Computers in Biology and Medicine

Computers in Biology and Medicine 36 (2006) 871–892

www.intl.elsevierhealth.com/journals/cobm

Building and using a medical ontology for knowledge management and cooperative work in a health care network

Rose Dieng-Kuntz*, David Minier, Marek Růžička, Frédéric Corby, Olivier Corby, Laurent Alamarguy

INRIA, ACACIA Project, 2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex, France

Received 5 April 2005; accepted 5 April 2005

Abstract

In the context of a health care network, we describe our method for reconstitution of a medical ontology via the translation of a medical database (DB) towards RDF(S) language. Then we show how we extended this ontology among others through natural language processing of a textual corpus. Then, we present the construction of a tool called "Virtual Staff", enabling a cooperative diagnosis by some of the health care network actors, by relying on this medical ontology and on the creation of SOAP and QOC graphs.

Keywords: Ontologies; Knowledge management; Semantic web; Cooperation; Medical application; SOAP; QOC

1. Context: needs in a health care network

The project "*Ligne de Vie*" (*Life Line*), in collaboration with the Nautilus society and the SPIM laboratory, aims at developing a knowledge management tool for a health care network [1–3].

According to article L6321.1 of the French Code of Public Health, "Health networks aim at supporting access to care, coordination, continuity or interdisciplinarity of the medical taking in care, in particular among those which are specific to some populations, pathologies or medical activities. They ensure a taking in care, adapted to the needs of the person as well for education to health, for prevention, for

* Corresponding author. Tel.: +33 4 92 38 78 10; fax: +33 4 92 38 77 83.
E-mail addresses: Rose.Dieng@sophia.inria.fr (R. Dieng-Kuntz), Oliver.Corby@sophia.inria.fr (O. Corby).

0010-4825/$ - see front matter
doi:10.1016/j.compbiomed.2005.04.015

diagnosis and for care. They can take part in public health actions. They proceed to evaluation actions in order to guarantee quality of their services. They can be constituted of liberal health professionals, job doctors, health establishments, health centres, social or medico-social institutions and organisations with medical or social vocation, as well as users' representatives."

In order to analyse the concrete needs of health care network in general, we had interviews with a physician and with a nurse. Specialised in a particular domain or in a specific pathology, a health care network is a health network gathering all the actors intervening in the care or follow-up processes. The objective of the network is to ease (a) communication and collaboration among these actors in spite of their physical distance, (b) the regular follow-up of the patient and (c) the respect of best practices inside the network. The patient must be guided towards relevant medical actors, who must be informed about the patient's state and who may gather (sometimes virtually, through synchronous or asynchronous communication tools) in order to work on the patient's record. For example, cystic fibrosis demands daily care throughout the patient's life and requires several kinds of professionals: paediatricians, physicians, gastro-enterologists, chest specialists, nurses, physiotherapists, psychologists, dieticians, social workers, as well as the patient, his/her family, the school doctor or the job doctor. The network must ease knowledge sharing about the patient record, among all these actors from various competence domains, with a user-tailored presentation of the information.

The ideas proposed by Nautilus SARL for launching the project *Ligne de Vie* stem from 15 years of experiences in health domain and from contacts with a network dedicated to diabetes: Nautilus offers a software for management of electronic medical record, Episodus, relying on a problem-oriented vision of the patient record and articulated around the notion of "Life Line" enabling to represent the life of the patient from his/her birth till his/her death with all the health problems encountered by this patient. As long as the patient still suffers from a problem, this one remains open. When the patient gets completely cured from this problem, the problem is closed.

The technical choices of the project *Ligne de Vie* are motivated as follows:

- The need to build *a referential common to the network actors* led us naturally to rely on a medical ontology. Moreover, this ontology should help, through inferences, to improve information search on the documents shared or accessible by the network members.
- The need to take into account *actors from various competence domains* incited us to study *viewpoint modelling* [4] in the ontology, in the patient records, or in the presentation of the results to the user after his/her query.
- The need of cooperative work between the actors of a network led to the idea of our partner Nautilus to develop a software called *Virtual Staff* in order to enable the members of the network to *visualise their collective reasoning*: in order to diagnose the pathology of the patient (according to the symptoms expressed by the patient, the observations or analyses of the doctor and the already known health problems of this patient), or in order to determine the best possible therapeutic procedures. This Virtual Staff should supply the users with *a service of support to cooperative reasoning*, during the phases of elaboration of diagnosis or therapeutic decision and *a service of constitution of an organisational memory* [5]—the memory of decisions of the community constituted by the members of the health care network.

This article focuses on the work of the Acacia team on the medical ontology and on the Virtual Staff: we will present our method for reconstituting the Nautilus ontology—this method relies on a translation from a medical DB towards the RDF(S) language; then we will show how we extended this ontology after analysis of a textual corpus. Then, we will describe the Virtual Staff based on this ontology. Last, we will conclude on the interest of this work and on our further work planned.

2. Reconstitution and extensions of the medical ontology Nautilus

2.1. Role of the ontology

In the project *Ligne de Vie*, the ontology aims at modelling some knowledge on general medicine or specialised medicine, on health care networks and their actors: the ontology will represent the conceptual vocabulary common to the actors of a health care network, and will be used in the Virtual Staff and for information retrieval. This ontology will be the kernel of a *medical semantic Web dedicated to the health care network*.

The medical community has long been sensitised to the need of modelling its knowledge and of making its terminologies explicit. Therefore, there exists several terminological or ontological resources in medical domain: GALEN [6,7], MENELAS [8,9], ON9 Library [10,11], SNOMED RT [12,13], UMLS [14] are ontologies/ontology libraries/thesauri/meta-thesauri that model a part of the medical domain. We could have relied on one of them but for reasons of collaboration, we had to use the Nautilus medical DB developed by our industrial partner.

In order to make the Nautilus ontology more understandable, it was important to represent it in a standard knowledge representation formalism: we chose RDF(S) [15] to which our semantic search engine CORESE [16–18] is dedicated. Therefore, we developed a translator from the Nautilus DB internal format towards RDF(S) language. Using a "reverse engineering" approach relying on the analysis of this DB coding principle, we decoded this DB in order to reconstitute a Nautilus ontology represented in RDF(S) so as to explore it and validate it via a semantic search engine, to annotate and search documents through this ontology, to guide the Virtual Staff, etc. Fig. 1 shows the architecture of the medical semantic Web dedicated to a health care network.

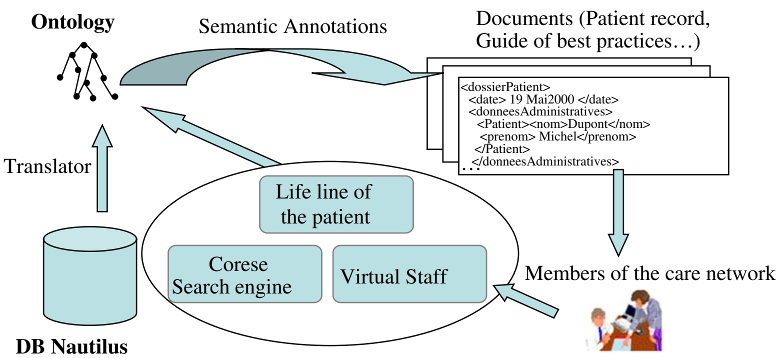

Fig. 1. Architecture of a medical semantic web for a health care network.

Table 1
Characters coding the root concepts in the Nautilus DB

Code	Root concept	Code	Root concept	Code	Root concept	Code	Root concept
A	Anatomy	B	Biology	C	Foreign body	D	Physician
E	Patient's state	F	Malformation	G	Diagnostic/therapeutic gesture	H	Person
I	Active principle	J	Classification description	K	Time	L	Administration
M	Movement, position	N	Treatment	O	Material	P	Pathology
Q	Physiology	R	Pathogenous agent	S	Symptom	T	Laboratory test
U	Location	V	Measurable data	W	System	X	Histology
Y	Psychological problem	Z	Document, visual aspect	0	Chapter	1	Pure adjective
2	Units	3	Calculation formula	4	Verb	5	Fear, discussion, request
6	Classification code	7	Drug distribution unit	8	Drug excipient		

2.2. Discussion: database vs ontology

A preliminary question was the status of the Nautilus DB: Could one regard it as an ontology? Some DB researchers assimilate a DB conceptual schema to an ontology. Admittedly, if an explicit principle is available of how the concepts of an ontology are coded into the internal format of a DB, one can regard this DB as an implementation of the ontology. But it is important to represent the ontology explicitly in a form understandable by a human user, and in a knowledge representation formalism. We thus looked upon Nautilus as a medical ontology encoded in an internal format but without a representation readable by a human user. The reconstitution of the Nautilus ontology thus required the reconstitution of the hierarchy of concepts and of the set of relations implicit in the DB.

2.3. Translation algorithm

The Nautilus DB is a Paradox database composed of:

- A table *Lexique.db* describing more than 36 000 medical terms.
- A file *Concept.txt* containing the list of the root concepts: *Anatomy*, *Physiology*, *Symptom*, *Pathology*, *Physician*, *Laboratory Test*, etc. Table 1 gives examples of characters coding such root concepts.
- A table *Savoir.db* describing the relations between the terms, with four types of relations: the ES specialisation relation (*is-a*), the E0 relation (non-transitive *is-a*), the AT relation (*located-on*) and the ME relation (*measured-with*).

Remark. In the Nautilus database, the E0 relation is used to indicate that a body part (such as eye, hand, foot. . .) or an organ (such as lung or kidney) is a "generic anatomic object", which means that there exists a "left eye" and a "right eye", etc.

To each term appearing in *Lexique.db* corresponds a single, six-character long, code (e.g. *PABCF*1): the first character thus enables to find the associated concept root (for example, the character P corresponds to the root concept *Pathology*). The first five characters of the code associated to a term characterise the concept; the sixth character enables to distinguish the synonymous terms. For example, the terms "*derivation abscess*" and "*fixation abscess*" coded, respectively, *PABCF*1 and *PABCF*2 in the *Lexique.db* table are two synonymous terms naming the same concept, coded *PABCF* (P indicating that this concept is attached to the root concept *Pathology*). The translation algorithm enables to reconstitute in the concept hierarchy that the *Pathology* root concept has as direct or indirect subconcept a concept named by the synonymous terms "*derivation abscess*" and "*fixation abscess*".

Using the *Concept.txt* file, the translator translates the root concepts into RDFS classes that are direct subclasses of the *conceptNautilus* class. The codes of the *Lexique.db* table, without ES relation in the *Savoir.db* table, are also translated into RDFS classes that are direct subclasses of the *conceptNautilus* class.

The use of the *Savoir.db* table enables to gradually reconstitute the concept hierarchy by specifying the specialisation links between the concepts. If the *Savoir.db* table describes a relation ES between Code_1 and Code_2, the translator generates a *SubclassOf* link between the concepts $\mathrm{Concept}_1$ and $\mathrm{Concept}_2$ (i.e. the RDFS classes generated by translation of Code_1 and Code_2): $\mathrm{Concept}_1$ becomes a subclass of $\mathrm{Concept}_2$ in the RDFS class hierarchy.

Similarly, the use of AT relations in the *Savoir.db* table enables to determine the domain and range of the RDF property "*located-on*", and to generate the RDF triples $Concept_1$ *located-on* $Concept_2$.

Lastly, the use of the *Lexique.db* table, which specifies the synonymous terms corresponding to the same code, enables to represent through RDFS labels the various synonymous terms naming the concept associated with this code.

The whole translation algorithm is detailed in [1]. It relies on an intermediate format of graphs and is decomposed into four steps:

(1) Translation of Nautilus concept and relation hierarchies into graphs.
(2) Translation of these graphs into RDF(S).
(3) Direct translation into RDF(S) of all the triples, except those containing ES specialisation relation or ME relation (*measured-with*).
(4) Specific processing of the triples including the ME relation.

Fig. 2 sums up the principle of translation from Nautilus DB to Nautilus ontology; Fig. 3 shows examples of extracts of the Nautilus tables, *Lexique.db* and *Savoir.db*, while Fig. 4 shows the extract of the ontology obtained after translation and Fig. 5 the corresponding RDF statements generated by the translator.

2.3.1. Discussion

The Nautilus ontology was aimed to be used in software of our industrial partner, compatible with the Nautilus database. Therefore in our translation we had to respect the initial Nautilus database as much as possible, in particular its initial relations: thus the ontology obtained after translation can have only the four relations initially available in Nautilus database since we could not add new relations during the translation process.

Moreover, if we consider that the choices for determining the contents and the structure of an ontology depend on the aimed application, Nautilus ontology has the concepts and relations needed for its application in the Episodus and Virtual Staff software. As described in Section 3.5, after the trans-

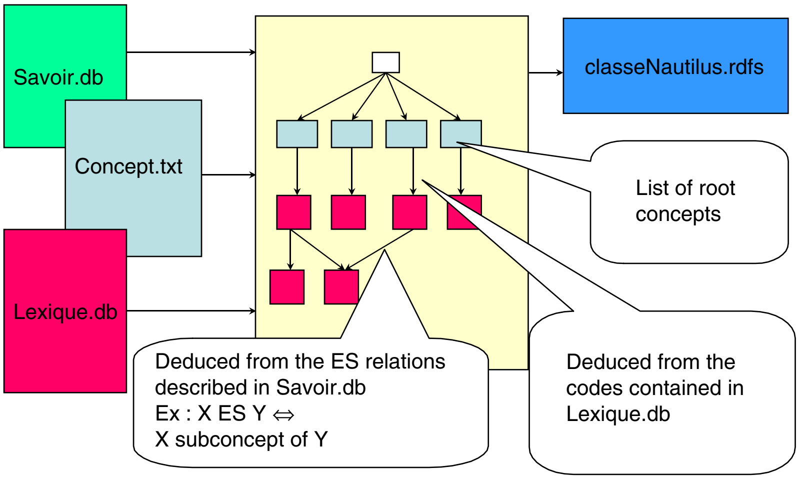

Fig. 2. Translation from Nautilus DB to the Nautilus ontology.

In Lexique.db Table of Nautilus DB:

LIBELLE	CODE
Bioprothese	**A** BIOP1
Heterogreffe valvulaire	**A** BIOP2
Bioprothese cardiaque	**A** OPRO1

In Savoir.db Table of Nautilus DB:

Qualified	Link	Qualifying
AOPRO	ES	ABIOP

Fig. 3. Extract of Nautilus DB tables.

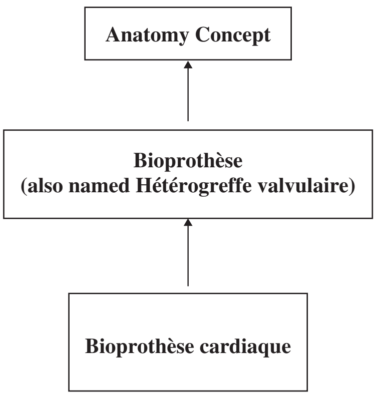

Fig. 4. Extract of the Nautilus ontology generated.

```
<lv:Concept rdf:ID="A">
    <rdfs:label xml:lang= "fr"> Anatomie </rdfs :label>
<rdfs:subClassOf rdf:resource="#ConceptNautilus"/>
</lv :Concept>

<lv:Concept rdf:ID="ABIOP">
    <rdfs:label xml:lang= "fr"> Bioprothese </rdfs :label>
    <rdfs:label xml:lang= "fr">Heterogreffe valvulaire</rdfs :label>
    <rdfs:subClassOf rdf:resource="#A"/>
</lv :Concept>

<lv:Concept rdf:ID="AOPRO">
    <rdfs:label xml:lang= "fr"> Bioprothese cardiaque </rdfs :label>
    <rdfs:subClassOf rdf:resource="# ABIOP "/>
</lv :Concept>
```

Fig. 5. Extract of the RDF(S) statements generated.

lation process, we added new relations needed for expressing the graphs handled in the Virtual Staff software.

When we analyse existing medical ontologies such as Galen [6,7], ON9 [10,11] or Snomed RT [12,13], they all offer high-level concepts and high-level relations, while the Nautilus ontology has no such top level. So, it could be interesting to connect Nautilus to such existing ontologies.

2.4. *Verification and validation of the ontology*

We distinguish, on the one hand, *automatic checking* through automated processing performed by the *translation program* and, on the other hand, *human validation* by the doctors or by the ontologist through *visualisation and navigation in the ontology*.

2.4.1. *Checking the coherence of the ontology*

The translation program carries out some coherence tests, which enabled us to detect various errors in the initial Nautilus database: (a) relations inducing cycles for the specialisation relation; (b) specialisation relations between a concept and itself; (c) redundancies.

Some of these errors were modelling errors and others corresponded to non-documented implementation tricks.

2.4.2. *Validation of the ontology with the CORESE search engine*

The interest of the translation of the Nautilus DB is to enable to check the ontology, once represented in RDF(S). Indeed, the user can browse the ontology using CORESE, our semantic search engine dedicated to RDF(S) [16–18]. Displaying the concepts via their labels instead of their code in the Nautilus DB makes the ontology more understandable for the doctors and its validation more friendly. CORESE interface could thus allow a team of doctors to visualise the ontology and to validate it by detecting the errors and by suggesting corrections.

By carrying out this validation within the Acacia team and with our Nautilus partner, we could detect several errors in the concept hierarchy. For example, in Fig. 6, we located several problems in the

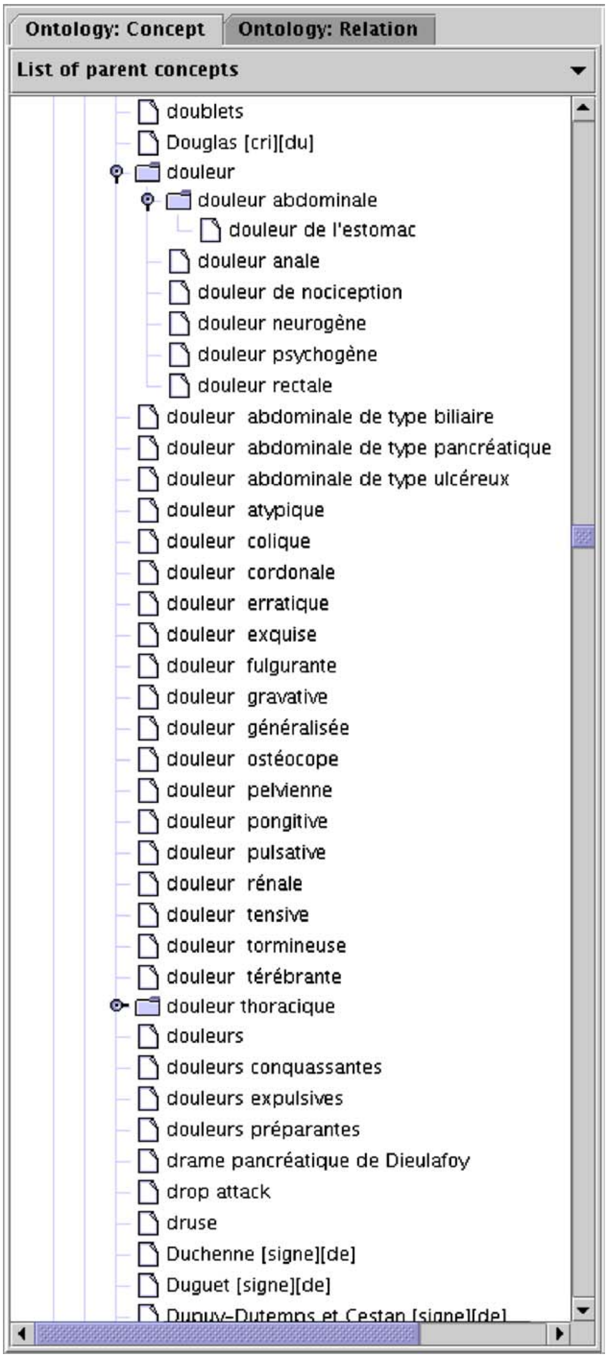

Fig. 6. Ill-formed hierarchy of "douleur" (pain).

conceptualisation of pain:

- *Problems of redundancy*: two concepts represent the concept of "*douleur*" (*pain*) while one single concept would be sufficient.
- *Structuring error*: the concepts *douleur abdominale de type biliaire* (*abdominal pain of biliary type*), *douleur abdominale de type pancréatique* (*abdominal pain of pancreatic type*), *douleur abdominale de type ulcéreux* (*abdominal pain of ulcerous type*) should have been direct subconcepts of the concept "*douleur abdominale*" (*abdominal pain*). It reveals that some ES relations had been forgotten in the *Savoir.db* table of the Nautilus DB.
- *Mixture of several points of view in modelling*: several viewpoints are mixed in the constitution of the hierarchy: some concepts characterise the pain by the part of the body concerned (e.g. *douleur abdominale* (*abdominal pain*), *douleur pelvienne* (*pelvic pain*), *douleur rénale* (*renal pain*), *douleur thoracique* (*thoracic pain*)) while others characterise it by its nature (e.g. *douleur exquise* (*exquisite pain*), *douleur fulgurante* (*fulgurating pain*)...).

Another advantage of CORESE search engine is to provide an environment to answer requests on the ontology—requests useful for the validation: since some concepts of the Nautilus ontology describe the anatomy, CORESE—which can find concepts via the terminology—can answer requests such as "*Which are the concepts having in their labels the word* '*abdominal*'?"

According to our partner, apart from the ontology validation, the queries of a doctor will rather relate to the ontology-based annotations on some medical documents or to the instances: e.g. "*Who are the patients whose medical record is annotated by a surgery for curing a pathology located on stomach?*" The inferences of CORESE based on the ontology Nautilus will enable to find the record of a patient who was operated for a stomach cancer.

2.5. Extensions of the Nautilus ontology

The Nautilus DB is used in the Episodus software for creating and editing the patient medical record. The concepts allowing to describe this patient record are thus useful when reasoning for diagnosis or for choice of the treatment. In the context of a tool aimed at supporting a health care network, it seems interesting to extend the Nautilus ontology by concepts enabling to describe a health care network and its actors, etc. For building these extensions, we rested on the corpus-based knowledge acquisition methodology [19] for semi-automatic acquisition relying on linguistic tools while being controlled by the expert and the knowledge engineer.

We thus constituted a corpus of documents on health networks: documents used for the preparation of the project "*Ligne de vie*", documents found on the Web about health care networks. Because of the disparate nature of these documents (scientific articles, vulgarisation articles, etc., and either in French or in English), the discourse structuring and the linguistic quality of the corpus were heterogeneous. Terminological extraction with a linguistic tool, Nomino [20], allowed to extract noun syntagms constituting candidate terms that we analysed, filtered, structured or gathered together. We thus built concept hierarchies on health networks, on health centres and on the patient record. Fig. 7 shows the *Network* hierarchy obtained from these candidate terms. Relying on the only terms attested by the corpus, it thus depends on the quality and completeness degree of the corpus. But, using the same method as [21], this hierarchy is easy

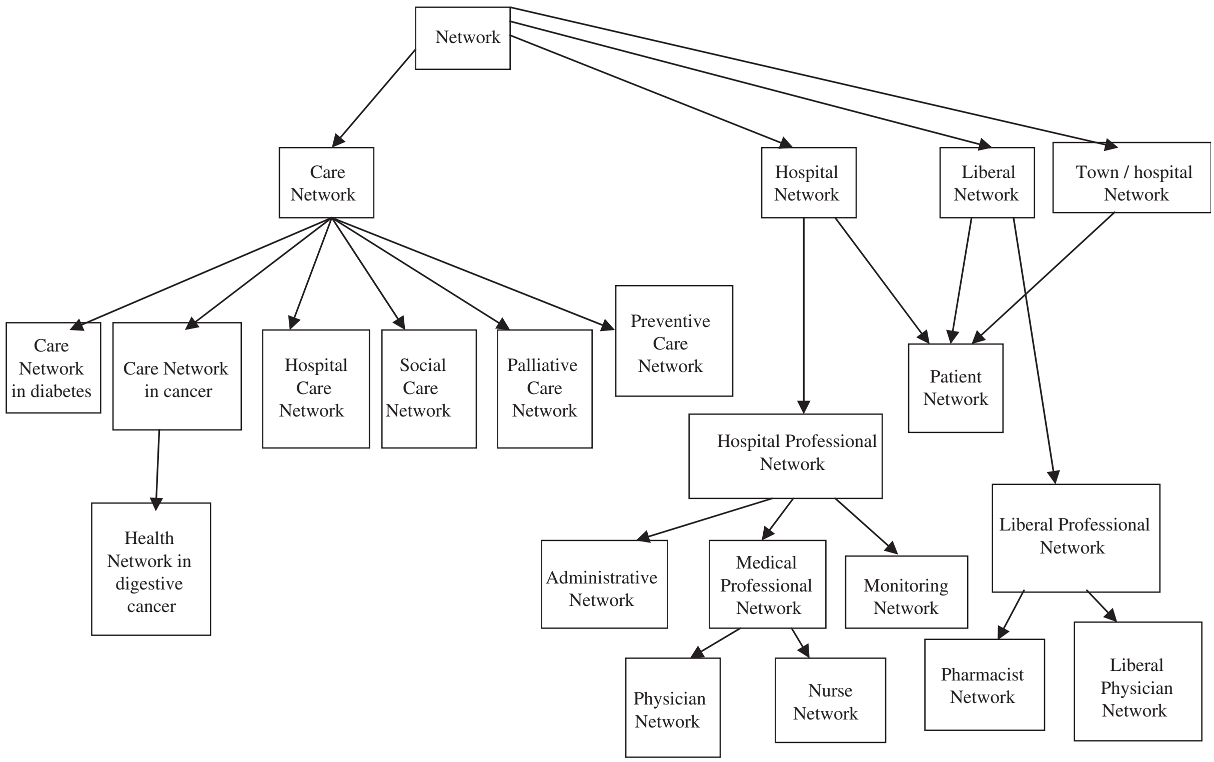

Fig. 7. Network hierarchy built from the list of candidate terms.

to enrich automatically using heuristic rules such as:

R1: If the term X in the list of candidate terms starts with the expression ≪Care Network in≫
Then suggest to create the concept X as a subconcept of the concept ≪Care Network≫.
R2: If Y is a subconcept of Pathology in the ontology,
Then suggest to create the concept ≪Care Network in Y≫ as subconcept of the concept ≪Care Network≫.

Since the term "*Care Network in Cancer*" was extracted by the linguistic tool, the rule $R1$ enables to create the concept "*Care Network in Cancer*" as a subconcept of "*Care Network*".

In the same way, the rule $R2$ would enable to create the concept "*Care Network in Cystic Fibrosis*" as subconcept of "*Care Network*", since the Nautilus ontology includes the concept "*Cystic Fibrosis*" as a subconcept of *Pathology*.

Lastly, to describe the hierarchy of the health actors, we integrated a part of the Canadian National Classification of the Professions in the Nautilus ontology by relying on the concepts common to this classification and to Nautilus.

As a conclusion, the extended Nautilus ontology was obtained from heterogeneous information sources (a database, a textual corpus, a classification) and by several ontologists, with a risk of non-homogeneity in its structuring, and a strong need of a validation by doctors. We will now show how this ontology is used in the Virtual Staff, software aimed at supporting collaborative work in the health network.

3. Representation and use of the Virtual Staff

3.1. Objectives of the Virtual Staff

In the hospital, the unity of location and time allows the doctors to meet as a staff in order to discuss about the decisions to take. In a health care network, the Virtual Staff aims to be a collaborative work supporting tool, allowing the real time update and history of therapeutic decisions. As an electronic board where each one can note information readable by the other members of the team, it constitutes a discussion support that may be synchronous (if the participants take part in the discussion at the same time or in the same place) or asynchronous (if each one accesses it at the moment appropriate to him/her). Starting from the patient's health problems, the members of the team will formulate diagnostic hypotheses and proposals for a treatment. Via this Virtual Staff, the health care team will connect the various elements of the patient record useful for the discussion, and thus will converge in an asynchronous way towards the definition of new health problems and of new therapeutic actions. The formulation of diagnostic hypotheses is a priori reserved to the medical actors, whereas the discussion on the treatment could sometimes imply non-medical professionals (for example, a welfare officer could emit arguments against the choice of a heavy treatment incompatible with housing conditions of the patient).

3.2. Weed's SOAP model

In Virtual Staff, the dependencies between the various diagnostic and therapeutic hypotheses can be represented through a graph using the concepts defined in the Nautilus ontology. The doctor will reason by linking the health problems to the symptoms, the clinical signs and the observations in order to propose health care procedures.

The Virtual Staff can thus rely on the SOAP model (subjective, objective, assessment, plan) used by the medical community [22]. In this model:

- the S nodes describe current symptoms and clinical signs of the patient,
- the O nodes describe analyses or observations of the physician,
- the A nodes correspond to the diseases or health problems of the patient,
- and the P nodes correspond to the procedures or action plans set up in order to solve the health problems.

This SOAP model is used in the medical community to structure a patient record. Therefore, its use to structure the doctor's reasoning—that relies on the same concepts—seems natural.

When a patient consults his/her doctor for new symptoms, the physician will create an instance of Virtual Staff. The system will then initialise a SOAP graph with all currently open pathologies and prescriptions for this patient (see Fig. 8): the initial A and P nodes automatically added in the graph are the pathologies and the care procedures already existing and open in the patient's life line. The doctor can then reason

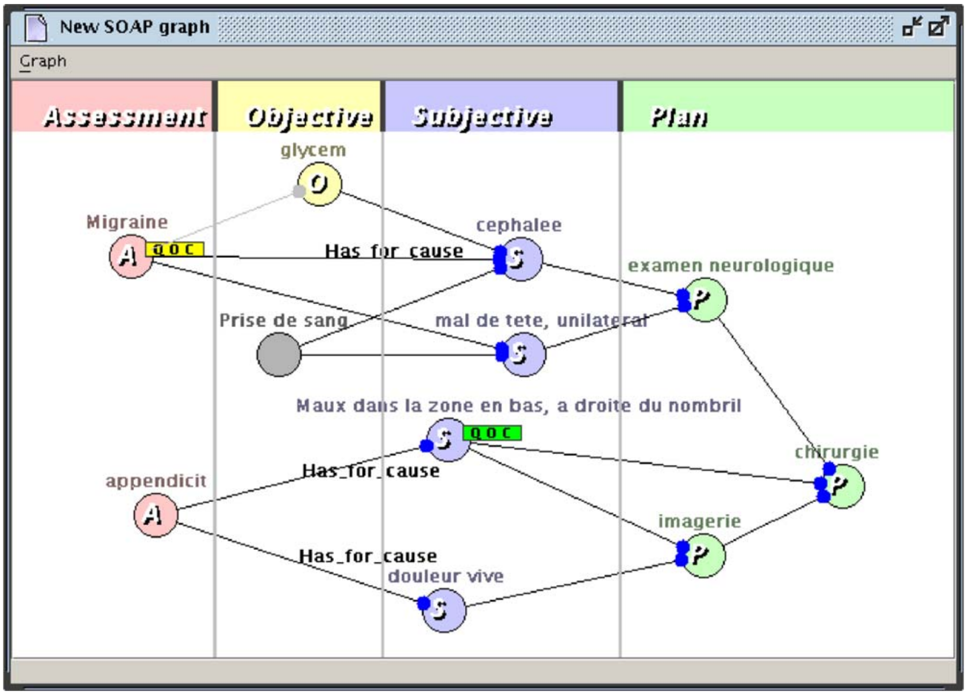

Fig. 8. SOAP model representing a medical record.

in order to add, in case of need, new A nodes (i.e. new pathologies diagnosed) or new P nodes (i.e. new action plans) in order to diagnose and treat the new health problem of the patient.

3.3. QOC model (question–options–criteria)

Sometimes, the physician may need to visualise all the possible solutions and the arguments in their favour or against them. The QOC model (question–options–criteria) [23], used by CSCW community for support to decision-making or for design rationale in a project, can then be useful. In this model, a question Q corresponds to a problem to solve. To solve the question Q, several options are thought out, with, for each option, the criteria in its favour and the criteria against it: each option is thus connected positively or negatively to criteria. The QOC graph is reduced to a tree if no criterion is linked to several options.

Two types of questions are possible for the Virtual Staff:

- *Diagnosis of a pathology* (i.e. find the right A in the SOAP model): Which pathology explains the clinical signs of the patient?
- *Search of a prescription* (i.e. find the right P in the SOAP model): Which action plan will enable to treat the diagnosed pathology?

In the Virtual Staff, among the criteria to be satisfied, there are the patient's symptoms and the doctor's observations: for a question about the patient's pathology, each possible option will be linked by a positive

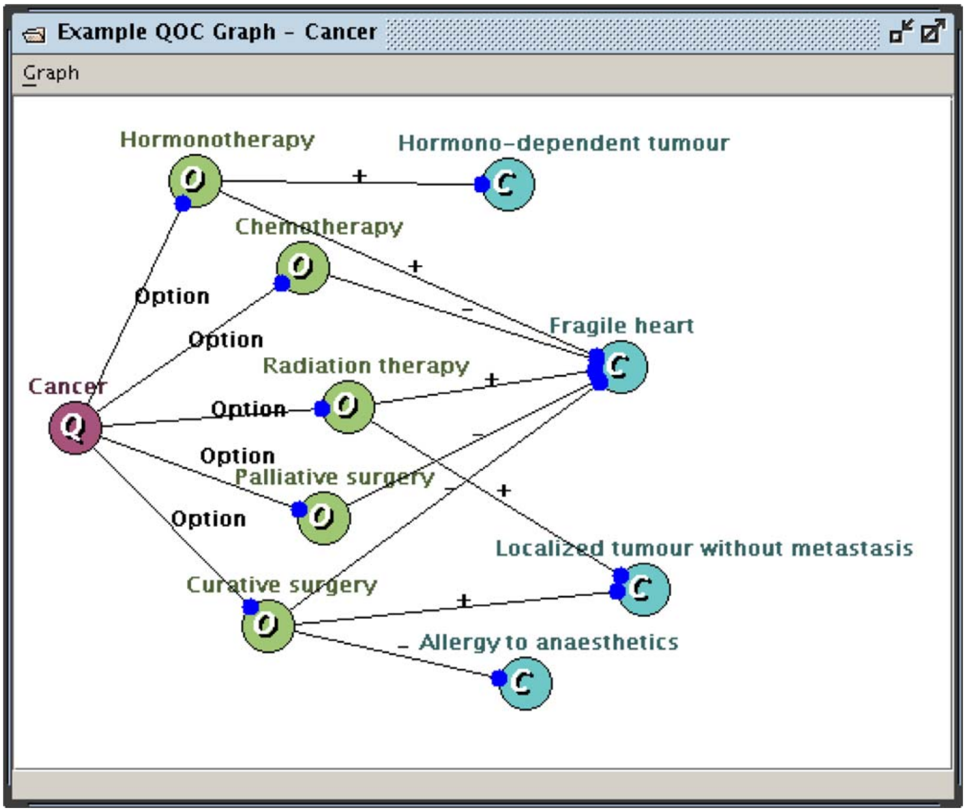

Fig. 9. QOC graph for choice of a therapeutic protocol a cancer.

influence link to the symptoms and observations compatible with this option, and by a negative link to the symptoms or observations rather incompatible with this pathology. The criteria will thus consist of S or O nodes of the SOAP model but they may also sometimes correspond to A or P nodes, if some diseases are incompatible or if some care procedures are exclusive.

For support to decision on a treatment to cure the diagnosed pathology, the options will be the various possible treatments, each one connected by a positive link to the criteria encouraging to choose it and by a negative link to the criteria inciting to reject it.

For example, Fig. 9 shows a QOC graph for choosing between an hormonotherapy, a chemotherapy, a radiotherapy, a palliative surgery and an eradicating surgery in order to treat a cancerous tumour, taking into account several criteria such as some characteristics of the tumour (e.g. hormonally dependent tumour, localised tumour without metastasis) or of the patient (e.g. fragile heart, allergy to anaesthesia).

3.4. Architecture of the Virtual Staff

Fig. 10 shows the whole architecture of the Virtual Staff, with its connections with other components, in particular the Nautilus ontology, the Episodus software and the CORESE semantic search engine.

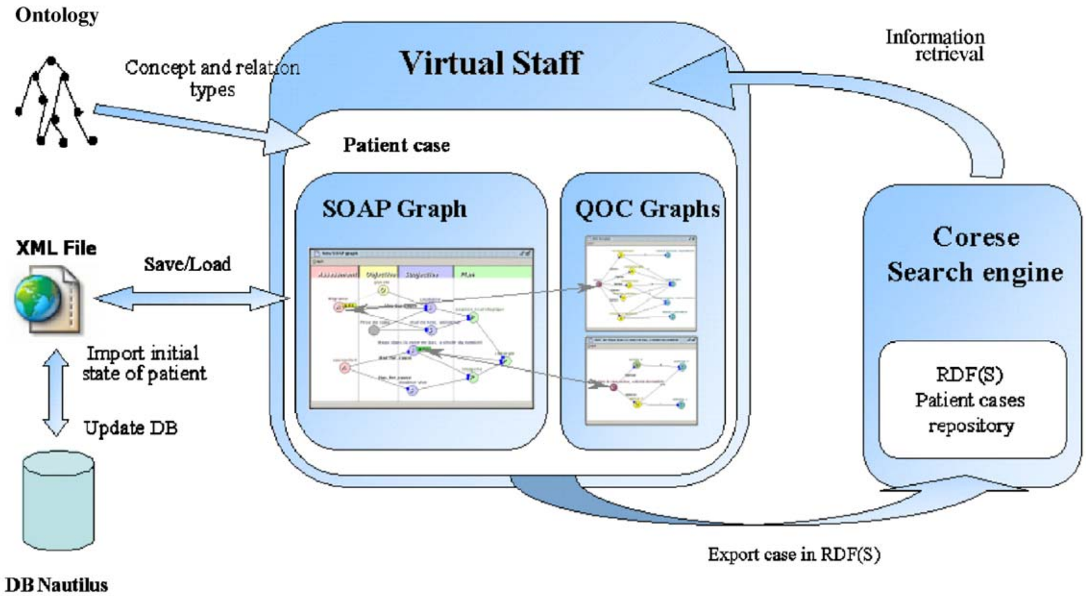

Fig. 10. Architecture of the Virtual Staff and its associated components.

3.5. Knowledge representation for the Virtual Staff

In the Virtual Staff, we thus combine both models: SOAP to visualise the medical record and QOC in phase of decision to choose between pathologies or between action plans.

These graphs can be represented through conceptual graphs [24], built by using the concepts and relations of the Nautilus ontology. Due to the correspondence between conceptual graphs and RDF(S) language [16–18], SOAP or QOC graphs can also be represented in RDF(S). Using the Nautilus ontology, the system can propose a list of possible concepts to help the user to build the SOAP and QOC graphs (see Fig. 11):

- the S nodes will correspond to instances of selected concepts among the subconcepts of *Symptom*,
- the O nodes will be chosen among subconcepts of *Laboratory-Test*, *Pathogenous-Agent* or *Foreign-Body*,
- the A nodes among those of *Pathology*, *Malformation* or *Psychological-Problem*,
- and the P nodes among those of *Treatment* or *Diagnostic-Therapeutical-Gesture*.

The arcs between the nodes will correspond to relations among concepts:

- *Symptom has_for_cause Pathology*;
- *Pathology has_for_consequence Symptom*;
  Malformation has_for_consequence Symptom;
  Psychological-Problem has_for_consequence Symptom;

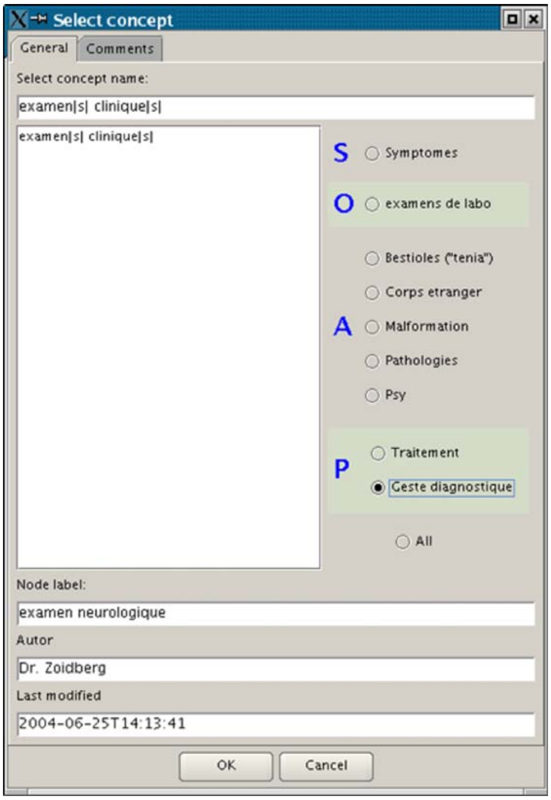

Fig. 11. Support of the Nautilus ontology for building a SOAP graph.

- *Pathology confirmed_by Laboratory-Test*;
- *Pathology treated_by Treatment*;
  Symptom treated_by Treatment;
  Malformation treated_by Treatment;
  Psychological-Problem treated_by Treatment.

In the same way, for QOC graphs aimed at determining the right pathology, the Options will be selected among the subconcepts of *Pathology*, and the Criteria among those of *Symptom*, *Laboratory-Test*, *Pathology* or *Treatment*.

To determine the right treatment, the Options will be selected among the subconcepts of *Treatment*, and the Criteria among those of *Symptom*, *Laboratory-Test*, *Pathology* or *Treatment*.

The arcs between the nodes of a QOC tree can be interpreted by "Question *option* Option" or by "Option *has-positive-criterion* Criterion" or by "Option *has-negative-criterion* Criterion".

We thus extended the Nautilus ontology with relations useful for representing SOAP graphs (*has_for_cause, has_for_consequence*, *confirmed_by*, *treated_by)* or QOC graphs (*option*, *has-positive-criterion*, *has-negative-criterion)*.

To express certainty degrees on a diagnosis of disease or priority degrees between the possible treatments, the physician can also indicate qualitative weights (such as ++ or −−) on some arcs between the nodes of SOAP or QOC graphs. An arc linking an option to a criterion and weighted by "++" (resp. −−) will mean that this criterion is a very strong argument in favour of (resp. against) this option. The quantitative weights studied thoroughly in [1] appeared too complex to be used by the health actors that were the intended end-users of the Virtual Staff. Therefore, we preferred a simple solution based on qualitative weights.

Let us note that the Virtual Staff is not at all an expert system. It only allows the care team to *visualise the reasoning and the decision-making process*; the reasoning is carried out by the members of the network and the inferences based on the ontology enable the system to filter the choices offered to the user. The role of the QOC graph can be compared to a guide of best practices [25]. But the QOC graph relies on the specific data of the concerned patient and not on generic data. To the nodes or arcs of the SOAP and QOC graphs, one can associate medical documents such as guides of best practices: an argument on the choice of a given treatment can be connected by a hypertext link to a guide describing the criteria of selection of this treatment.

Remarks.

- If one integrates in the Nautilus ontology the list of symptoms associated to a pathology, a query to CORESE about all the pathologies possible for a given symptom of the patient could be useful for helping in the diagnosis of the disease while building the QOC graph.
- By the same way, if one integrates in Nautilus the possible treatments for each pathology, as described in a guide of best practices, CORESE will be able to suggest the list of the possible prescriptions for the QOC graph.
- If the Nautilus ontology is extended by the drugs and their side-effects or contra-indications, CORESE can suggest that a given drug may be the cause of a given symptom or indicate a criterion against the choice of a given drug as a treatment of a pathology.

3.6. Interactions with the Virtual Staff

The first version of the Virtual Staff was evaluated by our industrial partner. It led to several improvements of the Virtual Staff interface. In particular, the new interface (cf. Fig. 12) helps the user to really follow the SOAP and QOC model: the Virtual Staff window is now divided into several zones, each dedicated to the kind of nodes that can be created according to the SOAP model or to the QOC model. It thus prevents the user from creating any graph, which would not respect the SOAP model or the QOC model.

At the opening of a Virtual Staff session, the list of the different participants is stored, with their names and their role in the network (i.e. this role corresponds to a subconcept of *Health-actor* in the Nautilus ontology). When one of these participants makes a contribution (i.e. adds a node or an arc in the SOAP graph or in a QOC graph), the identity of this contributor is stored and the nodes or arcs such added are visualised in the colour associated to this contributor. It thus enables to recognise directly, from the colour of the nodes or of the arcs, which physician or which member of the network argued about a given possible pathology or suggested a given treatment (see Fig. 12).

After a Virtual Staff session, the SOAP and QOC graphs created are saved and stored in an XML file, so that they can be exchanged with the Episodus software of our industrial partner. More precisely, the system saves the incremental changes starting from the Virtual Staff state at the opening of the session.

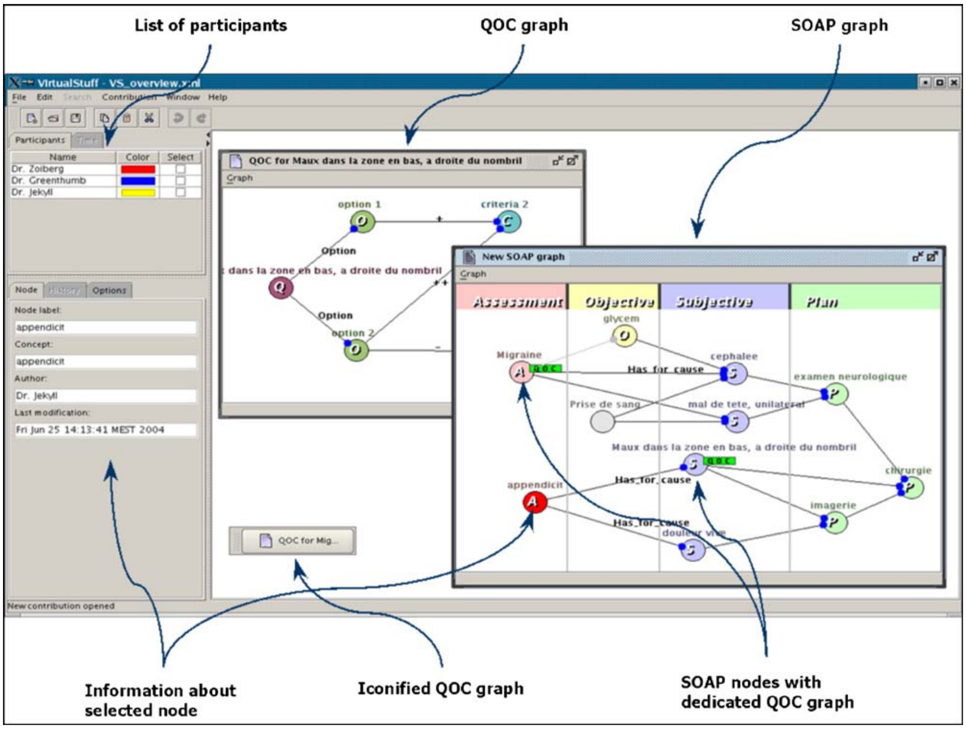

Fig. 12. Graphical User Interface of the Virtual Staff.

Moreover, the SOAP and QOC graphs are also saved in RDF(S) files so that the base of patient cases can be queried through CORESE for a reasoning similar to case-based reasoning: e.g. it would be possible to query about all the past patients for which a QOC graph had supported a decision between a chemotherapy and a hormonotherapy, and to ask which (positive or negative) arguments had been exchanged for leading to this decision.

3.7. Example

Let us detail as example the discussion summarised in Fig. 9.

Suppose that a localised tumour without metastasis has been discovered in the breast of a patient, whose life line dossier indicates that she is allergic to anaesthesia and has a fragile heart. The Virtual Staff can be used either in a synchronous way with all the different specialists gathered together at the same location, or in an asynchronous way with these different specialists disperse in various locations.

The surgeon stresses that a possible solution is *surgery* since the patient has a localised-tumour-without-metastasis (positive criterion in favour of surgery) but that her fragile heart and her allergy-to-anaesthesia are negative criteria against surgery.

The radiotherapist emphasises that another solution is *radiotherapy* for which the localised-tumour-without-metastasis is also a positive criterion. He notices that there is no negative argument against radiotherapy.

For the chemotherapist, two other solutions can also be thought out:

- *Hormonotherapy* since the results of the laboratory test prove that the patient's tumour is hormonally dependent, but her fragile heart could be considered as a negative criterion against hormonotherapy.
- *Chemotherapy* but the patient's fragile heart is also a negative criterion against chemotherapy.

Finally, taking into account all these possible options and their positive and negative arguments, the physicians can decide collectively to choose a *radiotherapy* as treatment for this patient's breast cancer.

This simplified example shows how the Virtual Staff may support or keep track of the discussions of several specialists about the therapeutic decisions about a patient's case.

4. Conclusions

4.1. Summary of the work presented

In this paper, we presented a translator of an ontology coded in an internal format of a database towards the RDF(S) language, in order to check and validate the ontology by visualising it via the CORESE engine, in a human-understandable form. The translation algorithm depends on the internal format of the database but the generic idea of building an ontology by decoding a database knowing its principle of coding, and then representing it in a standard formalism, is interesting for companies having DBs from which they wish to reconstitute an ontology.

Moreover, this ontology was extended by using a knowledge acquisition method based on corpus analysis through a linguistic tool, as in [19] and by integrating an existing classification. This constitution of an ontology from heterogeneous sources (database, textual corpus, classification) could be compared with approaches such as the ONIONS method [10,26].

We also specified and developed in JAVA a Virtual Staff, with:

- SOAP graphs describing the links between diagnostic and therapeutic hypotheses, symptoms and observations,
- and QOC graphs for support to decision-making.

The nodes of both kinds of graphs are typed by the concepts of the ontology. The arcs are typed by relations added to the Nautilus ontology for enabling to represent SOAP and QOC graphs.

Such a combination of these SOAP and QOC models with an ontology is original and illustrates the interest of an ontology to help the user to visualise a reasoning or a decision-making process. As IBIS method is close to QOC method, we can compare the Virtual Staff to gIBIS, which already relies on graph visualisation, and offers argumentation on decisions [27]. But our originality is to combine both SOAP graphs and QOC graphs and to rely on a medical ontology for building and handling these graphs. Another cooperative tool for a health care network is WebOnColl, a Web-based medical collaborative environment relying on user profiles and virtual workspaces [28].

The ontology and the Virtual Staff were validated by our Nautilus partner. As noticed earlier, the validation through CORESE enabled us to detect some errors in the initial ontology.

Notice that, even though the Virtual Staff was implemented with the Nautilus ontology (for collaboration reasons), it would be possible to adapt the Virtual Staff to another medical ontology such as UMLS meta-thesaurus.

4.2. Discussion

From *organisational viewpoint*, the organisation constituted by a health care network can be considered as a virtual enterprise, with a rather informal structure, and its members constitute a community gathered by a common objective (i.e. offer the best health care and follow-up for the patients), each member having also more specific objectives due to his/her profession (e.g. doctor vs nurse vs social worker) and to his/medical specialty. The kind of cooperation in this organisation may also depend on the kind of network: some networks are dedicated to a heavy pathology (e.g. diabetes, oncology, etc.), and gather members from different professions and different specialties, while other networks rather gather the same kind of health centres or professionals (e.g. hospital professional network, liberal practitioner network, pharmacist network, nurse network, etc.) and others will be dedicated to a type of patient (new born, old people). The kind of interactions will of course differ according to the type of network. Our approach, based on support to cooperative reasoning, seems useful for a health care network dedicated to a heavy pathology.

From a *cognitive viewpoint*, the members of the network build themselves a mental representation of the patient's case. The graphs handled in the Virtual Staff aim at enabling a participant to visualise partly this representation, to share it with other participants and to make it evolve through cooperation with other participants. SOAP model seems relevant for medical reasoning and QOC for representing diagnosis and therapeutic decisions since QOC model is known as useful for design rationale of a design project: patient's health care and follow-up can be considered as a therapeutic project in which some members of the health care network will take part.

From *technological viewpoint*, in addition to SOAP and QOC graphs, we mainly rely on semantic Web technologies:

- ontologies for representing the concepts shared by the network members,
- RDF(S) for representation of the ontology and of the annotations on the patient's record,
- CORESE semantic search engine for querying the ontology and the RDF annotations (in particular on one or several Virtual Staffs concerning one or several patients),
- XML for interchange between the Virtual Staff and the Episodus software describing the life line of the patient.

Our approach can be seen as tackling knowledge management from several viewpoints:

- *Organisational memory:* we build a memory of a specific community constituted by the health care network members: these members may be individual or organisations (e.g. hospital, health centre). More precisely, we build a memory of health projects about patients: each life line of a patient is considered as the trace of a medical project, with events, phases, actors playing a role in this project. Tractability of the project decision rationale is tackled by QOC graphs.
- *Support to cooperative work:* our approach follows the suggestions of [29] for taking inspiration of CSCW for heath care support.
- *Case-based reasoning:* even though we do not use classic case-based reasoning techniques [30], querying through the semantic search engine CORESE on past patient cases and past Virtual Staff sessions in order to have suggestions for a new patient case aims at the same objective as case-based reasoning.

4.3. Further work

As further work, we plan to extend the Nautilus ontology as suggested in Section 3.4, provided that such extensions are accepted by the physicians (cf. according to our industrial partner, it seems that some physicians do not appreciate to have "intelligent" suggestions from a system and prefer to choose and decide everything themselves).

Last, we will proceed to its evaluation by physicians taking part in an actual network (probably in diabetes): it will probably lead to other improvements of the Virtual Staff interfaces so as to enable friendly interactions between the end-user (a medical professional) and the system, by taking into account the context of this user.

Acknowledgements

We thank very much Philippe Ameline, the author of the Nautilus database who launched the project "*Ligne de Vie*", who had the initial idea of the Virtual Staff and who interacted very fruitfully with us for specification and validation of our work on the Nautilus ontology and on the Virtual Staff. We also thank Pr P. Degoulet, Dr. Dericco, Mrs. Labelle and Mrs. D. Sauquet, for our fruitful discussions, and the "Ministère de la Jeunesse, de l'Education Nationale et de la Recherche" for funding the project "*Ligne de Vie*".

References

[1] D. Minier, F. Corby, R. Dieng-Kuntz, O. Corby, P.-H. Luong, L. Alamarguy, Rapport intermédiaire du contrat Ligne de Vie, Octobre, 2003
[2] R. Dieng-Kuntz, D. Minier, F. Corby, M. Ružicka, O. Corby, L. Alamarguy, P. Luong, Medical ontology and virtual staff for a health network, in: E. Motta, N. Shadboldt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web: Proceedings of the 14th International Conference, EKAW'2004, October 5–8, Whittlebury Hall, UK, Lecture Notes in Computer Science Publisher, vol. 3257, Springer, Heidelberg, 2004, pp. 187–202 (ISBN: 3-540-23340-7).
[3] M. Ružicka, R. Dieng-Kuntz, D. Minier, Virtual staff—software tool for cooperative work in a health care network, INRIA Research Report RR-5390, November, 2004.
[4] M. Ribière, R. Dieng-Kuntz, A viewpoint model for cooperative building of an ontology, in: U. Priss, D. Corbett, G. Angelova (Eds.), Conceptual Structures: Integration and Interfaces, Proceedings of the Tenth International Conference in Conceptual Structures (ICCS'2002), July 15–19, Lecture Notes in Computer Science, vol. 2393, Springer, Borovetz, Bulgaria, 2002, pp. 220–234.
[5] R. Dieng-Kuntz, N. Matta (Eds.), Knowledge Management and Organizational Memories, Kluwer Academic Publishers, Dordrecht, 2002.
[6] A. Rector, A. Gangemi, E. Galeazzi, A. Glowinski, A. Rossi-Mori, The Galen model schemata for anatomy: towards a re-usable application—independent model of medical concepts, in: P. Barahona, M. Veloso, J. Bryant (Eds.), Proceedings of Medical Informatics in Europe (MIE'94), 1994, pp. 229–233.
[7] A. Rector, J. Rogers, P. Pole, The Galen high level ontology, in: J. Bender, J.P. Christensen, J.R. Scherrer, P. Mcnair (Eds.), Proceedings of Medical Informatics in Europe MIE 96, IOS Press, Amsterdam, 1996, pp. 174–178.
[8] P. Zweigenbaum, B. Bachimont, J. Bouaud, J. Charlet, J.-F. Boisvieux, Issues in the structuring and acquisition of an ontology for medical language understanding, Methods Inf. Med. 34 (1/2) (1995) 15–24.
[9] P. Zweigenbaum, Consortium Menelas, Menelas: Coding and Information Retrieval from Natural Language Patient Discharge Summaries, Advances in Health Telematics, IOS Press, Amsterdam, 1995, pp. 82–89.
[10] A. Gangemi, D. Pisanelli, G. Steve, Ontology integration: experiences with medical terminologies, in: N. Guarino (Ed.), Formal Ontology in Information Systems, IOS Press, Amsterdam, 1998, pp. 163–178.

[11] D.M. Pisanelli, A. Gangemi, G. Steve, An ontological analysis of the UMLS metathesaurus, J. Amer. Medical Inform. Assoc. 5 (1998) 810–814.

[12] R.A. Côté, D.J. Rothwell, J.L. Palotay, R.S. Beckett, L. Brochu (Eds.), The Systematized Nomenclature of Human and Veterinary Medicine: Snomed International, College of American Pathologists, Northfield, IL, 1993.

[13] K. Spackman, K. Campbell, R. Côté, SNOMED RT: a reference terminology for health care, in: D.R. Masys (Ed.), The Emergence of Internetable Health Care: Systems that Really Work, Proceedings of the 1997 AMIA Annual Symposium, Nashville, TN, October 25–29, Hanley & Belfus, Philadelphia, PA, 1997, pp. 640–644.

[14] National Library of Medicine, UMLS Knowledge Source, 14th ed., Doc. National Institute of Health–National Library of Medicine, Bethesda, MD, USA, January, 2003

[15] O. Lassila, R.R. Swick, Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, < http://www.w3.org/tr/rec-rdf-syntax/>, 22 February 1999.

[16] O. Corby, R. Dieng, C. Hébert, A conceptual graph model for W3C resource description framework, in: B. Ganter, G.W. Mineau (Eds.), Conceptual Structures: Theory, Tools, and Applications, Proceedings of ICCS'2000, August 13–17, Lecture Notes in Artificial Intelligence, vol. 1867, Springer, Darmstadt, Germany, 2000, pp. 468–482.

[17] O. Corby, C. Faron-Zucker, Corese: a corporate semantic web engine, in: Proceedings of the International Workshop on Real World RDF and Semantic Web Applications in Conjunction with 11th International World Wide Web Conference, Hawaii, May 2002 < http://paul.rutgers.edu/~kashyap/workshop.html>.

[18] O. Corby, R. Dieng-Kuntz, C. Faron-Zucker, Querying the Semantic Web with the CORESE search engine, in: R. Lopez de Mantaras, L. Saitta (Eds.), Proceedings of the 16th European Conference on Artificial Intelligence (ECAI'2004), August 22–27, IOS Press, Valencia, 2004, pp. 705–709.

[19] N. Aussenac-Gilles, B. Biébow, S. Szulman, Revisiting ontology design: a methodology based on corpus analysis, in: R. Dieng, O. Corby (Eds.), Knowledge Engineering and Knowledge Management: Methods, Models and Tools, Lecture Notes in Artificial Intelligence, vol. 1937, Springer, Berlin, 2000, pp. 172–188.

[20] L. Dumas, A. Plante, P. Plante, ALN: Analyseur Linguistique de ALN, version 1.0, ATO, UQAM, 1997.

[21] J. Golebiowska, R. Dieng-Kuntz, O. Corby, D. Mousseau, SAMOVAR: using ontologies and text-mining for building an automobile project memory, in: R. Dieng-Kuntz, N. Matta (Eds.), Knowledge Management and Organizational Memories, Kluwer Academic Publishers, Boston, 2002, pp. 89–102.

[22] L.D. Weed, The problem oriented record as a basic tool in medical education, patient care and clinical research, Ann. Clin. Res. 3 (3) (1971) 131–134.

[23] A. MacLean, R. Young, V. Bellotti, T. Moran, Questions, options, and criteria: elements of a design rationale for user interfaces, Int. J. Human-Computer-Interact. 6 (3/4) (1991) 201–250.

[24] J.F. Sowa, Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, Reading, MA, 1984.

[25] G. Georg, B. Séroussi, J. Bouaud, Dérivation d'une base de connaissances à partir d'une instance GEM d'un guide de bonnes pratiques médicales textuel, in: R. Dieng-Kuntz (Ed.), Actes des 14èmes Journées Francophones sur l'Ingénierie des Connaissances (IC'2003), Presses Universitaires de Grenoble, Laval, 2003.

[26] A. Gangemi, D. Pisanelli, G. Steve, An overview of the ONIONS project: applying ontologies to the integration of medical terminologies, Data Knowl. Eng. 31 (2002).

[27] J. Conklin, M. Begeman, gIBIS: a hypertext tool for exploratory policy discussion, Commun. ACM (1998) 140–152.

[28] C.E. Chronaki, D.G. Katehakis, X.C. Zabulis, M. Tsiknakis, S.C. Orphanoudakis, WebOnCOLL: medical collaboration in regional health care networks, IEEE Trans. Technol. Biomedicine 1 (4) (1997) 257–269.

[29] W. Pratt, M. Reddy, D. McDonald, P. Tarczy-Hornoch, J. Gennari, Incorporating ideas from computer-supported cooperative work, J. Biomed. Inform. 37 (2) (2004) 128–137.

[30] M. Moussavi, A case-based approach to knowledge management, in: D.W. Aha (Ed.), Proceedings of the AAAI'99 Workshop on Exploring Synergies of Knowledge Management and Case-Based Reasoning, AAAI Press, Orlando, FL, 1999, Technical Report WS-99-10.

Rose Dieng-Kuntz graduated from Ecole Polytechnique and ENST, was Doctor in Computer Science at University of Paris-Sud, INRIA research director, ACACIA team leader at INRIA-Sophia-Antipolis, organizer of 24 conferences/workshops, member of 55 program committees, co-chair of EKAW'2000, area chair of ECAI'2002, conference chair of COOP'2004, FAIA series editor at IOS Press, and published 8 books (as author or editor) and 120 papers on Knowledge Engineering, Knowledge Management, Ontologies, Semantic Web, Conceptual Graphs and Multi-Agent Systems.

David Minier obtained his Graduate Diploma in Computer Science Engineering at ESIEA (Ecole Supérieure d'Informatique, d'Electronique et d'Automatique—France) and was in 2003 engineer on the Life Line project in the ACACIA team. Since June 2004, he has been an R&D engineer for a CRM editor.

Marek Růžička obtained his M.Sc. in Information and Knowledge Engineering from the University of Economics—Prague, in 2003. In 2004, he was expert engineer on the Life Line project in the ACACIA team. He is currently a Ph.D. student in Computer Science at the UEP, on user-driven transformation of textual documents. His main interest research areas include semantic analysis of textual knowledge, semantic web and ontology engineering. He is (co-)author of eight reviewed papers.

Frédéric Corby worked for 13 years in a child hospital laboratory and obtained his diploma of DUT in computer science. In 2003, he was trainee in the ACACIA team on the Life Line project. He prepares a professional diploma on network programming.

Olivier Corby, with a Ph.D. in Computer Science from the University of Nice-Sophia Antipolis, performs research on software development environment for knowledge modelling and on semantic web technologies for corporate memory management. Co-developer of SMECI, developer of NMS and WebCOKACE, he is the designer and main developer of CORESE semantic search engine and semantic web server. He published about 60 articles in journals, books, conferences, or workshops, is co-author of 2 books, was co-chair of EKAW'2000 and member of 25 program committees of conferences or workshops.

Laurent Alamarguy received his M.Sc. degree in Linguistics from Blaise Pascal University (Clermont-Ferrand, France) and is currently a Ph.D. student in the ACACIA team on computational linguistics. His research interests include corpus-based Natural Language Processing, lexical semantics in functional genomics, text annotation and biomedical ontology modelling.