications in psychiatry with textual data: an ontology for psychiatry including social concepts

Marion RICHARD [a,d, 1], Xavier AIMÉ [a,d], Marie-Odile KREBS [b,d] and Jean CHARLET [a,c,d]

[a] *INSERM, U1142, LIMICS, F-75006, Paris, France;*
Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France;
Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430, Villetaneuse, France.
[b] *Laboratoire de Pathophysiologie des Troubles Psychiatriques, Centre Hosp. Sainte-Anne*
[c] *Assistance publique - Hôpitaux de Paris, France*
[d] *GDR Psychiatrie CNRS 3557, Paris, France.*

Abstract. We propose a modular approach to develop an ontology of psychiatry, ONTOPSYCHIA, based on Patient Discharges Summaries (PDS) and divided into three modules (i.e. social, mental disorders and treatments). We decided to take into account the social aspects of the patient life described in PDS to consider information such as family history, social environment or education.

Keywords. Ontology for psychiatry; classification systems; data interoperability; knowledge representation in social life;

Introduction

Descriptions of cases with mental disorders are various, heterogeneous and encompass many aspects from personal history, family history and many facets of dysfunctions and subjective symptoms. The large amount of information in psychiatry, stresses the need for improvement of our capability to detect, quantify the behaviour and model the symptoms or dysfunctions associated with psychiatric disorders. Tackling the challenge of modeling information is one field of knowledge engineering. Ontologies are created for many purposes, including semantic researches, data retrieval, formal reasoning, etc.

We propose ONTOPSYCHIA, an ontology for psychiatry. The use of ONTOPSYCHIA, associated with dedicated tools, will allow (1) to perform semantic research in Patient Discharges Summary (PDS), (2) to represent the comorbidity, (3) to index PDS for the constitution of cohorts, and (4) to identify resistant patient's profiles. We also postulate that this approach can help reaching a consensus on descriptive categories of mental disorders.

We decided to take into account the social aspect of patient life in our ontology, to be able to correlate social context/event and mental health.

[1] Corresponding Author.

1. Materials and methods

We follow the methodology illustrated by Charlet et al. [1], which combines a top-down and a bottom-up approach. This method allows an access to terms which represent concepts in use. During the construction of the ontology we focus on information contained in PDS (top-down approach) and return to domain classifications only for the enrichment phase (bottom-up approach). We first collect and anonymize 8,000 PDS of Sainte-Anne Hospital (the largest psychiatric hospital based in Paris, France). In order to anonymize the PDS, we used MEDINA [2] software (which has been developed following HIPPAA law). In order to extract specialized terms of the corpus, we used MELT [3] and YATEA [4]. After the extraction we faced an exhaustive list of 20,000 terms related to the psychiatry domain. We chose a modular approach to deal with the amount of knowledge available in psychiatry, and decided to build three modules: (1) social aspects of a patient life, (2) mental disorders/illnesses, and (3) treatments. Each concept is denoted both by a preferred label in English and in French, and by alternative labels (synonymous terms, acronyms). The labels come from the list of extracted terms and from nomenclature (DSM, ICD, and SNOMED). The choice of the preferred label was done consistently with the evaluation of the lexical prototypicality as suggested by [5].

We decided to include the social aspects of patient life in our ontology, to correlate social context/event and mental health. We based on two resources: SNOMED 3.5FR/Social which aims at representing the social aspects that may influence the patient's health and treatment; The *Family Health History* ontology (FHH) [6] conceptualizes (1) members of family and relations between them, (2) and health states with diagnosis.

2. Results and Discussion

Currently, we are completing the module about social life (cf. Table 1 & Table 2). The last step is the conceptualization validation by psychiatrists and neuropsychologists of Sainte-Anne Hospital. We divided the modeled knowledge in four major classes: (1) *Attribute* – adjectives describing more precisely the concepts; (2) *Concept of social life* – to organize all concepts extracted from the corpus and dealing with social aspects (education, social situation, civil situation, etc.); (3) *Human being* – to represent all human entities; and (4) *Groups* – to formalize the primary considered group, such as *company*, *institution*, *service*. Finally we use PELLET[2] reasoner to assess the module consistency.

Table 1. Social module metrics

Number of concepts	Maximum depth	Maximum number of siblings	Average number of siblings	Maximum depth
1331	6	14	5	6

[2] http://clarkparsia.com/pellet/download

Table 2. Statistical about social module and each main class[3,4]

Social module	Classes	Equivalent Classes	Objects Properties
OntoPyschia	1331	454	274
Concept about social life	742	280	105
Human Being	302	224	226
Group	132	24	9

Firstly, module's subject is represented by the number of concepts under "Concept about social life". Secondly, we notice that concepts under "human being" are highly connected, which can be explained by connections between individuals: 141 object properties are only describing relations between family members, around 70 are connected with concepts about social life and others are connected with others human being who are not family members (but friends, boyfriends, etc.). We also notice that almost all classes under "human being" are defined, because a person's description always depends on several concepts (e.g. every female/male person has an axiom to identify the gender). Thirdly we can note that "attribute" sub-class contains essentially undefined classes and only few objects properties. This is because all these concepts are adjectives used to define other classes.

We based the initiation of the validation phase on SNOMED 3.5FR/Social. However, we quickly needed to move away from this modeling, because of association of concepts that do not satisfy our expectations (e.g. "mendiant" (~ beggar) and "fainéant" (~ lazy person) are considered equivalent).

In opposite, we took all FHH defined classes because they all model family members.

Data in the psychiatry domain are still difficult to synthesize, but the social module shows that we can expect to build a model that fit our objectives. Now, we will develop the two other modules on (i) diseases and (ii) treatments.

References

[1] Charlet, J.; Bachimont, B. & Jaulent, M.-C. Building medical ontologies by terminology extraction from texts: an experiment for the intensive care units. *Computers in biology and medicine, Elsevier,* 2006*, 36*, 857-870

[2] Grouin, C. & Névéol, A. De-identification of clinical notes in French: towards a protocol for reference corpus development. *Journal of biomedical informatics, Elsevier,* 2013

[3] Denis, P. & Sagot, B. Coupling an annotated corpus and a lexicon for state-of-the-art POS tagging. *Language resources and evaluation, Springer,* 2012*, 46*, 721-736

[4] Aubin, S. & Hamon, T. Improving term extraction with terminological resources. *Advances in Natural Language Processing, Springer,* 2006, 380-387

[5] Aimé, X. & Charlet, J. Preferred label validation by Lexical Prototypicality Gradient: a use case on a Rare Diseases Ontology *on Capturing and Refining Knowledge in the Medical Domain (K-MED 2012),* 2012, 36

[6] Peace, J. & Brennan, P. Ontological representation of family and family history. *AMIA... Annual Symposium proceedings/AMIA Symposium. AMIA Symposium,* 2006, 1072-1072

[3] We do not deal the issue of Disjoint Classes in this paper because modeling them at this stage would be premature.

[4] Statistics about Sub-Classes are not provided because as currently they are indifferently defined as axioms or equivalent classes. A cleaning exercise will be performed before the uploading on BioPortal.