 DOI 10.1109/MODELS-C.2019.00110

A Semantic Model To Fight Social Exclusion

Victor-Alejandro Ortiz[†], Montserrat Estañol[*†], Maria-Cristina Marinescu[†],
Maria-Ribera Sancho[*†], Ernest Teniente[*], and Carmen Rueda[‡]
[†]Barcelona Supercomputing Center, Barcelona, Spain. E-mail: {victor.ortiz, maria.marinescu}@bsc.es
[*]Universitat Politècnica de Catalunya, Barcelona, Spain. E-mail: {estanyol, ribera, teniente}@essi.upc.edu
[‡]Àrea de Drets Socials, Ajuntament de Barcelona, Barcelona, Spain. E-mail: crueda@bcn.cat

***Abstract*—This work presents a semantic model meant to help with the identification and prediction of individuals at risk of social exclusion. The model is based on the self-sufficiency matrix, a tool that evaluates a person's self-sufficiency in different areas, and that is used by Barcelona's City Council. Existing data sources can then be mapped to this model, in order to analyze, query, and visualize the data.**

***Keywords*—semantic technologies, modeling, social exclusion, self-suffiency matrix**

I. Introduction

Inequality and exclusion are typical of human societies. The role of social workers is to understand the situation of individuals that ask for help - their social, cultural, economic context, and the severity of their situation - then act accordingly to help them (this process generally involves an extensive interview). Different skills/experience level may be decisive in how the social worker is able to identify key problems and focus on them. Our goal is to provide them with tools to assist them and ensure standardization and quality across interviews.

To enable analysis and prediction, the first step is data integration from heterogeneous sources, e.g. text/ CSV files, relational/ vertical and NoSQL databases [1].These sources may change across different urban settlements or over time and in this context, semantic technologies can show their advantages and make data integration scalable and cost-effective.

Our main contribution is a domain ontology for social exclusion, through which we can integrate all types of heterogeneous data sources by mapping (potentially) changing data schemas to the unchanging model. One of our goals is creating a model useful to the Social Services of Barcelona's City Council. This work is part of the LinDaFIX project[1], whose ultimate aim is to improve the identification and prediction of individuals at risk of social exclusion.

II. Background

A. Self-sufficiency matrices

One of the challenges faced by public service providers is being able to demonstrate efficiency and effectiveness of the outcomes of their policies [2]. One way of doing so is by using a self-sufficiency matrix (SSM). Self-sufficiency is defined as the ability of individuals to provide for themselves, and the matrix defines several evaluation areas, such as finances or work and education. The matrix is meant to provide a holistic view of the individual at a certain point in time as recommendations for the social workers. This approach to evaluation has been shown to be reliable and effective [2].

B. Semantic technologies

Semantic technologies are currently the most effective alternative to model complex systems and enable cross-domain integrated access to heterogeneous data sources. A semantic model can be viewed as a common, domain-specific vocabulary, which may be reused across cities and extended with new categories without modifying the application or data sources. Differently from databases and ETL processes, semantic models are easy to maintain and refine, independently of the data. Using technologies such as Ontology-Based Data Access (OBDA) [3] it is possible to keep the data in the source, therefore, changes to the data are available immediately. The connection between model and data is achieved by semi-automatically mapping the data schema to the model [4] [5].

III. System Architecture

The LinDaFIX project spans several areas: semantic technologies, machine learning, automated reasoning, and natural language processing. Fig. 1 shows the general framework of the project's goals. Here we focus on the data integration via semantic technologies, and the modeling of inequality indicators as part of the domain ontology. A subset of these will be based on the SSM. Once data is mapped to the model it can be analyzed with statistical methods, queried and visualized, and used to compute the indicators.

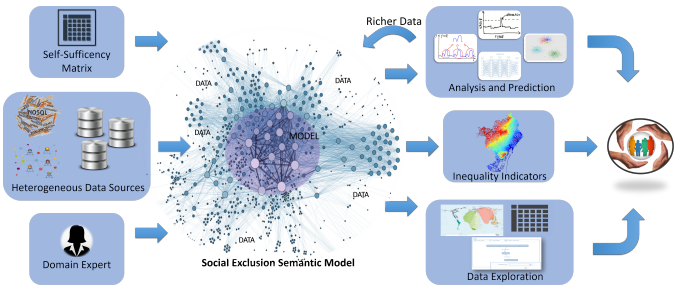

Fig. 1. General LinDaFIX framework

IV. Methodology

One of the areas in which the Social Services Department of Barcelona currently needs help is in generating individual assessments using SSMs. The purpose is twofold: to have a better understanding of where to focus during the interview,

[1]LinDaFIX is funded by RecerCaixa, a program which backs research projects that can transform society and offer solutions to social problems.

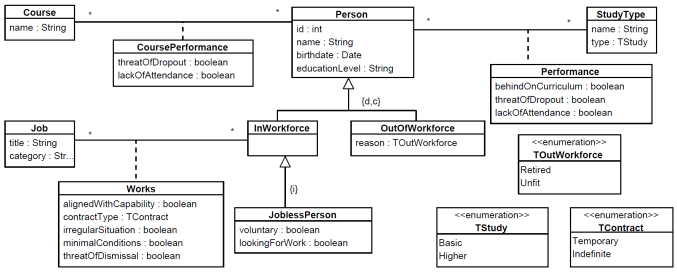

Fig. 2. UML class diagram representing a fragment of our semantic model.

and to standardize the evaluation process to ensure that all social services centers and professionals apply the same criteria and treat all their users equally.

This is why, in addition to a more generic ontology based on the literature, we have decided to start from the SSMs to build a more practical model. After concluding the iterative evaluation of this model with the Social Services Department, we will contrast and complement it with the generic one.

Fig.2 shows a fragment of this semantic model, corresponding to the *Work and Education* area, represented in a UML a class diagram. This model will be later translated to OWL, with those modifications necessary to account for the differences in the open vs closed world semantics. Its main concept is the *Person*. We distinguish between people who are in the workforce (*InWorkforce*) or not (*OutOfWorkforce*). Someone may not be in the workforce due to retirement or medical reasons. For individuals who are in the workforce but jobless we store additional information on whether they are unemployed by choice, or if they are currently looking for a job. For the employed individuals, the model stores information such as the contract type, whether the job fulfills a minimal set of safety conditions, or if there is a threat of dismissal, as all of these may put a person at risk of exclusion. This information is represented via relationships and attributes in association classes. We also keep track of many other parameters such as education level and any kind of education program that they may be attending.

The model also has additional restrictions which are not shown in the diagram, and that we represent using OCL. e.g., *an InWorkforce person can only be jobless if and only if he does not have a job*[2]. Another interesting point to consider are those restrictions that may only be applicable to certain regions or countries. For instance, in Mexico, we could state that *an InWorkforce person who is in an irregular situation (`irregularSituation == true`) at work, is considered to have an indefinite contract (`contractType == TContract::Indefinite`).*

Once the City Council's data sources are fully mapped to the model, it will become possible to automatically fill in the values of the individuals' SSM as recommendations for the social workers, while users and tools will be able to access data from heterogeneous data sources in a unified way, through the ontology.

[2]In OCL: `context InWorkforce inv: (self.job->isEmpty() implies self.oclIsTypeOf(JoblessPerson)) and (self.oclIsTypeOf(JoblessPerson) implies self.job->isEmpty())`

V. Related work

Application of semantic web in the social goods domain is encouraged by hosting workshops in recognized conferences (e.g. ISWC), nonetheless, as far as we can tell there is no contribution on poverty, social exclusion, or social services. Works like [6] propose using ontologies to describe the poverty domain and promote the discussion about data analysis to support policy makers in the generation of a public plan to help prevent poverty and studies such as [7] use machine learning algorithms to predict the risk of suffering chronic social exclusion. Regardless of that, we are not aware of any approach that uses semantic modelling in conjunction with machine learning techniques applied to heterogeneous sources of social data to identify groups and individuals in a vulnerable situation - which is our ultimate goal.

VI. Potential practical impact

The impact of the larger (LinDaFIX) project in terms of improving the life of the citizens is straightforward - protecting the most vulnerable members of the society and striving towards more equal opportunities. The semantic model we are presenting in this paper is the enabler for the integration of heterogeneous data sources via a common vocabulary. Integrated heterogeneous data enables querying and analysis in a cost-effective manner. This impacts the quality of the data and improves the precision of detecting and predicting vulnerability situations. For social services, it means reaching more people with fewer resources, where it most matters.

From a purely technical point of view, building a domain ontology is not new. It is however a novel approach in terms of advancing the state-of-practice in the social services domain, especially so because of the practical methodology (based on SSMs) and the feedback we get as result of our collaboration with the Social Services Department of the City Council.

Acknowledgments: This work is partially supported by the Semiotic project, funded by Ministerio de Economia, Industria, y Competitividad (TIN2016-78473-C3-2-R).

References

[1] B. Kolev, P. Valduriez, C. Bondiombouy, R. Jiménez-Peris, R. Pau, and J. Pereira, "Cloudmdsql: querying heterogeneous cloud data stores with a common language," *Distributed and Parallel Databases*, vol. 34, no. 4, pp. 463–503, 2016.

[2] T. Fassaert, S. Lauriks, S. van de Weerd, J. Theunissen, M. Kikkert, J. Dekker, M. Buster, and M. de Wit, "Psychometric properties of the Dutch version of the Self-Sufficiency Matrix (SSM-D)," *Community Mental Health Journal*, vol. 50, no. 5, pp. 583–590, Jul 2014.

[3] D. Calvanese, P. Liuzzo, A. Mosca, J. Remesal, M. Rezk, and G. Rull, "Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire," *Engineering Applications of Artificial Intelligence*, vol. 51, pp. 212–229, 2016.

[4] S. Pavel and E. Jerome, "A Survey of Schema-Based Matching Approaches," *Journal on Data Semantics IV*, vol. 3730, no. May, 2005.

[5] G. Kappel, H. Kargl, G. Kramler, A. Schauerhuber, M. Seidl, M. Strommer, and M. Wimmer, "Matching Metamodels with Semantic Systems - An Experience Report," in *Model Management und Metadaten-Verwaltung - BTW 2007 Workshop, Aachen, Germany*, no. May, 2007.

[6] A. Panori, C. Kakderi, and P. Tsarchopoulos, "Designing the Ontology of a Smart City Application for Measuring Multidimensional Urban Poverty," *Journal of the Knowledge Economy*, 2017.

[7] E. Serrano, M. C. Suárez-Figueroa, J. González-Pachón, and A. Gómez-Pérez, "Toward proactive social inclusion powered by machine learning," *Knowledge and Information Systems*, vol. 58, no. 3, pp. 651–667, 2019.