

Construction d'une Base de Connaissance Biomédicale Utilisant la Méthode des Scripts
Valérie Bertaud, D.M.D., Jeremy Lasbleiz, M.D., Régis Duvauferrier, M.D., Nicolas Garcelon, Anita Burgun, M.D., Ph.D.

EA 3888, School of Medicine, IFR 140, University of Rennes I, France

Le projet BiKiNi (Biomedical Knowledge Network) est destiné à gérer des connaissances biomédicales en permettant de recueillir les expertises, de les relier en réseaux de connaissance, de les représenter sous forme de graphes.

Nous avons étudié comment la théorie des scripts peut être employés pour fusionner la connaissance de plusieurs sources afin d'établir une base de connaissance unifiée. Nous proposons et évaluons des outils pour aider à construire et utiliser le réseau de connaissances.

MATERIELS ET METHODES:

Limitation du domaine: Nous décrivons ici une expérience réalisée dans le domaine des douleurs des dents et des gencives en urgence odontologique : D19 TEETH/GUM SYMPTOM/COMPLAINT dans l'ICP2 (International Classification of Primary Care 2nd edition).

Pour le recueil des expertises et l'actualisation de la base de connaissance, il utilise les modalités de construction des tests de concordances des scripts (TCS) (Fig.1).

Sources des connaissances: Les connaissances ont été collectées auprès d'experts (un médecin généraliste, un rhumatologue et un radiologue), dans la littérature [2] et dans des recommandations de bonne pratique [3], en **langage naturel**, suivant la logique des scripts.

La théorie des scripts [1] est issue de la psychologie cognitive. Elle postule que les praticiens, lorsqu'ils doivent résoudre un problème (poser un diagnostic, décider d'éventuels examens complémentaires ou choisir un traitement), utilisent des réseaux de connaissances préenregistrés (les scripts) afin de comprendre la situation clinique et de déterminer la ligne de conduite appropriée. Notre idée est d'utiliser ces scripts pour construire une base de connaissance qui pourra ensuite être utilisée pour aider à la décision clinique ou pour évaluer des étudiants.

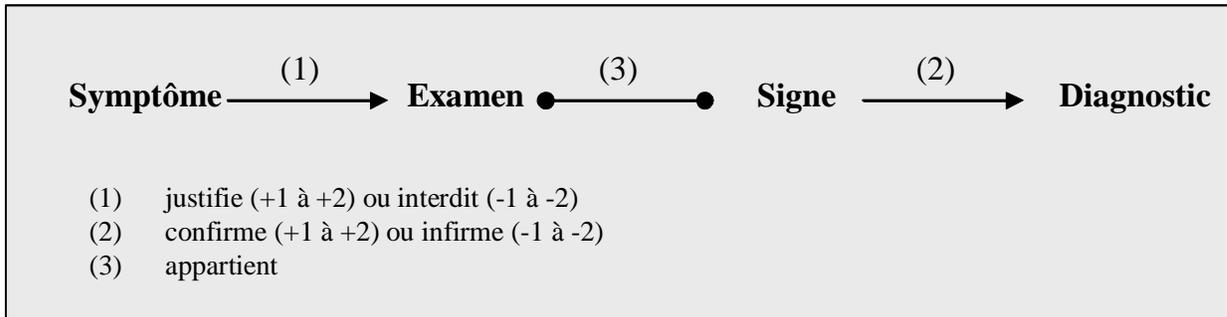
Des examens d'imagerie médicales sont envisagés en fonction de signes (Table 2). des hypothèses diagnostiques sont définies en fonction de différents signes (Table 1).

Si vous pensez à (option diagnostique)	et qu'alors vous trouvez (nouvelle information obtenue par un examen clinique ou par un test para clinique)	l'effet sur l'hypothèse diagnostique est le suivant
Diagnostic	Type d'examen : signe	-2 -1 0 +1 +2
-2 : l'hypothèse est pratiquement éliminée ; 1 : l'hypothèse devient moins probable ;0 : l'information n'a aucun effet sur l'hypothèse ; +1 : l'hypothèse devient plus probable ;+2 : il ne peut s'agir pratiquement que de cette hypothèse		

Fig 1 Exemple d'un questionnaire de recueil de connaissances sémiologiques

Si vous pensez faire (option d'examen d'imagerie diagnostique ou interventionnelle)	et qu'alors vous trouvez (nouvelle information obtenue par l'anamnèse, un examen clinique ou par un autre examen)	l'effet sur la nécessité de demander ce test est le suivant (entourer)
Examen d'imagerie	Symptômes	-2 -1 0 +1 +2

-2 : absolument contre-indiqué ; 1 : très peu utile ou plutôt néfaste ; 0 : non pertinent dans cette situation ; +1 : utile et souhaitable ; +2 : indispensable		



Indexation semi-automatique des termes employés : Pour relier entre elles les connaissances, il utilise l'indexation MESH. Cette étape permet de découvrir les expressions identiques au niveau sémantique alors quelles sont différentes au niveau lexical. On peut ainsi fusionner les expressions identiques et donc éviter les répétitions et identifier des convergences dans le réseau sémantique.

Un rapprochement avec le MeSH a été effectué afin de **normaliser les termes employés**. La méthode est partiellement automatisée. Le NOMINDEX du Laboratoire d'informatique médicale de Rennes. Un script en langage Perl enregistre les termes obtenus avec leur nomenclature source dans une base de données relationnelle (Oracle®). Tous les termes MeSH obtenus ont été vérifiés manuellement. Une **indexation manuelle** a été réalisée pour les termes non mappés automatiquement. Deux nœuds sont considérés comme similaires s'ils ont plus de 90% de termes MeSH en commun.

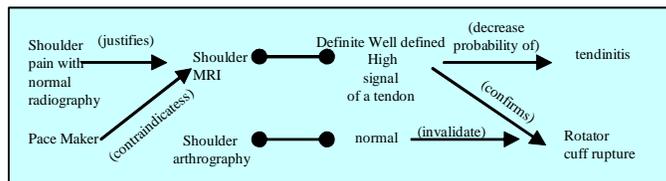


Figure 2: In this example, key words chosen by the user express « Well defined High signal of a tendon ». Paths which contain « Well Define High signal of a tendon » are displayed and diagnostic items and indication items which are linked to selected diagnostic items.

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

Visualisation graphique: Le logiciel Graphviz (Graph Visualization Software) [6] construit des graphes orientés interactifs.

Interrogation Medline: L'indexation par les mots MeSH permet de créer automatiquement des requêtes et d'interroger directement PubMed.

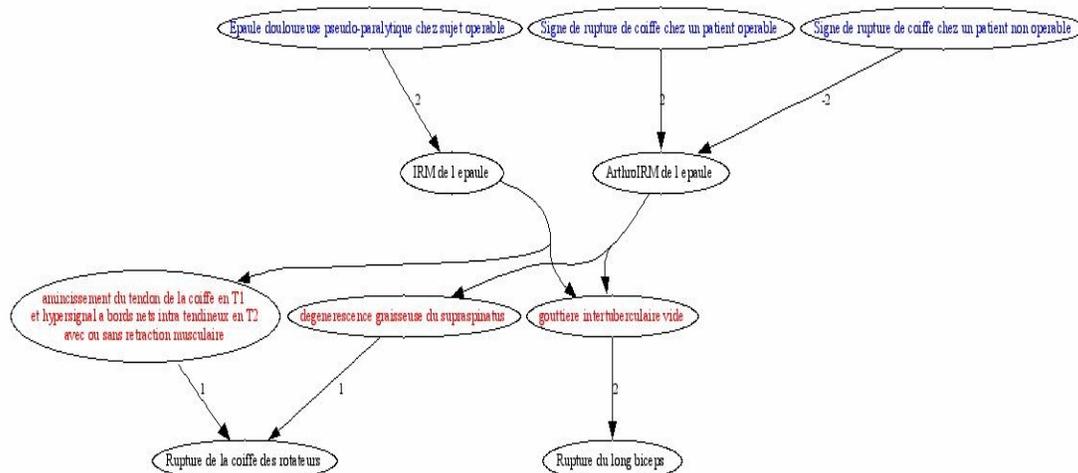


Fig 2 : Représentation graphique des expertises pour la rupture de la coiffe des rotateurs de l'épaule en IRM

RESULTATS:

The experts and the books provided 136 diagnostic items and 92 indication items. We extracted from books 49/136 diagnostic items and 42/92 indication items. There was 1 double and 1 triple within the indication items. Within the diagnostic items, we found 3 doubles and 1 triple. Those repetitions have been deleted. Once that step is done, there are 131 diagnostic items (in the step with radiological findings) and 89 indication items (in the step with clinical or para-clinical RFE) (Fig.4). The 89 indication items gathered 20 different procedures indicated in 59 RFE. The 131 diagnostic items gathered 55 different diagnoses which linked to 72 different radiological findings. In the indication items, there were 6 contraindicated diagnostic procedures linked to the RFE, 2 were harmful, 14 unuseful, 30 desirable, 36 indispensables.

In the diagnostic items, 21 diagnoses were invalidated, 7 with decreased probability, 17 with unchanged probability, 45 with increased probability, 41 were validated. We may add that radiological findings were described on 9 main types of diagnostic procedures.

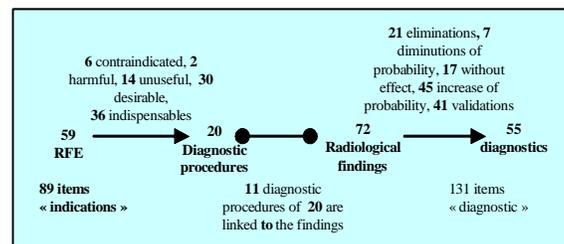


Fig 4 : Global analysis of relations in the semantic network.

The computerized mapping of experts terms toward MeSH terms provided 231 different MeSH terms. The distribution of these different terms according to the four types of nodes is the following:

- 123 in the RFE nodes
- 40 in the diagnostic procedure nodes
- 83 in the radiological findings nodes
- 79 in the diagnostic nodes

A total of 2553 MeSH terms were used to represent the terms present in the knowledge base. 63 % of the terms were obtained by automated indexing using NOMINDEX, 24% were added manually, 13% correspond to context specific terms (e.g., diagnosis).

Analyzing the knowledge base

New links appeared in the convergence procedure. There are convergences for 30/59 RFE, for 15/20 diagnostic procedures and for 32/55 diagnostics. The network also “folds back”: six concepts are found both in findings and diagnosis. In theory, there are 3 ways of folding back the network.

In practice, circularity mainly concerns “RFE”/ “Diagnosis” and “RFE”/ “Diagnosis”/ “Radiological findings”. The fold “RFE”/ “radiological finding” is essentially observed for normal examinations (Fig 5). **If we compare the convergences obtained manually versus the convergence obtained automatically, we get for the**

diagnostic procedures 15 convergences versus 15/20, for the RFE 19 convergences versus 32/55, for the diagnosis 22 versus 30/59.

Discussion

Nous avons montré que les scripts, initialement développés pour l'enseignement, peuvent être utilisés pour modéliser la connaissance dans un objectif différent.

Les principaux avantages de l'approche par la méthode des scripts et de la visualisation graphique sont :

- Elle permet de représenter les règles et les concepts.
- Etant donné qu'elle respecte le fonctionnement du raisonnement médical (raisonnement hypothético-déductif), elle permet de représenter la connaissance provenant d'experts, de livres, de recommandations de bonne pratique.
- Elle permet d'extraire des connaissances concernant les différents types de raisonnements médicaux : le diagnostic, la prescription, le traitement.
- La mise en correspondance des expressions libres avec le vocabulaire normalisé (MeSH) fournit une représentation unifiée des expressions et des mots utilisés. Par conséquent, il permet d'identifier des convergences au sein du réseau.
- Il est possible de créer une visualisation graphique qui répond aux besoins des utilisateurs.
- L'indexation des règles à l'aide des termes MeSH permet l'interopérabilité de la base de connaissance avec la littérature scientifique (Medline).
- La mise en correspondance avec le vocabulaire MeSH et l'interface graphique aident l'utilisateur à interroger rapidement Medline.

An important limitation of our system is the low performance of automatic indexing of terms. The mapping tool that we have used to perform that task extracts only 63% of the MeSH concepts that were qualified relevant. Even if 13% more concepts may be inferred automatically, the remaining 24% concepts must be added manually. However, where this mapping tool is insufficiently powerful for items written in French, the situation may be different for English. The use of the MetaMap (6) program in English combined with the Restrict to MeSH algorithm (7) of the NLM would certainly be more powerful.

La méthode peut être appliquée à d'autres domaines médicaux. Nous l'avons d'ailleurs également testée pour le diagnostic dans le domaine des urgences odontologiques. De plus, nous envisageons d'utiliser cette méthode pour construire, utiliser, partager et maintenir des guides de bonne pratique informatisés connectés à la littérature grâce à l'indexation MeSH. [7].

Connecting the guidelines to the scientific literature is indeed essential. Clinical Practice Guideline integrated with the Evidence-Based Medicine are now the basis of the Computerized Assisted Medical decision-making systems. The semantic network created with our method consists of independent units comparable to MLMs (Medical Logic Modules) of Arden Syntaxe used in the GLIF (9) (Guideline Interchange Format). The translation of the rules produced by our method can easily follow the specifications of the GEM (Guideline Elements Model) allowing to structure the textual data in the guidelines (10).

Les applications du logiciel BiKiNi sont l'aide au diagnostic médical, l'aide à la prescription de procédures diagnostiques ou thérapeutiques (guidelines), l'aide à la recherche de références bibliographiques (moteur de recherche), l'aide à la génération de test d'évaluation des étudiants (TCS).

Les avantages principaux du modèle proposé sont de représenter aussi bien les règles que les concepts et de faire apparaître la circularité de ceux-ci, de permettre l'extraction de ces connaissances directement auprès des experts et indirectement à partir d'ouvrages de référence, d'extraire les connaissances dans les différents types de raisonnements médicaux à savoir la prescription, le diagnostic ou le traitement, d'utiliser le Mesh pour établir les

convergences du réseau sémantique et les utiliser, de permettre la visualisation du réseau créé et d'être adapté à une maintenance compatible avec l'évolution rapide de la médecine.

Conclusion: This method makes it possible to collect rules directly from experts as long as it follows the course of medical reasoning (hypothetico-deductive reasoning). The indexation of items with MeSH Terms makes it easier to interface the knowledge base to the scientific literature. The method can be applied to other medical areas. It can be suited to guidelines.

Charlin B., Gagnon R., Sibert L. Van der Vleuten C. – Le test de concordance de scripts, un instrument d'évaluation du raisonnement clinique. *Pédagogie médicale* 2002 ; 3 : 135-144
Pour un enseignement stratégique. L'apport de la psychologie cognitive. Jacques Tardif. Les éditions Logiques, 1997.

BIBLIOGRAPHIE:

- [1] B. Charlin, J. Tardif, H.P.A. Boshuizen. Script and medical diagnosis knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 2000 ; 75 : 182-190.
- [2] P. Artis, Ch. Amory, Ch. Charton, P. Watteau, N. Nevels, J. Liberman, E. Fizon. Urgences dentaires dans la pratique quotidienne -Editions Techniques- *Encycl. Med. Chir. (Paris-France) . Stomatologie-Odontologie II. 23-750-A-10. 1994. 9p.*
- [3] Recommandations et références dentaires. Prescription d'antibiotiques en odontostomatologie. ANAES. 1976
- [5] UMLS Knowledge Source Server Version 4.2.3 available at : http://umlsks.nlm.nih.gov/kss/servlet/Turbine/template/admin,user,KSS_login.vm
- [4] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp.* 2001;;17-21.
- [6] O. Bodenreider, J.A. Mitchell. Graphical visualization and navigation of genetic disease information. *AMIA Annu Symp Proc.* 2003; 792
- [7] R.A. Greenes, A. Boxwala, W.N. Sloan, L. Ohno-Machado, S.R.A. Diebbel. A framework and tools for authoring, editing, documenting, sharing, searching, navigating, and executing Computed-Based Clinical Guidelines. *Proc AMIA Symp.* 1999; 261-5.