

# Conception et visualisation d'un graphe de connaissances à partir de données en hématologie : prise en charge de l'anémie chez l'adulte

S. Hodroj<sup>1</sup>, S. Despres<sup>1</sup>, J. Nobecourt<sup>1</sup>, F. Cymbalista<sup>2</sup>

<sup>1</sup> Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé – LIMICS, INSERM, UMR 1142, F- 93000, Bobigny, France.

<sup>2</sup> Université Sorbonne Paris Nord, Signalisation, Microenvironnement et Hémopathies Lymphoïdes B, SIHMEL, INSERM, UMR 978, F-93000, Bobigny, France.

2023

## Résumé

Plusieurs systèmes informatisés d'aide à la décision médicale (SADM) ont été développés pour aider les praticiens dans leurs démarches diagnostiques afin d'aboutir à une meilleure prise en charge. Cet article présente les travaux de construction d'un graphe de connaissances dans le cadre d'un SADM dédié aux médecins généralistes ou spécialistes non hématologues, dont l'objectif essentiel est de déterminer les situations urgentes en hématologie. En premier lieu, nous rappelons les notions de base concernant les différentes approches des SADM. Après avoir identifié les questions de compétences et précisé le besoin, nous décrivons la construction de ce graphe à partir des dernières données scientifiques concernant la prise en charge de l'anémie et la topologie proposée. Nous passons enfin à la validation du graphe ainsi que sa visualisation.

## Mots-clés

SADM, graphe de connaissances, visualisation, hématologie, anémie.

## Abstract

Several clinical decision support systems (CDSS) have been developed to help practitioners in their diagnostic procedures in order to achieve the best care. This article presents the construction work of a knowledge graph within the framework of a CDSS dedicated to non-hematologist physicians whose main objective is to determine the urgent situations in hematology. First, we recall the basic notions concerning the different approaches of CDSS. After having identified the skills issues and specified the need, we describe our first work on the construction of this graph based on the latest scientific data concerning the management of anemia and the proposed topology. We finally pass to the validation of the graph as well as its visualization.

## Keywords

CDSS, knowledge graph, visualization, hematology, anemia.

## 1 Introduction

L'hématologie est un domaine médical assez spécialisé qui regroupe des diagnostics de gravité très variable, et des urgences vitales qui doivent être diagnostiquées et prises en charge. En outre, les pathologies malignes hématologiques sont peu fréquentes et les médecins généralistes doivent pouvoir les identifier pour adresser rapidement le patient dans un centre spécialisé. A l'inverse, ils doivent avoir des notions claires du degré de gravité de la pathologie afin de ne pas adresser inutilement aux urgences des patients qui peuvent bénéficier d'une prise en charge en consultation programmée. Une difficulté réside également dans la distinction entre les cas qui peuvent être pris en charge par le médecin traitant et ceux qui relèvent d'un spécialiste.

Suite à ce constat, une des hématologues de l'hôpital Avicenne<sup>1</sup> a eu l'idée de créer un Système d'Aide à la Décision Médicale (SADM) en hématologie, destiné à être utilisé par les généralistes ou médecins spécialistes non hématologues lorsqu'ils sont en consultation. La finalité du projet est de fournir aux généralistes et aux médecins non hématologues un outil ergonomique capable d'identifier (i) les situations d'urgence, (ii) les situations où le généraliste peut prendre en charge le patient, et (iii) le moment où il doit, s'il doit, passer la main à un centre spécialisé.

Dans ce papier, nous décrivons nos premiers travaux sur la construction d'un graphe de connaissances à partir des dernières données scientifiques en hématologie que nous avons limitées au chapitre Anémie (ECN-Item R2C 213). Ces données proviennent des recommandations nationales, ou des sociétés savantes.

Dans ce qui suit, la section 2 présente l'état de l'art des principales notions relatives aux SADM avec leurs approches et leurs différents modes de fonctionnement afin d'identifier les fonctionnalités correspondant à notre projet. Dans la section 3, nous présentons les scénarios d'usage, et nous

1. <https://www.aphp.fr/contenu/hopital-avicenne-4>

mettons en évidence un ensemble de questions de compétences représentatives des besoins des experts en termes d'utilisation du SADM. Puis nous décrivons le modèle sémantique du graphe de connaissances, en détaillant le processus que nous avons utilisé pour la génération de ce graphe. Ensuite, nous présentons des requêtes *SPARQL* traduisant les questions de compétences dont la réponse peut être recherchée dans le graphe de connaissance produit, validant ainsi celui-ci. Enfin, la section 4 contenant la conclusion et les perspectives.

## 2 Etat de l'art

Avec le développement de l'ordinateur, des systèmes informatisés d'aide à la décision médicale, qu'elle soit diagnostique ou thérapeutique, ont été développés en fournissant au médecin les connaissances médicales dont il a besoin pour structurer sa démarche diagnostique afin d'aboutir à la meilleure prise en charge.

Selon [1], les SADM sont des outils informatiques dont le but est de fournir aux cliniciens en temps et lieux utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées, correctement filtrées et présentées, afin d'améliorer la qualité des soins et la santé des patients, en produisant les diagnostics probables de leur état clinique (aide au diagnostic) et/ou les traitements qui lui seraient adaptés (aide à la thérapeutique).

Parmi les SADM actuellement disponibles, il faut distinguer les systèmes « spécialisés » fournissant une aide dans un domaine donné, des systèmes « généralistes » qui recouvrent toute la médecine et sont évidemment plus délicats à construire [2].

### 2.1 Les différentes approches des SADM

Deux types d'approches sont utilisées pour concevoir les SADM : numérique et logico-symbolique. Les premiers SADM utilisaient des approches numériques essentiellement probabilistes. Ces approches avaient de bonnes performances. Pourtant, la pauvreté d'interface, l'absence de justification, et la présentation des résultats sous la forme de diagnostics probables n'ont pas convaincu les médecins, et ces systèmes n'ont pas été utilisés en routine clinique, mais restent toujours utilisés dans le cas de la médecine prédictive et du calcul de score. Ensuite, les SADM ont mis en œuvre l'approche logico-symbolique (décider à partir des connaissances). Généralement, ces systèmes sont constitués de trois modules principaux, la base de connaissances, la base de faits, et le moteur d'inférence [2] [3].

Dans les approches logico-symboliques, plusieurs formalismes de représentation des connaissances sont utilisés dont le pouvoir d'expressivité varie. Il s'agit essentiellement de règles de production (de la forme « Si condition Alors action »), d'arbre de décision constitué de nœuds décisionnels, et des différents types de réseaux (sémantiques, bayésiens, neurones, etc.). Actuellement les sources de connaissances utilisées sont fondées sur les recomman-

dations de pratique clinique (RPC) qui sont principalement des documents textuels. Les formalismes de représentation des connaissances ont évolué pour tenir compte des spécificités des RPC. Ainsi, après l'arbre de décision qui permet de représenter les RPC sous la forme d'une arborescence des situations cliniques possibles pour une pathologie donnée, plusieurs autres formalismes de représentation des RPC ont été élaborés : *Arden*, *EON*, *GLIF*, *Asbru*, *SAGE*, *GUIDE*, *PRODIGY*, *PROforma*, etc., ou même bien les ontologies [2] [3].

### 2.2 Les modes de fonctionnement des SADM

Certains SADM relèvent des approches « documentaires » où le médecin utilise au besoin le système en mode volontaire guidée. L'avantage de ces approches réside dans le fait que le médecin garde le contrôle du raisonnement, et dans le fait qu'elles ne dépendent pas de l'existence de données patient codées. En revanche, les inconvénients résident dans le temps mis pour trouver la réponse qui peut ne pas être compatible avec la durée d'une consultation et par le fait que le médecin soit conscient d'être en difficulté et décide de rechercher effectivement les informations nécessaires.

D'autres SADM suivent des approches « automatiques », où le raisonnement est conduit par le moteur d'inférence qui reconnaît la situation clinique d'un patient sur la base des données patient structurées extraites du dossier informatisé, et déclenche automatiquement la structure de connaissances appropriée pour identifier et proposer les décisions recommandées, qu'elles soient diagnostiques ou thérapeutiques. Néanmoins, les inconvénients de ces approches résident dans l'exclusion du raisonnement des médecins, « la fatigue aux alertes » engendrée par le manque de spécificité des alertes, ainsi que la qualité du codage de l'état du patient qui conditionne la qualité de l'aide à la décision.

A la frontière de ces modes de fonctionnement, certains systèmes proposent des approches mixtes. Ces approches laissent le choix de l'une ou l'autre des approches, ce qui convient avec la pratique clinique, où les médecins souhaitent conserver la partie noble de l'activité médicale, c'est-à-dire, le raisonnement [2] [3].

Plusieurs systèmes d'aide à la décision existent et constituent un support de choix en matière d'accessibilité, d'interactivité et de flexibilité. Certains d'entre eux proposent une aide à la décision utilisant des formulaires « intelligents » fondés sur des algorithmes décisionnels conçus à partir de données issues des recommandations médicales. Ils recréent le cheminement du raisonnement médical et donnent « en quelques clics » la réponse au problème spécifique posé. Toutefois, contrairement aux SADM classiques, certains systèmes ne sont pas intégrés aux logiciels métier du praticien : ils ne peuvent donc être sollicités qu'à la demande de celui-ci, ils n'ont pas accès aux données du dossier médical informatisé et la recommandation produite n'est pas directement transmise au logiciel de prescription [3].

## 3 Construction du graphe

### 3.1 Spécification du besoin

Il s'agit, durant cette étape, de préciser selon les objectifs et le contexte, l'approche du SADM et le mode d'exploitation de connaissances en adéquation avec notre projet. Ainsi l'approche utilisée sera logico-symbolique, et l'exploitation de connaissances sera réalisée en adoptant une approche documentaire en mode volontaire guidée, avec inclusion de système d'alerte pour les urgences (ex : ASTI Mode guidé<sup>2</sup>, Antibiocliv<sup>3</sup>, etc.). Quant à la navigation, elle se fera sous forme graphique et la forme textuelle restera réservée aux justifications et aux ressources documentaires. Parmi les systèmes existants, « Antibiocliv » et « ASTI mode guidé » sont ceux qui se rapprochent le mieux de ce que nous souhaitons faire. Ils possèdent des caractéristiques que nous souhaitons développer dans le domaine de l'hématologie : utilisation en consultation, interactivité permettant l'accès aux traitements et à des recommandations nationales offrant aux médecins la possibilité d'approfondir les propositions faites s'il le souhaite, et surtout la trace de la démarche décisionnelle qui apparaît sur l'interface sous forme d'un récapitulatif ou d'un résumé de la situation actuelle. En revanche, notre approche cherche à créer un parcours interactif fondé sur un arbre décisionnel édité par un collègue d'hématologues qui suit en majorité une démarche sous forme textuelle de question-réponse. En outre, cet outil sera spécifiquement conçu pour le domaine de l'hématologie et précisément pour la prise en charge de l'anémie. A notre connaissance, il n'existe pas de SADM s'intéressant uniquement à ce domaine et répondant aux questions de compétences des futurs utilisateurs. L'objet du travail présenté dans cet article est la construction du graphe de connaissances qui permettra de répondre aux questions de compétences. Il servira par la suite de support à la navigation dans la démarche décisionnelle. Cette construction du GC constitue une étape primordiale dans l'élaboration du système d'aide à la décision en hématologie.

### 3.2 Scénarios d'usage

Nous avons retenu deux scénarios d'usage motivant notre modélisation : identification des situations où le médecin généraliste peut prendre en charge le patient, et identification des situations d'urgence où il doit passer la main à un service spécialisé. Ainsi une grande urgence est une anémie mal tolérée avec instabilité hémodynamique. Cette tolérance dépend généralement de la concentration de l'Hb, la rapidité d'installation, et du terrain sous-jacent. Parmi les mécanismes particuliers considérés comme des urgences majeures figure l'hémolyse aiguë où le médecin généraliste doit demander le test de Coombs en urgence et passer la main à un spécialiste. Dans la section 3.5 nous utilisons cet exemple d'hémolyse aiguë dans la requête *SPARQL* afin d'aboutir à la prise en charge convenable selon l'arbre décisionnel.

2. <http://ics.upmc.fr/ASTI/private/content/index.html>

3. <https://antibiocliv.com/>

### 3.3 Questions de compétences

Afin de déterminer la spécificité des connaissances à représenter, nous avons collecté et explicité un ensemble de questions de compétences (QC) formulées par les experts en hématologie et en médecine générale dans le but de comprendre précisément leurs attentes et afin d'apporter à ces derniers une réponse adéquate en termes d'exploration des liens entre les concepts du domaine. En résumé, les questions de compétence auxquels le graphe doit permettre de répondre concernent essentiellement :

- QC1. Quelles sont les situations d'urgence hématologique en fonction d'un signe ou d'une pathologie identifiée ?
- QC2. Quelles sont les situations où le généraliste peut prendre en charge personnellement le patient ?
- QC3. Quelles sont les situations où le généraliste doit passer la main à un spécialiste ?
- QC4. Quels sont les examens d'orientation nécessaires dans la démarche diagnostique ?
- QC5. Quels traitements prescrire lorsqu'une prise en charge en hématologie n'est pas indiquée ?

### 3.4 Conception itérative et visualisation du modèle conceptuel

La construction du graphe s'est déroulée en plusieurs étapes. Nous avons sélectionné en accord avec les experts les dernières données scientifiques en hématologie fondées sur les recommandations de sociétés savantes ou d'agences nationales (Guide HAS, ECN item 213, etc.) pour constituer notre corpus.

La construction du modèle est itérative et réalisée selon une méthode agile. Elle se déroule en deux étapes principales représentées dans la figure 1 : (i) analyse des textes conduisant à un premier modèle et (ii) confrontation du modèle au cours d'entretiens avec les experts.



FIGURE 1 – Méthodologie de construction du graphe

La numérisation du corpus a permis de constituer un vocabulaire du domaine qui a ensuite été utilisé lors de la construction du modèle afin d'assurer l'interopérabilité sémantique. L'alignement sémantique aux terminologies est fait manuellement en utilisant le serveur multi-terminologique *HeTOP*. Pour les concepts médicaux, *SNO-MED CT*, la terminologie clinique la plus complète à ce jour a été utilisée, alors que pour les concepts biologiques nous avons choisi le vocabulaire de référence *LOINC* [4].

Les connaissances décrites dans le manuel d'hématologie de l'ECN (essentiellement utilisées dans l'étape diagnostique) sont synthétisées sous forme d'algorithmes de PEC (cf. Figure 2(a)).

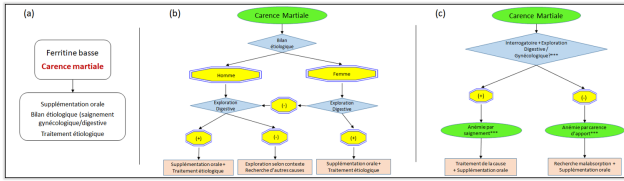


FIGURE 2 – Modification itérative de l'arbre décisionnel

Comme tout algorithme, il porte sur de textes statiques qui engendrent en fonction de l'état initial de la situation un certain déroulement d'instructions. Il s'agit d'une représentation de connaissances compilées des experts qui n'est pas directement exploitable pour construire le modèle. Les entretiens avec l'hématologue et le médecin généraliste nous ont permis d'explicitier les raccourcis de raisonnement contenus dans les algorithmes et de les confronter à la pratique quotidienne du généraliste jusqu'à l'obtention de graphes enrichis par les annotations issues des échanges avec l'hématologue et le médecin généraliste.

La figure 2 illustre le déroulement itératif de la construction du modèle. Nous prenons l'exemple de la « carence martiale » de l'algorithme des anémies microcytaires (Figure 3.5 de l'ECN item 213) pour présenter quelques étapes du travail d'acquisition des connaissances du domaine, jusqu'à la construction du modèle final de connaissances.

Dans cet exemple, l'algorithme issu des recommandations présenté dans la figure 2(a) suggère la supplémentation orale avec bilan et traitement étiologique, mais ne donne pas de détails sur la prise en charge ni sur l'ordre chronologique de la démarche. La figure 2(b) représente le premier modèle construit dans lequel nous avons ajouté le critère de genre dans la démarche thérapeutique ainsi que la démarche à faire en cas du bilan étiologique négatif. Finalement, nous avons obtenu le modèle représenté dans la figure 2(c), qui a révélé l'importance de l'interrogatoire et a précisé, dès le début, le bilan étiologique à faire, avec deux déductions introduites qui sont l'anémie par saignement et l'anémie par carence d'apport.

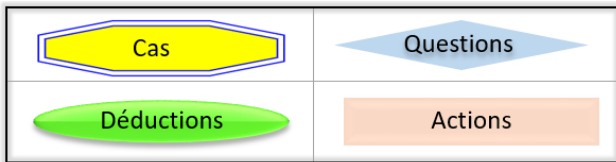


FIGURE 3 – Topologie des nœuds

Le modèle conceptuel obtenu est visualisé sous forme d'un arbre de décision. Nous codons l'arbre de décision sous la forme d'un graphe orienté. Les sommets portent les noms des noeuds (cf. Figure 3), et les arcs les relations entre ces sommets. Le chemin entre deux noeuds traduit un patron. Au cours de la construction du modèle, nous avons identifié des régularités conduisant à des patrons de construction servant à la représentation du raisonnement de l'expert : «

séquence avec cas simples » (Question -> Ensemble de cas -> Résultat), « séquence avec cas complexes » (Question -> Cas complexes issus de l'expertise -> Résultat -> Action à déclencher) et « séquences de Combinaison » se terminant normalement par le type 2[5].

L'utilisation systématique de ces patrons nous permet de montrer l'absence de certains types de nœuds comme l'exemple illustré dans la figure 2(b) où nous avons un passage direct d'une question à une autre en cas de l'exploration digestive négative. Il en va de même pour les enchaînements non respectés, les connaissances non élicitées, et les connaissances présentes mais n'intervenant pas au niveau indiqué. Ce qui explique la construction de plusieurs versions du graphe qui correspondent à l'approfondissement des connaissances sur le domaine (cf. Figure 1).

Ainsi, le contenu des algorithmes construits par les hématologues a servi pour identifier les verrous freinant le bon usage des recommandations des hématologues[5].

### 3.5 Validation du graphe de connaissances

Une étape de validation est nécessaire afin de détecter les incohérences dans le graphe et de vérifier s'il répond aux questions de compétences posées. Le graphe de connaissances est représenté sous la forme de fichiers RDF et RDF/S. Le fichier RDF/S reprend la topologie des noeuds présentés dans la figure 3 et le typage des liens. Il décrit les liens entre deux noeuds. Dans l'extrait que nous avons représenté dans la figure 4, nous cherchons *via* une requête SPARQL s'il existe une déduction liée à une action qui est une situation d'urgence. Nous utilisons la version 4.4 de Corese<sup>4</sup>.

```

Corese 4.4.0 - Inria UCA I3S - 2022-06-06
File Edit Engine Debug Query User Query Template Display Shacl Shex Event Explain
System Shacl editor Turtle editor Query1 x Query2 x Query3 x +
Query Modifier Shacl Shex Header Log Push Copy Browse to SPARQL
select.rq
1 prefix hemato: <http://www.limics.fr/hemato.rdfs#>
2 select ?nomDeduction ?nomAvis ?natureUrgence where {
3 ?uneDeduction ?r ?avis
4 ?avis hemato:estUneUrgence ?k
5 ?uneDeduction hemato:name ?nomDeduction
6 ?avis hemato:name ?nomAvis
7 ?k hemato:name ?natureUrgence
8
9

```

FIGURE 4 – Exemple d'une requête SPARQL liée à une situation d'urgence

La figure 5 montre le résultat obtenu par cette requête. Il suggère comme action de faire le test de Coombs en urgence et de demander l'avis d'un spécialiste hématologue en réponse à la déduction qui est l'hémolyse dans cet exemple.

Graph	XML/RDF	Table	Validate
1	?nomDeduction Hémolyse	?nomAvis Avis Hémato/Interniste	?natureUrgence A Coombs En Urgence ! A

FIGURE 5 – Résultat de la requête SPARQL liée à une situation d'urgence

4. <https://corese.inria.fr/>

