

# Classification des comptes-rendus mammographiques à partir d'une ontologie radiologique en OWL

Amel Boustil\*, Sahnoun Zaidi \*\*,  
Ziad Mansouri \*, Christine Golbreich\*\*\*

\* Département d'informatique, Université de Skikda, BP 26 Route El-hadaik 21000, Skikda, Algérie

boustil1710@yahoo.fr

\*\* Département d'informatique, Université Mentouri Constantine, LIRE, Constantine 2500  
Sahnounz@yahoo.fr

\*\*\* Université de Rennes 1, 35043 Rennes, France  
Christine.Golbreich@univ-rennes1.fr

**Résumé.** Dans cet article, nous proposons un système de classification des comptes-rendus mammographiques, reposant sur une ontologie radiologique décrivant les signes radiologiques et les différentes classes de la classification ACR des systèmes BIRADS dans le langage OWL. Le système est conçu pour, extraire les faits issus des textes libres de comptes-rendus en étant dirigé par l'ontologie, puis inférer la classe correspondante et en déduire l'attitude à tenir à partir de la classification ACR. Ce travail présente la construction d'une ontologie radiologique mammaire dans le langage OWL et son intérêt pour classer automatiquement les comptes-rendus de mammographies.

## 1 Introduction

Les comptes-rendus de mammographies écrits en texte libre sont difficiles à interpréter et à analyser par un programme machine. La difficulté est liée à la nature informelle de ces comptes-rendus. Trouver un processus qui permet de structurer les comptes-rendus et donner une représentation formelle de leur contenu est une tâche difficile vue la complexité du langage naturel et des connaissances médicales (Zweigenbaum, 1994).

L'objectif principal de cet article est de montrer une utilisation possible dans le domaine médical des ontologies formelles en OWL, le langage standard d'ontologie du Web (OWL, 2004). Ce travail vise à fournir un outil d'aide à l'interprétation des comptes-rendus médicaux mammographiques et à leur classification. Il a consisté d'abord à concevoir et réaliser une ontologie regroupant tous les concepts du domaine : concepts radiologiques, concepts pathologiques, et différentes classes ACR. Les classes ont été définies à partir de la classification dite ACR (ACR, 2000) et ont été représentées dans le langage OWL DL en utilisant l'éditeur Protégé et son plugin OWL (Holger, 2004). Notre système a pour tâche, d'extraire les faits correspondant au contenu des comptes-rendus de mammographies, puis, d'inférer la classe pathologique correspondante selon la classification ACR en utilisant le raisonnement par subsumption, et d'en déduire la conduite à tenir.

L'interprétation des comptes-rendus utilisée dans le cadre de ce travail est basée sur l'exploitation de la logique de description comme langage d'ontologie, comme (Golbreich, 2003) qui reposait sur C-Classic pour l'indexation et la recherche d'images, et non les graphes conceptuels comme dans le système Menelas (Zweigenbaum, 1994) ou le traitement du langage naturel comme dans le système MedLee (Nilesh et al., 1995). L'avantage de notre méthode est double. Elle permet d'une part l'utilisation du standard du Web Sémantique OWL pour la description des connaissances, donc le partage et l'interopérabilité, et d'autre part des services puissants de raisonnement basés sur la logique de description.

L'idée principale est de suivre dans chaque phrase du compte-rendu radiologique donné en texte libre la trace des concepts, instances, et propriétés de l'ontologie. Si l'extracteur localise dans une phrase bien identifiée des concepts, instances ou propriétés, il va essayer de déterminer les relations entre eux, à partir des modèles de classes fournis par l'ontologie. Ensuite, un raisonneur pour la logique de description OWL DL identifie pour l'individu associé au compte-rendu la (les) classes possibles et l'attitude à tenir qui en résulte.

Le reste de l'article est organisé comme suit : dans la deuxième section, nous décrivons la Classification ACR introduite par les systèmes BIRADS et nous motivons notre besoin de construire une ontologie radiologique dans le format OWL (§2). Ensuite nous décrivons l'ontologie ACR en OWL (§3). La dernière section présente un aperçu de l'architecture globale du système et des fonctionnalités attendues (§4).

## 2 Les systèmes BIRADS et la classification ACR

La mammographie permet d'établir un diagnostic et de donner des indications pronostiques concernant des lésions observées sur des images mammographiques. Son rôle principal est la détection précoce du cancer du sein chez des patientes asymptomatiques (ACR, 2000).

Les règles qui permettent d'établir une conclusion diagnostique et/ou pronostique à partir de caractéristiques morphologiques observées dans les images mammographiques sont publiées dans le cadre de systèmes de classification. Les systèmes BIRADS de l'American College of Radiology (ACR, 2000) standardisent une classification des images mammographiques en six catégories en fonction du degré de suspicion de leur caractère pathologique. Ils ont pour objectif de standardiser le compte-rendu grâce à une structure claire et un lexique bien défini, afin d'éviter les erreurs d'interprétation. De ce fait, construire une ontologie partageable entre pathologistes et radiologues dispersés géographiquement est crucial. L'ontologie doit regrouper des concepts utilisés par les radiologues dans leur description physique des signes radiologiques issus de deux types d'examen complémentaires : l'échographie et la mammographie ainsi que leur morphologie (forme, taille, localisation, nombre, contour, densité, etc.), des concepts décrivant les anomalies pathologiques mammaires (Kystes, carcinomes, etc.) et les différentes classes données par la classification ACR décrivant la catégorie et l'attitude à tenir.

Dans une première étape, une construction manuelle ou supervisée, élaborée suite à une analyse approfondie de la classification normalisée ACR, une consultation d'une base de comptes-rendus mammographiques et des documents biomédicaux, combinée à des interviews d'experts, a permis d'identifier les principales classes du domaine impliquées dans la classification ACR et de définir une taxonomie des concepts du domaine présentée dans la figure 1. La deuxième étape a été de définir la représentation logique en OWL DL des différentes classes ACR et autres concepts.

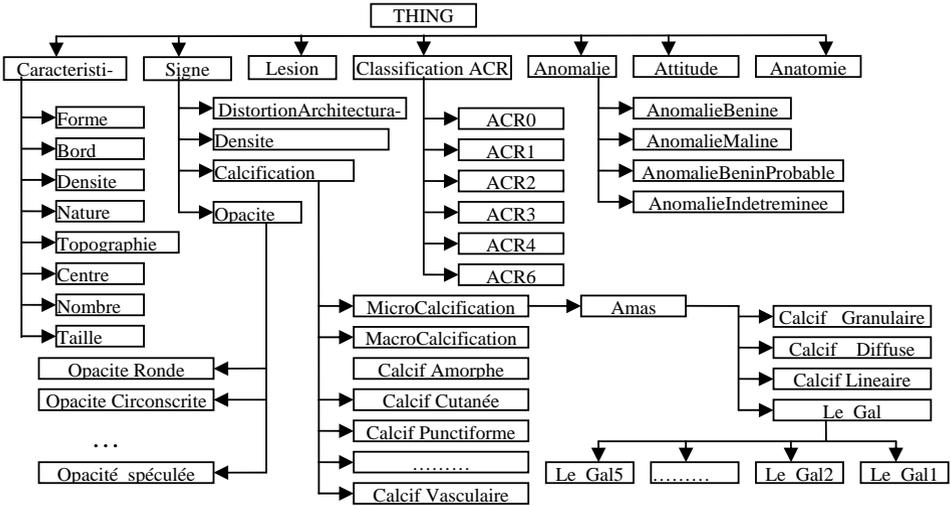


FIG. 1 – Taxonomie partielle de l’ontologie ACR

### 3 L’ontologie ACR en OWL

En OWL, chaque concept est représenté par une classe ‘définie’ ou ‘primitive’. Il existe deux types de conditions, les conditions nécessaires et les conditions nécessaires et suffisantes (CNS). Une condition nécessaire est représentée en OWL par un axiome `subclassOf`:  $A \subset B$  indique que si un individu appartient à la classe A, il vérifie nécessairement la condition exprimée par B (B peut être une classe anonyme). Une condition nécessaire et suffisante est représentée en OWL par un axiome `equivalentClass`:  $A \equiv B$  où B est une CNS d’appartenance à la classe A. Dans ce cas la classe A est dite ‘définie’.

(a) **Les classes primitives** : 6 Classes ‘primitives’ ont été définies pour les concepts racines, au premier niveau de la figure 1: Anomalie, Attitude, Caractéristique Physique, Lésion, Signe, Anatomie. Notre premier travail a été d’identifier ces concepts et de trouver une taxonomie (Noy, 2001) assez puissante des signes de diagnostic en se basant sur les caractéristiques physiques e.g., morphologie, puis de donner les formules logiques combinant les signes radiologiques et les lésions correspondantes pour exprimer les conditions nécessaires et suffisantes ou nécessaires (Golbreich et al 2004) pour chaque Classe ACR.

(b) **Les propriétés** : puis nous avons défini les propriétés. Un concept peut être relié à d’autres concepts de l’ontologie par des propriétés OWL comme la forme, le bord, le signe, représentées par des `objectProperty` en OWL e.g., `hasForm`, `hasBord`, `hasSign`, etc. Le tableau 1 décrit le domaine de départ et d’arrivée (`domain`, `range`) de quelques relations. Il est également possible de spécifier certaines caractéristiques d’une propriété, comme par exemple son inverse à l’aide de la primitive OWL `inverseOf`

Propriété	Domain	Range	Inverse
hasAnomaly	ACR	Anomalie	
hasAttitude	ACR	Attitude	
hasForm	Calcification Opacite	Forme	
hasSign	Anomalie	Signe	IsSignOf
hasDensity	Anomalie	Densite	
hasOpacity	Anomalie	Opacite	
isSignOf	Signe	Anomalie	hasSign

TAB. 1 – Quelques exemples de propriétés de l’ontologie ACR.

(c) **Les classes ‘définies’** : Toutes les classes ACR sont représentées par des classes définies e.g., la classe ACR 2 ci-dessous.

**ACR 2 : Il existe des anomalies bénignes ne nécessitant ni surveillance ni examen complémentaire :**

[L1] • Opacité ronde avec macrocalcifications (adénofibrome ou kyste)

[L2] • Ganglion intramammaire

[L3] • Opacité(s) ronde(s) correspondant à un/des kyste(s) typique(s) en échographie

[L4] • Image(s) de densité graisseuse ou mixte (lipome, hamartome, galactocèle, kyste huileux)

[L5] • Cicatrice(s) connuet(s) et calcification(s) sur matériel de suture

[L6] • Macrocalcifications sans opacité (adénofibrome, kyste, adiponécrose, ectasie canalaire sécrétante, calcifications vasculaires, etc.)

[L7] • Microcalcifications annulaires ou arciformes, semi-lunaires, sédimentées, rhomboédriques

[L8] • Calcifications cutanées et calcifications punctiformes régulières diffuses

La classe ACR2 est interprétée comme suit. Chaque ligne L1 à L8 est considérée comme décrivant une sous-classe de la classe ACR2, décrite par le biais des autres concepts et propriétés. Par exemple, la ligne L1 correspond à une anomalie de type 1. L’image présente une Anomalie 1 si et seulement si le compte-rendu indique qu’il existe un signe radiologique de type Opacité ronde et une Macrocalcification, d’où la définition. Un kyste ou un adénofibrome donne une anomalie 1, d’où l’implication :

$$\text{Kyste} \sqcup \text{Fibroadenome} \sqsubset \text{Anomalie}_1$$



FIG. 2 – Conditions Nécessaires Et/Ou suffisantes De Anomalie1

Les autres anomalies sont représentées de la même manière, et la classe ACR2 est définie comme l’union des sous-classes d’anomalies de 1 à 8 :

$$\text{ACR2} \equiv \text{Anomalie}_1 \sqcup \text{Anomalie}_2 \sqcup \text{Anomalie}_3 \sqcup \text{Anomalie}_4 \sqcup \text{Anomalie}_5 \sqcup \text{Anomalie}_6 \sqcup \text{Anomalie}_7 \sqcup \text{Anomalie}_8$$

L’axiome suivant exprime que les anomalies de la classe ACR2 impliquent ni surveillance ni examen complémentaire :  $\text{ACR2} \sqsubset \neg(\text{Surveillance} \sqcup \text{ExamenComplementaire})$

La vérification de la consistance et la classification de l'ontologie sont réalisées automatiquement à l'aide de Racer (Haarslev et al., 2001) dans l'environnement Protégé OWL.

## 4 L'application

L'idée principale de notre application réside dans l'extraction d'instances et de leurs propriétés à partir des comptes-rendus en utilisant des techniques de traitement du langage naturel comme dans (Ricky et al., 2001). Contrairement à cette approche nous exploitons une représentation formelle extraite des comptes-rendus et le raisonnement ontologique par Racer pour identifier les classes ACR des comptes-rendus. L'architecture globale et ses différents composants sont décrits figure 3 :

**L'analyse structurelle** : l'analyse structurelle a pour rôle de reconnaître la structure du compte-rendu : entête, dates, informations patients, contenu, etc.

**L'analyse lexicale** : Le but de cette analyse est d'extraire tous les concepts et les propriétés décrits dans le champ contenu. Une représentation intermédiaire en XML a été attribuée.

**L'analyse sémantique** : Le but de cette analyse est de trouver les liens entre les concepts et leurs propriétés en utilisant l'ontologie. La tâche la plus sophistiquée est de déterminer le domaine (et co-domaine) de chaque propriété.

**L'analyse inférentielle** : A partir du fichier XML résultat de l'analyse sémantique, nous construisons la conjonction des concepts et propriétés et testons si l'individu créé appartient à une Classe ACR ou non. Ce mécanisme correspond à identifier les classes de l'ontologie auquel un individu appartient et se fait à l'aide d'un raisonneur comme Racer.

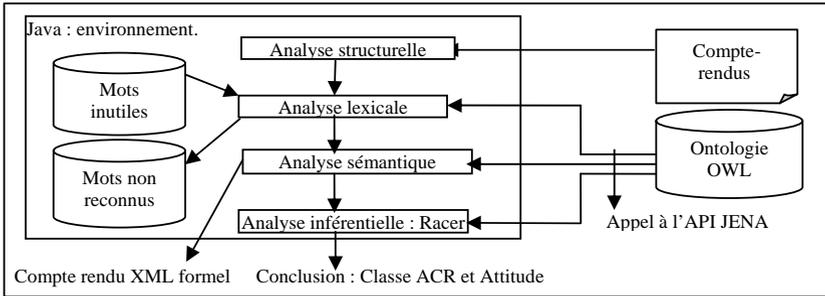


FIG. 3 – L'architecture globale du système

## 5 Conclusion

Dans le présent travail, nous avons conçu un système permettant la classification des comptes-rendus mammographiques en utilisant une ontologie formelle pour classer les images radiologiques-mammographiques à partir des signes radiologiques observés et d'une Classification normalisée dite ACR. L'ontologie est représentée en OWL DL à l'aide de Protégé OWL. Un raisonneur comme Racer sera utilisé pour identifier les classes possibles d'appartenance pour les images présentant des signes d'anomalies. L'application est développée en Java en utilisant plusieurs API. Les analyseurs lexical, syntaxique et sémantique

sont déjà prêts, l'analyseur inférentiel est en cours. Une phase de test est prévue pour évaluer l'approche sur une base de comptes-rendus réelle. L'idée à plus long terme est de combiner techniques d'analyse d'image et symbolique en vue d'un outil d'aide aux radiologues.

## Références

- ACR Classification (2000). [http://www.anaes.fr/anaes/Publications.nsf/nPDFFile/R\\_E\\_LILF-5CJKY4/\\$File/ACR.pdf?OpenElement](http://www.anaes.fr/anaes/Publications.nsf/nPDFFile/R_E_LILF-5CJKY4/$File/ACR.pdf?OpenElement).
- Golbreich C., Mercier, S. (2004). Construction of the dialysis and transplantation ontology, advantages, limits, and questions about Protégé OWL. 7th International Protégé Conference, Bethesda.
- Golbreich C. (2003). Vers un moteur de recherche évolué de documents multimédia par le contenu. Rapport interne Université Rennes 2, 2000. Bulletin de l'AFIA N°55, 2003.
- Haarslev V., Möller R. (2001). Description of the RACER System and its Applications. Description Logics.
- Holger K. (2004). The Protégé OWL Plugin. 7th International Protégé Conference, Bethesda.
- Nilesh L. Jain, D.Sc, Carol Friedman (1995). Identification of Findings Suspicious for Breast Cancer Based on Natural Language Processing of Mammogram Reports. Proc AMIA Annu Fall Symp. 829-33.
- Noy N. and D.L. McGuinness (2001). Ontology developemnt 101: A guide to creating your first ontology. Tehnical Report SMI-2001-0880, Stanford University. CA 94305.
- OWL Web Ontology Language Reference (2004). W3C Recommendation 10 February. <http://www.w3.org/TR/owl-ref/>
- Ricky K., Taira, Stephen G. Soderland, and Rex M. Jakobovits (2001). Automatic Structuring of Radiology Free-Text Reports, Radiographics, 21:237-245.
- Zweigenbaum P., Consortium Menelas (1994). MENELAS: An Access System for Medical Records Using Natural Language. Computer Methods and Programs in Biomedicine, 45: 117-120.

## Summary

This paper, presents a system for automatically classifying mammography reports, based on a radiological ontology, which describes the radiological signs and classes of the ACR classification in the OWL language. Our system is designed to extract facts from the full text of mammography reports driven by the ACR ontology, then to deduce the relevant classes and the corresponding attitude. We show how to construct a radiological ontology in OWL and how to use it for the classification of mammography reports.