

# Towards a Specialized Knowledge Base in Medicine: The BIKiNi (Biomedical Knowledge Network) Project

Valerie Bertaud <sup>ab</sup> DMD, Nicolas Garcelon <sup>b</sup>, Franck Marin <sup>a</sup>, Ihssene Belhadj <sup>a</sup>, Anita Burgun <sup>ab</sup> M.D Ph.D, Regis Duvauferrier <sup>ab</sup> MD.

<sup>a</sup> EA 3888, School of Medicine, IFR 140, University of Rennes 1, France

<sup>b</sup> CHU, Rennes, France

**OBJECTIVES:** The script concordance test (SCT)[1] was originally designed to assess clinical skills and measure the degree of concordance between the clinical reasoning of both examinees and a panel of experts[2]. We have investigated how a script concordance approach could be used to merge the knowledge of several experts in order to build a specialized knowledge base.

**METHODS: Defining the field of knowledge:** Shoulder symptom or complaint (L08 in the Internationale Classification of Primary Care:ICPC2).

**Eliciting knowledge from experts:** The experts (a physician, an orthopedist and a radiologist) knowledge was collected in natural language with two questionnaires (Tab. 1-2).

**Automating the generation of links:** A mapping to MeSH (NOMINDEX[3]) was performed in order to unify the terms used by the experts. The MeSH indexings were combined in order to identify similarities in the semantic network. The method is automated thanks to a Perl script. If the number of MeSH terms in common is higher than 90%, then these two nodes are considered to be similar. An example of convergent nodes is given in Fig. 1.

**Evaluation:** Starting from the information provided by experts, the knowledge base was indexed in parallel manually by one of the authors (RD). The evaluation consisted of comparing the convergent nodes in the two networks.

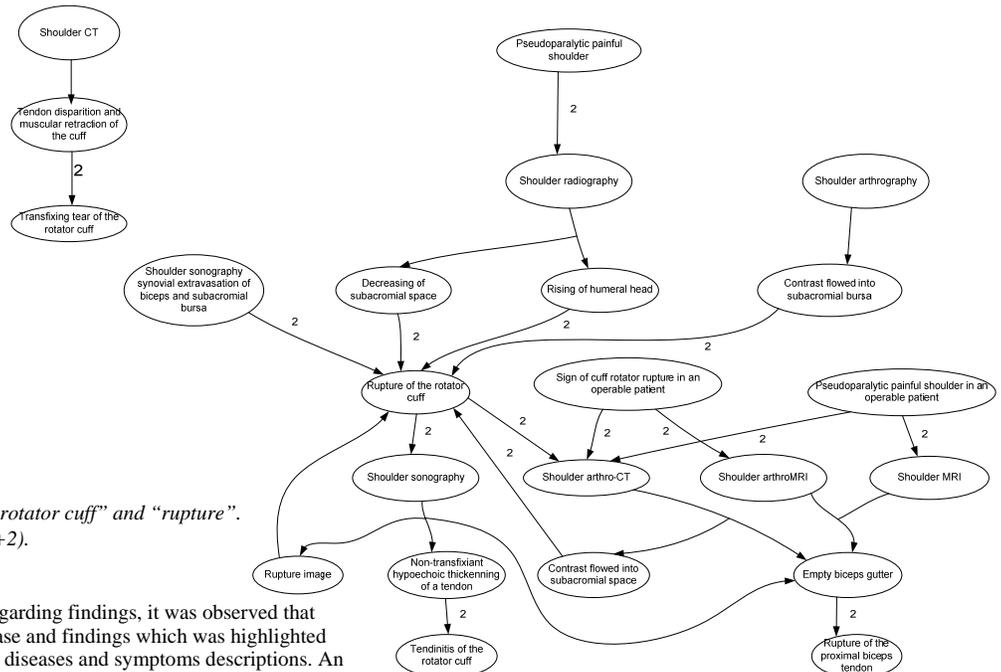
We have used the software Graphviz (Graph Visualization Software) with Dot language. Dot draws directed graphs[4].

**RESULTS:** The experts and the books provided 131 diagnosis items and 89 indication items; 20 different procedures; 59 Reasons for encounter; 55 different diagnoses and 72 different radiological findings. The computerized mapping of experts terms toward MeSH terms provided 231 different MeSH terms.

New links appeared in the convergence procedure. There are convergences for 30/59 RFE, for 15/20 diagnostic procedures and for 32/55 diagnosis. Six concepts are found both in findings and diagnosis. A previous diagnosis can be a finding for further interpretations.

The manually obtained convergences were compared to the automatically obtained convergence: for the diagnostic procedures 15 convergences (automatically) versus 15 (manually), for the RFE 19 convergences (manually) versus 32(automatically), for the diagnosis 22 (manually) versus 30 (automatically).

Fig.2: Retrieved graph for the keywords :” rotator cuff” and ”rupture”. Items with only very probable hypotheses (+2).



If you think of (diagnostic option)	and then you find (new information obtained in an imaging procedure)	The effect of this new information on the diagnostic hypothesis is
Diagnosis	Imaging finding	-2 -1 0 +1 +2
This new information turns the diagnostic option to be : -2 : almost eliminated ; 1 : less probable; 0 : no effect ; +1 more probable ; +2 : very probable		

Table 1 : The questionnaire of good practice assessment in prescription (indication item)

If you think of making (Diagnostic or interventional imaging procedure)	and then you find (new information obtained in questioning, in clinical examination or in another procedure)	The effect on the need for requiring this procedure is :
Imaging procedure	Patient's reason for encounter (RFE) or clinical finding	-2 -1 0 +1 +2
-2 : absolutely contraindicated ; -1 : not very useful or rather harmful; 0 : not relevant in this situation ; +1 : useful and desirable; +2 : essential		

Table 2 : The questionnaire for knowledge assessment of in radiological semiology (diagnosis item)

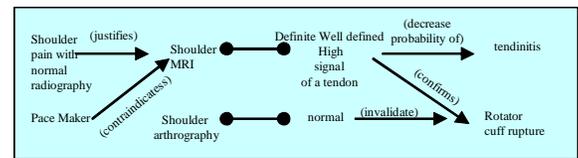


Figure 1: "Shoulder MRI", "Definite well defined high signal of a tendon" and "rotator cuff rupture" are convergent nodes

**DISCUSSION-CONCLUSION:** Regarding findings, it was observed that there was a continuum between concepts of disease and findings which was highlighted by the fact that terms were found at same time in diseases and symptoms descriptions. An important limitation of our system was the low performance of automatic indexing of terms. The mapping tool extracted only 63% of the relevant MeSH concepts. Even if 13% more concepts may be inferred automatically, the remaining 24% concepts had to be added manually.

The current study showed that the SCT approach initially developed for education can be adapted to represent knowledge for other purposes. The method allows to extract knowledge directly from the experts and indirectly from reference books.

Our system was tested in the domain of shoulder complaint and focused on medical imaging. The same method is applicable in other domains.

## REFERENCES:

- [1] Charlin B., Tardif J., Boshuizen H.P.A. – Script and medical diagnosis knowledge : theory and applications for clinical reasoning instruction and research. Acad Med 2000; 75 : 182-190
- [2] Charlin B., Roy I., Brailowsky C.A., Van der Vlieten C.P.M. – The script concordance test : a tool to assess the reflective clinician. Teach Learn Med Educ 2000; 12 : 189-195.
- [3] Happe A., Pouliguen B., Burgun A., Cuggia M, Le Beux P. Automatic concept extraction from spoken medical reports. Int J Med Inform. 2003 Jul;70(2-3):255-63.
- [4] M.N. Boulos, C. Russell, M. Smith. Web GIS in practice II: interactive SVG maps of diagnoses of sexually transmitted diseases by Primary Care Trust in London, 1997 - 2003. Int J Health Geogr. 2005 Jan 18; 4(1):4.