An ontological probabilistic approach to the breast cancer problem in semantic medicine

Bevilacqua V[1], Cucci F[1], Santarcangelo V[1], Mastronardi G[1]

Motivation

We show the application of semantic web and RDF (Resource Description Framework) for organization of Breast Cancer patient for diagnostic.The analysis is based on patient's genomic map, information about risk factors and family history.

Methods

The RDF data model consists of resources, properties and values. Properties are the relationships that bind together the resources and values, and are also identified by URIs. The RDF data model allows to define a simple model to describe the relationships between resources, in terms of properties identified by a name and its values. However, RDF data model provides no mechanism for declaring these properties, or to define relationships between these properties and other resources. For this task there is RDF Schema. RDF graphs can be useful to describe the patient as an URI with his properties and relations with other people which shares a family history (characterized by his health problems). This could be useful to have a complete clinical view of a person (this could be implemented thanks to the family doctors) and to help doctors who haven't information about a patient. In particular in this work we show the application of this semantic medicine approach is an experimental ontology for the Breast Cancer Diagnose. This ontology model information from a world focused on real cases of patients previously affected by breast cancer, whereas biological data of the patient (age, menopause, etc..) and specific characteristics of disease (tumor size, lymph nodes involved, grade, location, etc.).A short list of general risk factors crucial for breast cancer is advantage age (65 age), late menopause (55 years or after), family history (particularly in cases where one or more close relatives degree (mother, sister and daughter) has been diagnosed with cancer at an early age); excessive consumption of alcohol; have had first child after thirty years old; -have been submitted recently to hormonal therapy based on only estrogen; early menarche (first menstrual period before 12 years old); obesity after menopause; BRCA1 e BRCA2 mutations. In our study we test our approach with Anna. Anna is a young woman of twenty-three, his grandmother died of breast cancer, two other aunts have died fifty years for ovarian cancer, and a third aunt of thirty-nine has breast cancer. The gynecologist suggests that in the family of the patient there is a mutation in one of two genes that control susceptibility to the development of

[1] Department of Electrical and Electronics Engineering - Polytechnic of Bari

breast or ovarian cancer: the gene BRCA1 and BRCA2. To Anna veins discovered a suspicious lump in her right breast and the doctor prepares a biopsy and asks Anna to perform a DNA test, to detect a genetic mutation. It's necessary to trace the exact form of the gene mutated in a parent carrier of a change and identify it. Anna learns that her aunt sick from a genetic test was a carrier of a rare mutation in BRCA1 and BRCA2 genes. Analysis of the DNA of Anna turns out that she carries the same mutation in the aunt. The RDF mapping can organize these information and prevents the loss of data fundamental for the study of pathologies in family. RDF, with the use of SKOS and OWL is used to model these scenarios and organize important data. It can be interrogated by SPARQL queries.This information need to be interpreted and elaborated by an algorithm. This algorithm can be static (his rules are defined) or dynamic (his rules are obtained by a RDF Knowledge Base). The output is a prediction about the probability to get Breast Cancer.So far we have considered values defined for properties, but the diagnosis of diseases: symptoms, causes and consequences are characterized by a degree of uncertainty, which complicates the conceptualization of these sectors in formal ontology and thus limits the ability of understanding at the level of machine. Then it's necessary to declare properties used to express probabilistic knowledge, generic and concrete, of fundamental importance for the functioning of probabilistic reasoning. Each factor of risk has an own range in which is defined and a probability (pi) associated. Person considered can have some probability (pi) related with its characteristics, in fact, it has a probability p1 associated at his age, a probability p2 associated at the age of menarche if it is in the range [10-12], a p3 probability associated at his age of menopause (if it's over 60), a p4 probability associated to a positive outcome of a Last Echograpy, a p5 probability based on the birthday of its first daughter. These information are used by the algorithm to analyze the RDF Data of the Patient and produce a prediction about the probability to get Breast Cancer.

Results

Possible results of this approach could be an implementation of a system that through an user-friendly GUI can create, have access, modify and update the RDF document associated to every patient. This system can be implemented integrating the swi-prolog parser.

Availability

http://www.vitoantoniobevilacqua.it/

Contact e-mail

bevilacqua@poliba.it

Supplementary information

available on http://www.vitoantoniobevilacqua.it News: BITS 2010 materials